Check for updates

RESEARCH ARTICLE

REVISED Evaluation of electrocardiogram: numerical vs. image data for emotion recognition system [version 2; peer review: 2 approved, 1 approved with reservations]

Sharifah Noor Masidayu Sayed Ismail [ID]1, Nor Azlina Ab. Aziz [ID]2,
Siti Zainab Ibrahim1, Sophan Wahyudi Nawawi3, Salem Alelyani4,5,
Mohamed Mohana [ID]4, Lee Chia Chun6

1Faculty of Information Science & Technology, Multimedia University, Bukit Beruang,, Melaka, 75450, Malaysia
2Faculty of Engineering, Multimedia University, Bukit Beruang, Melaka, 75450, Malaysia
3School of Electrical Engineering, Faculty of Engineering, Universiti Teknologi Malaysia, Skudai, Johor Bahru, 81310, Malaysia
4Center for Artificial Intelligence, King Khalid University, Abha, 61421, Saudi Arabia
5College of Computer Science, King Khalid University, Abha, 61421, Saudi Arabia
6Hexon Data Sdn Bhd, Kuala Lumpur, 59200, Malaysia

Abstract

Background: The electrocardiogram (ECG) is a physiological signal used to diagnose and monitor cardiovascular disease, usually using 2-D ECG. Numerous studies have proven that ECG can be used to detect human emotions using 1-D ECG; however, ECG is typically captured as 2-D images rather than as 1-D data. There is still no consensus on the effect of the ECG input format on the accuracy of the emotion recognition system (ERS). The ERS using 2-D ECG is still inadequately studied. Therefore, this study compared ERS performance using 1-D and 2-D ECG data to investigate the effect of the ECG input format on the ERS.
Methods: This study employed the DREAMER dataset, which contains 23 ECG recordings obtained during audio-visual emotional elicitation. Numerical data was converted to ECG images for the comparison. Numerous approaches were used to obtain ECG features. The Augsburg BioSignal Toolbox (AUBT) and the Toolbox for Emotional feature extraction from Physiological signals (TEAP) extracted features from numerical data. Meanwhile, features were extracted from image data using Oriented FAST and rotated BRIEF (ORB), Scale Invariant Feature Transform (SIFT), KAZE, Accelerated-KAZE (AKAZE), Binary Robust Invariant Scalable Keypoints (BRISK), and Histogram of Oriented Gradients (HOG). Dimension reduction was accomplished using linear discriminant analysis (LDA), and valence and arousal were classified using the Support Vector Machine (SVM).
Results: The experimental results show 1-D ECG-based ERS achieved

Open Peer Review

Approval Status ? ✔ ✔

	1	2	3
version 2 (revision) 30 May 2022		✔ view	✔ view
version 1 04 Nov 2021	? view	✘ view	? view

1. **Umesh Chandra Pati** [ID], National Institute of Technology, Rourkela, India

2. **Marios Fanourakis**, University of Geneva, Geneva, Switzerland

3. **Md. Asadur Rahman** [ID], Military Institute of Science and Technology (MIST), Dhaka, Bangladesh

Any reports and responses or comments on the article can be found at the end of the article.

65.06% of accuracy and 75.63% of F1 score for valence, and 57.83% of accuracy and 44.44% of F1-score for arousal. For 2-D ECG-based ERS, the highest accuracy and F1-score for valence were 62.35% and 49.57%; whereas, the arousal was 59.64% and 59.71%.

Conclusions: The results indicate that both inputs work comparably well in classifying emotions, which demonstrates the potential of 1-D and 2-D as input modalities for the ERS.

Keywords

Emotion recognition, electrocardiogram, numerical ECG, image ECG, DREAMER

This article is included in the Research Synergy Foundation gateway.

Corresponding authors: Sharifah Noor Masidayu Sayed Ismail (sharifahayuismail@gmail.com), Nor Azlina Ab. Aziz (azlina.aziz@mmu.edu.my), Siti Zainab Ibrahim (sitizainab.ibrahim@mmu.edu.my)

Author roles: Sayed Ismail SNM: Data Curation, Formal Analysis, Investigation, Methodology, Software, Visualization, Writing – Original Draft Preparation; **Ab. Aziz NA**: Conceptualization, Data Curation, Funding Acquisition, Investigation, Project Administration, Resources, Software, Supervision, Validation, Writing – Review & Editing; **Ibrahim SZ**: Conceptualization, Data Curation, Investigation, Project Administration, Resources, Software, Supervision, Validation, Writing – Review & Editing; **Nawawi SW**: Funding Acquisition, Supervision, Validation; **Alelyani S**: Funding Acquisition, Supervision, Validation; **Mohana M**: Funding Acquisition, Supervision, Validation; **Chia Chun L**: Funding Acquisition, Supervision, Validation

Introduction

Medical professionals have been actively using electrocardiogram (ECG) wave images as a tool for monitoring[1–3] and diagnosing cardiovascular diseases,[4–6] such as heart attacks, dysrhythmia, and pericarditis, with some reported accuracy of more than 99% in the past decade. Fundamentally, ECG is used to measure electrical activity in the human heart by attaching electrodes to the human body. Due to the continual blood pumping action to the body, the electrical activity of the heart can be found in the sinoatrial node. The electrocardiogram signal is composed of three basic components: P, QRS, and T waves (Figure 1). P waves are produced during atrium depolarization, QRS complexes are produced during ventricular depolarization, and T waves are produced during ventricle recovery.

Today's ECG devices have advanced from large and immobile to compact, wearable, and portable. Additionally, the signal accuracy of portable devices is comparable to that of traditional medical devices and can be used for the same purposes as traditional devices, including the study of human emotions. Many studies have proven that ECG which is associated with autonomic nervous system's (ANS) physiological responses can be used to identify human emotions.[8–11] Different emotions influence human heart activities differently; these influences may be hidden in the ECG wave and can be detected through closer monitoring of the main features of ECG, namely, heart rate (HR) and heart rate variability (HRV).

Previous research on human emotions has primarily relied on either direct analysis of 1-D data[12–14] or the conversion of 1-D data to a 2-D spectral image[15] prior to identifying the emotions. Despite this, majority of the portable devices record the ECG signal as images (2-D images) in a PDF file rather than as raw numerical data (1-D data).[16–18] The example of a PDF-based 2-D ECG is depicted in Figure 2. Due to this problem, researchers were required to convert the PDF file of the ECG into 1-D data before performing further emotion analysis, adding complexity to the pre-processing process. On this account, given the positive results obtained in monitoring and diagnosing cardiovascular-related diseases, the efficacy of 2-D ECG in emotion studies also warrants further investigation.
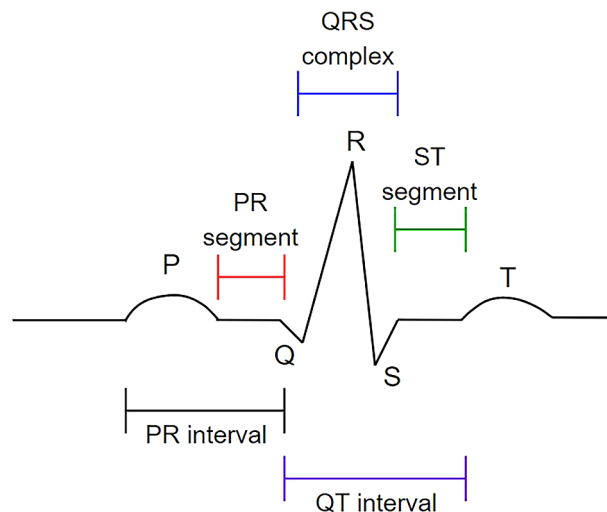

Figure 1. The P wave, QRS complex, and T wave in the standard electrocardiogram (ECG). This figure has been reproduced with permission from Ref. 7.
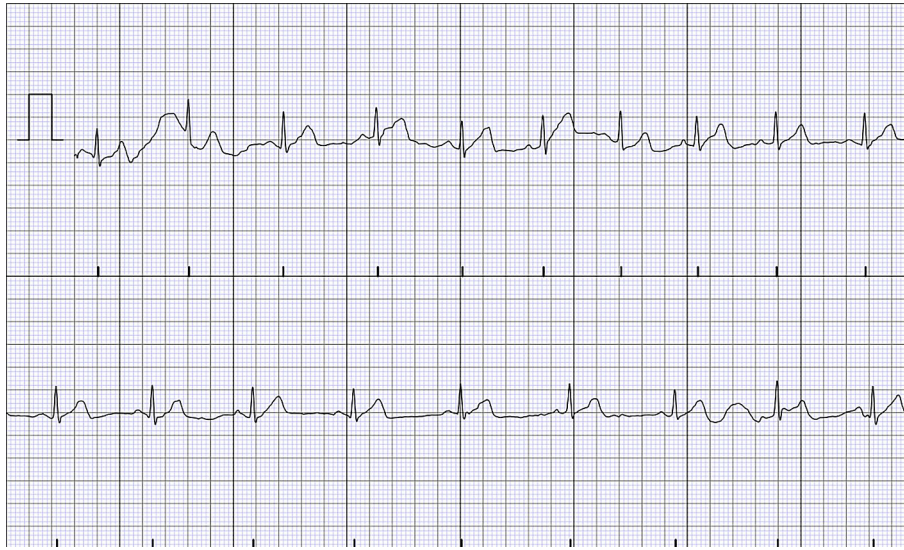
Figure 2. The example of 2-D ECG in a PDF file.

As far as our knowledge is concerned, despite numerous attempts to recognize emotions using ECG signals, the effects of employing various types of ECG inputs to recognise emotions in the emotion recognition system (ERS) have yet to be closely studied. In addition, there is no consensus on whether or not the type of ECG input format affects the emotion classification accuracy. Therefore, to address this gap, the contribution of this study is to compare emotion classification performance using 1-D and 2-D ECGs to investigate the effect of the ECG input format on the ERS.

This study analysed ECG data from the DREAMER dataset, a multimodal database. In DREAMER, ECG signals were recorded from 23 participants using 18 audio-visual stimuli for the elicitation of various emotions. The Augsburg BioSignal Toolbox (AUBT)[19] and the Toolbox for Emotional Feature Extraction from Physiological Signals (TEAP)[20] were used to help extract features from the 1-D ECG. Prior to emotion classification, the dimension of the extracted ECG features was reduced using linear discriminant analysis (LDA). On the other hand, the 2-D ECG was obtained by converting the 1-D ECG, and six different feature extractors were used to extract features from the 2-D ECG, namely Oriented FAST and Rotated BRIEF (ORB), Scale Invariant Feature Transform (SIFT), KAZE, Accelerated-KAZE (AKAZE), Binary Robust Invariant Scalable Keypoints (BRISK), and Histogram of Oriented Gradients (HOG). The Support Vector Machine (SVM) classifier is used, and the ERS results for both ECG inputs are compared to examine the effect of signal input on ERS performance. The finding indicates no substantial difference between the two ECG inputs since both produce a promising outcome within same range of accuracy for emotion recognition.

The next section discusses related works. The following section describes the dataset and the proposed methods in depth. The results are then provided. Finally, the study is concluded in the final section.

Related works
Researchers in the emotion recognition field have been proposing multiple approaches using electrocardiogram signals. For instance, Minhad, Ali, and Reaz[21] used 1-D ECG to classify emotions of happiness and anger. They achieved 83.33% accuracy using the SVM classification method. Besides, Tivatansakul and Ohkura[22] used 1-D ECG from the AUBT dataset to detect emotions for the emotional healthcare system. K-Nearest Neighbour (KNN) successfully classified three emotions (joy, anger, and sadness) with an accuracy 85.75%, 82.75%, and 95.25%, respectively. The MPED database for ERS was proposed by Song *et al.*[23] using ECG numerical data to recognise discrete emotions (joy, humour, disgust, anger, fear, sadness, and neutrality). Attention Long Short-Term Memory (A-LSTM) was used as a feature extractor to extract the frequency and time-domain features from the physiological signal. The A-LSTM was used as a classifier along with SVM, KNN, and Long Short-Term Memory (LSTM). Averagely, A-LSTM achieved better results of 40% to 55% compared to those of other classifiers.

Katsigiannis and Ramzan[13] suggested that ERS should use low-cost and off-the-shelf devices to collect ECG signals based on numerical format. Their dataset was called DREAMER, and the classification using SVM with a radial basis function kernel successfully achieved 62.37% for valence and arousal. This dataset is adopted here. Additionally, numerous other researchers also used the ECG signals from the DREAMER dataset to perform emotion recognition.

For instance, 1-D ECG data from the DREAMER dataset is utilized by Wenwen He et al.[24] that suggested an approach for emotion recognition using ECG contaminated by motion artefacts. The proposed approach improved classification accuracy by 5% to 15%. Additionally, Pritam and Ali[25] also employed 1-D ECG from the DREAMER dataset to develop the self-supervised deep multi-task learning framework ERS, which consists of two stages of learning: ECG representation learning and emotion classification learning. The accuracy gained in this study was greater than 70%. Hasnul et al.[12] also used the 1-D ECG by DREAMER dataset to compare the performance of two feature extractor toolboxes. They noted that the dataset's size and the type of emotion classified might affect the suitability of the extracted features.

As mentioned before, the 2-D ECG was widely used for a variety of other purposes, including human authentication, ECG classification, and cardiovascular-related diseases. For example, Ref. 26 and Ref. 27 developed authentication systems based on printout-based 2-D ECG and 2-D ECG spectral images that achieved greater than 99% accuracy using CNN. Additionally, Klosowski et al.[28] reached the highest accuracy rate of 100% by classifying ECG signals into several categories, including normal ECG, brachy cardia, and premature ventricular contraction (PVC). Meanwhile, Ref. 4 and Ref. 29 employ 2-D ECG to detect and diagnose cardiovascular disease, specifically myocardial infarction (MI) and arrhythmia. Additionally, Mandal et al.[5] published a study comparing 1-D and 2-D ECGs for the diagnosis of Ventricular Arrhythmia. They concluded that both ECG inputs are effective at detecting the disease.

Despite rising popularity among medical practitioners in assessing patients' cardiac disease, 2-D ECG remains inadequate compared to 1-D ECG usage as a type of input in emotion recognition studies. As a result, the number of studies employing 1-D ECG in ERS is higher than that utilizing 2-D ECG in ERS. However, rather than employing a printout-based 2-D ECG, emotion researchers classified human emotions using 2-D ECG spectral images. For example, Ref. 15 determines the R-peaks of the electrocardiogram prior to generating the R-R interval (RRI) spectrogram. Following that, CNN was used to classify the emotions, with an accuracy rate greater than 90%. Elalamy et al.[30] used ResNet-50 to extract features from a 2-D ECG spectrogram. Then, Logistic Regression (LR) was employed as a classifier and achieved an accuracy of 78.30% in classifying emotions.

Table 1 summarises these works, including the reference to the work, the dataset details (number of participants, number of stimuli), the signal used, the ECG input, the purpose of the work, the features extracted, the classifiers, and their accuracy—the accuracy denoted by an asterisk (*) refers to the accuracy of works that do not mainly focus on ERS.

Although considerable research has been conducted using ECG for ERS, the majority of research has focused on 1-D ECG analysis rather than 2-D ECG analysis, despite the fact that systems based on 2-D ECG have achieved excellent results in detecting cardiovascular-related diseases and human authentication. Additionally, no comparison of 1-D and 2-D ECG input was found in the emotion studies. As a result, it is unknown whether the ECG input format has an effect on the ERS's emotional classification accuracy. The significance of this study is that it compares emotion classification performance between 1-D and 2-D ECGs to determine the effect of the ECG input format on the ERS.

Methods
In this section, the details of the dataset are described, and the experimental setup for 1-D and 2-D ECGs is explained. The current study began in September 2020. MATLAB version 9.7 was utilized for data conversion and feature extraction, whereas Python version 3.8.5 was used for feature dimension reduction (for 1-D ECG) and classification. The analysis code used in this study is available from GitHub and archived with Zenodo.[47]

The dataset (DREAMER)
This study used ECG signals from Katsigiannis and Ramzan[13] called DREAMER. The DREAMER dataset is a freely accessible database of electroencephalogram (EEG) and electrocardiogram (ECG) signals used in emotion research. However, EEG signals were removed from this study because the primary focus is on ECG signals. The ECG was recorded using the SHIMMER ECG sensor at 256 Hz and stored in 1-D format. The DREAMER dataset contains 414 ECG recordings from 23 subjects who were exposed to 18 audio-visual stimuli designed to evoke emotion. Each participant assessed their emotions on a scale of 1 to 5 for arousal, valence, and dominance. However, because this study was primarily concerned with arousal and valence ratings, participants' evaluations of dominance were discarded. The summary of the DREAMER dataset is tabulated in Table 2.

Experimental setup
1) 1-D ECG

The proposed ERS for 1-D ECG consists of three stages: feature extraction, feature dimension reduction, and emotion classification. The structure of the proposed 1-D ECG-based ERS is illustrated in Figure 3.

Table 1. The summary of existing works using 1-D and 2-D ECG input.

Ref	Dataset	Signal used	ECG input	Purposes	Feature extracted	Classifier	Result (%)
21	Own dataset 69 subjects, 20 stimuli	ECG	1-D	ERS	Statistical features from the time and frequency domains	SVM, NB, KNN, Gaussian	SVM – 69.23 NB – 53.83 KNN – 61.83 Gaussian – 70.00
22	AUBT	ECG	1-D	ERS	Local pattern description using Local Binary Pattern (LBP) and Local Ternary Pattern (LTP)	KNN	LBP – 84.17 LTP – 87.92
23	(MPED) 23 subjects, 28 stimuli	ECG	1-D	ERS	Statistical features from the time and frequency domains	SVM, KNN, LSTM, A-LSTM	SVM – 42.66 KNN – 40.02 LSTM – 49.77 A-LSTM – 51.66
13	(DREAMER) 23 subject, 18 stimuli	ECG, EEG	1-D	ERS	Statistical features from the time and frequency domains	SVM, KNN, LDA	Valence – 62.37 Arousal – 62.37
24	DREAMER	ECG	1-D	ERS	Statistical features from the time, frequency, time-frequency domains, and nonlinear analysis-related	SVM	Valence – 86.09 Arousal – 87.80
25	DREAMER	ECG	1-D	ERS	Deep-learning	Convolutional Neural Network (CNN)	Valence – 74.90 Arousal – 77.10
12	DREAMER	ECG	1-D	ERS	Statistical features from the time and frequency domains	SVM	Valence – 65.80 Arousal – 65.40
26	(MWM-HIT) 100 subjects	ECG	2-D	Authentication System	PQRST peaks	CNN	99.99*
27	PhysioNet dataset (Fantasia and ECG-ID)	ECG	2-D spectral	Authentication system	Spectrogram	CNN	99.42*
28	Own dataset generated by FLUKE "ProSim 4 Vital Sign and ECG Simulator"	ECG	2-D spectral	ECG classification	Instantaneous frequency and spectral entropy	LSTM	100*
4	Zhejiang dataset	ECG	2-D	Myocardial infarction screening	Object detection	DenseNet, KNN, SVM	DenseNet – 94.73* KNN – 89.84* SVM – 92.19*
29	MIT-BIH arrhythmia dataset	ECG	2-D spectral	Arrhythmia classification	Local features from 2-D images using deep learning	CNN	99.11*
5	Physiobank dataset	ECG	1-D and 2-D	Ventricular Arrhythmia detection	ECG beat images	SVM, Probabilistic Neural Network (PNN), KNN, Random Forest (RF)	99.99* (both are useful)
15	Own dataset 11 subjects, 6 stimuli	ECG and EEG	2-D spectral	ERS	Statistical features from the time and frequency domains (R-R interval spectrogram)	CNN	ECG – 91.67 EEG – 90.00
30	AMIGOS, DEAP	ECG, PPG, EDA	2-D spectral	ERS	Features extracted from spectrogram by ResNet-50	Logistic Regression	AMIGOS – 78.30 DEAP – 69.45

Table 2. The summary of the DREAMER dataset.

No of subject	23
No of videos	18 audio-visual stimuli
Type of stimuli	Audio-video
Used Signal (Hz)	ECG (256)
Rating scales	Valence, Arousal
Rating values	1–5
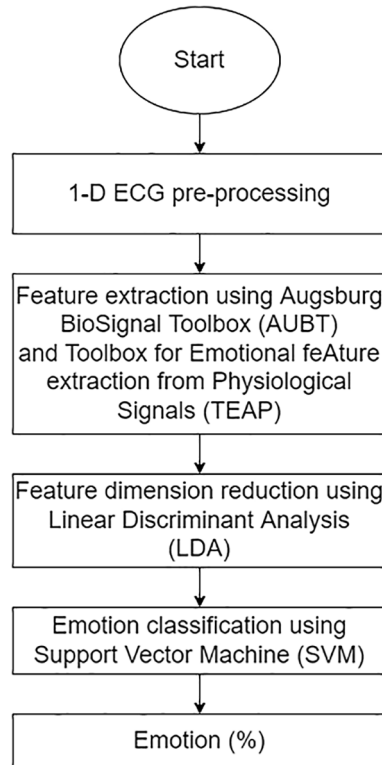
Figure 3. The overall structure of 1-D ECG-based ERS.

Two open-source toolboxes, namely, Augsburg BioSignal Toolbox (AUBT)[19] and Toolbox for Emotional feature extraction from Physiological signals (TEAP),[20] were employed to facilitate feature extraction from the ECG signals. AUBT provides tools for the analysis of physiological signals such as the ECG, RESP, EMG, and GSR. These tools are available for Windows with MATLAB 7.1. On the other hand, TEAP is compatible with the MATLAB and Octave software packages operating on Windows and can analyse and compute features from physiological data such as EEG, GSR, PPG, and EMG.

The AUBT and TEAP feature extractors were included with the Low Pass Filter (LPF), a filter meant to reject all undesirable frequencies in a signal. The LPF was one of the most widely used filters before the computation of statistical features for physiological signals.[31,32] As a result, automated 1-D ECG pre-processing utilizing LPF was performed in this study to reduce muscle and respiratory noise in ECG signals.

AUBT extracted 81 features in the time and frequency domains from each 1-D ECG signal, including the mean, median, and standard deviation for each PQRST wave, HRV, frequency spectrum range, and amplitude signal. Sixteen (16) statistical features are extracted, including mean, IBI, HRV, and multiscale entropy in the time and frequency domains. Table 3 and Table 4 provide abbreviations and descriptions of AUBT and TEAP features, respectively.

Table 3. Features extracted from Augsburg Bio-signal Toolbox (AUBT).

Features	Description
P, Q, R, S, T	P-, Q-, R-, S-, T-peaks (ECG)
HRV	Heart rate variability
Ampl	Amplitude signal
Mean	Mean value
Median	Median value
Std	Standard deviation
Min	Minimum value
Max	Maximum value
SpecRange	Mean of the frequency spectrum in a given range

Table 4. Features extracted from Toolbox for Emotional feature extraction from Physiological signals (TEAP).

Features	Description
meanIBI	Mean inter-beat interval
HRV	Heart Rate Variability
MSE	Multiscale entropy at 5 levels
sp0001/0102/0203/0304	Spectral power 0-0.1 Hz, 0.1-0.2 Hz, 0.2-0.3 Hz, 0.3-0.4 Hz
energyRatio	Spectral energy ratio between f<0.08 Hz/f>0.15 Hz and f<5.0 Hz
tachogram_LF/MF/HF	Spectral power in tachogram (HRV) for low, medium, and high frequencies.
tachogram_energy_ratio	Energy ratio for tachogram spectral content (MF/(LF+HF))

Additionally, to prevent the "Curse of Dimensionality," dimensionality reduction is employed to further reduce the number of high-dimensional features to low-dimensional features. The dimensions of the features were decreased using linear discriminant analysis (LDA), a well-known approach for reducing the dimensions of features.[33] LDA is a supervised algorithm that can reduce dimensionality while retaining as much class-discrimination information as possible. Following that, the low-dimensional features were fed into a Support Vector Machine (SVM) classifier for emotion classification. The following section will outline the classifying process.

2) 2-D ECG

The duration of the ECG recording varies according to the duration of the video (average = 199 seconds). As Katsigiannis and Ramzan proposed, this study analysed the final 60 seconds of each recording to allow time for a dominant emotion to emerge.[13] Following that, 1-D ECG was pre-processed using a simple MATLAB function by Ref. 34 to eliminate baseline wander caused by breathing, electrically charged electrodes, or muscle noise. The signal was then divided into four segments corresponding to 15 seconds each. Then, using MATLAB version 9.7, the 1-D ECG was transformed into a 2-D ECG (Figure 4). The image has a width of 1920 pixels and a height of 620 pixels.

Due to the fact that the 2-D ECG was converted to a rectangle shape, it is not easy to resize the photos to the standard input image sizes of 224×224 and 299×299. As a result, the converted 2-D ECG was resized to 60% of its original size using Python version 3.8.5. This scale percentage was chosen after considering the quality of the image, the type of feature extractor used, and the computational cost the system can afford. The coloured images were converted into greyscale images. Then, binarization of the image using an Otsu's automatic image thresholding method[35] was done. This method ascertains the optimal threshold values from pixel values of 0 to 255 by calculating and evaluating their within-class variance.[36]

The area of interest for 2-D ECG is laying on the PQRST waves, making the peaks detector the best approach to be employed. Therefore, six different feature extractors that could extract peaks, edges, or corners were applied to extract features from 2-D ECGs using Python version 3.8.5:
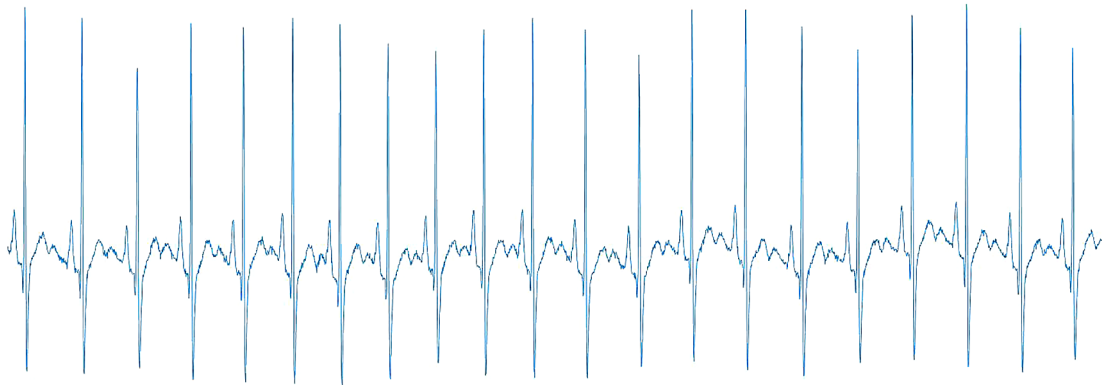
Figure 4. The 2-D ECG converted from 1-D ECG.

1. ORB[37]: ORB features are invariant to rotation and noise because they are a combination of Features from Accelerated Segment Test (FAST) detection and Binary Robust Independent Elementary Features (BRIEF) description methods.

2. SIFT[38]: SIFT identifies feature points by searching for local maxima on the images using Difference-of-Gaussians (DoG) operators. The description approach generates a 16x16 neighbourhood around each identified feature and sub-blocks the region. SIFT is also rotation and scale invariant.

3. KAZE[39]: KAZE is based on the scale of the normalised determinant of the Hessian Matrix, with the maxima of detector responses being captured as feature points using a moving window. Additionally, KAZE makes use of non-linear space via non-linear diffusion filtering to reduce noise while keeping the borders of regions in images.

4. AKAZE[40]: AKAZE is a more sophisticated version of KAZE that is based on the Hessian Matrix determinant. Scharr filters were employed to enhance the quality of the invariance rotation, rendering AKAZE features rotation- and scale-invariant.

5. BRISK[41]: While searching for maxima in the scale-space pyramid, BRISK detects corners using the AGAST algorithm and filters them using the FAST Corner Score. Additionally, the BRISK description is based on the recognised characteristic direction of each feature, which is necessary for rotation invariance.

6. HOG[42]: HOG is a feature descriptor that is used to compute the gradient value for each pixel. The image shape denoted the edge or gradient structure derived using a high-quality local gradient intensity distribution.

All of the extractors successfully extracted the ECG features, including the peaks, edges, and corners from the PQRST waves. The extracted features were then given to the classifier (SVM) to classify the emotions. Figure 5 illustrates the structure of the proposed 2-D ECG-based.

Support vector machine

Emotion classification was performed using SVM. The SVM works by separating the class data points and drawing a boundary called the hyperplane between them. Each hyperplane has what are known as "decision boundaries" to determine which side of each class resides. As reported in previous studies, SVM has a low computational cost and shows excellent performance in classifying emotions, as reported in previous studies.[13,21,24,43,44]

Experimental setting

The scale of self-assessed emotions, which ranges from 1 (lowest) to 5 (highest), was classified using a five-point scale with middle-point thresholds (an average rate of 3.8). As a result, scales four and five were assigned to the high class, while the remaining scales were assigned to the low class. This results in an imbalanced distribution of DREAMER classes: valence has 39% of high valence and 61% of low valence; arousal has 44% of low arousal and 56% of high arousal.
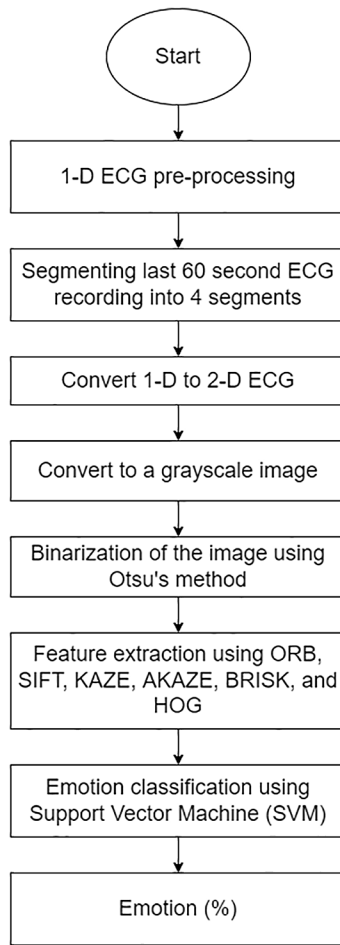
```
                    ┌─────────┐
                    │  Start  │
                    └─────────┘
                         │
                         ▼
        ┌─────────────────────────────────┐
        │      1-D ECG pre-processing      │
        └─────────────────────────────────┘
                         │
                         ▼
        ┌─────────────────────────────────┐
        │  Segmenting last 60 second ECG   │
        │   recording into 4 segments      │
        └─────────────────────────────────┘
                         │
                         ▼
        ┌─────────────────────────────────┐
        │      Convert 1-D to 2-D ECG      │
        └─────────────────────────────────┘
                         │
                         ▼
        ┌─────────────────────────────────┐
        │   Convert to a grayscale image   │
        └─────────────────────────────────┘
                         │
                         ▼
        ┌─────────────────────────────────┐
        │   Binarization of the image using│
        │          Otsu's method           │
        └─────────────────────────────────┘
                         │
                         ▼
        ┌─────────────────────────────────┐
        │   Feature extraction using ORB,  │
        │ SIFT, KAZE, AKAZE, BRISK, and    │
        │              HOG                 │
        └─────────────────────────────────┘
                         │
                         ▼
        ┌─────────────────────────────────┐
        │   Emotion classification using   │
        │  Support Vector Machine (SVM)    │
        └─────────────────────────────────┘
                         │
                         ▼
        ┌─────────────────────────────────┐
        │          Emotion (%)             │
        └─────────────────────────────────┘
```

Figure 5. The overall structure of 1-D ECG-based ERS.

The hyperparameters for SVM were tuned using an exhaustive parameter search tool, GridSearchCV, from Scikit-learn that automates the tuning procedure.[45] This study tuned only the parameters with a high and relative tuning risk and left the remainder at their default values because they are the least sensitive to the hyperparameter tuning process, as suggested by Weerts, Mueller, and Vanschoren.[46]

The dataset was split into a reasonable proportion of training and testing sets to evaluate the model's performance on new unseen data. This study used a stratified train-test split of 80:20 for the training and testing sets. This strategy guarantees the dataset's exact proportions of samples in each class are preserved.

Additionally, as we had a small dataset size, this study applied KFold Cross-Validation, with the number of folds set to 10, the most commonly used number in prior research to improve ERS performance. The experimental setting is tabulated in Table 5.

Table 5. The experimental setting values.

Setting		Value
SVM hyperparameter	Kernel	{linear, rbf}
	C	[0.1,1,10]
	Gamma	[0.1,1,10]
Train-test split		Stratified 70:30
Kfold cross-validation		10

Table 6. Testing emotion classification accuracy and F1-score for 1-D and 2-D electrocardiogram (ECG).

Type of ECG input	Feature extractor	Valence		Arousal	
		Accuracy	F1-Score	Accuracy	F1-Score
1-D	*AUBT and BioSig	62.37	53.05	62.37	57.98
	AUBT	63.86	72.73	*57.83*	*44.44*
	TEAP	*65.06*	*75.63*	54.22	42.42
2-D	ORB	61.75	47.76	56.33	40.59
	KAZE	*62.35*	*49.57*	56.33	40.59
	SIFT	61.14	46.4	56.33	40.59
	AKAZE	61.75	47.76	*59.64*	*59.71*
	BRISK	61.75	47.76	56.02	40.23
	HOG	61.14	46.4	56.33	40.59

Table 7. Computation Time for Each Feature Extractor using Support Vector Machine (SVM).

Type of ECG input	Feature extractor	Computational time (sec)	
		Valence	Arousal
1-D	AUBT	1.65	1.63
	TEAP	1.51	1.55
2-D	ORB	1473.07	1461.77
	KAZE	4486.27	6034.26
	SIFT	239.31	238.55
	AKAZE	2950.23	3308.23
	BRISK	4926.46	3610.64
	HOG	6516.30	6431.28

Results

The testing performance of the ERS in classifying emotions using two different types of ECG data, 1-D and 2-D, is summarised in Table 6. The result denoted by an asterisk (*) corresponds to the original DREAMER publication 13, whereas the best accuracy and F1 score for classifying valence and arousal were bolded and shaded.

For 1-D input, the features extracted using the TEAP feature extractor obtain the best valence performance with an accuracy of 65.06% and an F1 score of 75.63%. The best arousal performance is obtained using features extracted by the AUBT feature extractor, which has a 57.83% and a 44.44% F1 score.

On the other hand, the KAZE feature extractor achieves the best valence performance with 2-D input, achieving 62.35% accuracy and a 49.57% F1 score. Simultaneously, with 59.64% accuracy and a 59.71% F1 score, the AKAZE feature extractor achieves the best performance in arousal emotion.

For comparison purposes, the computation time for both ECG inputs was recorded and reported in Table 7. The average time required to compute 1-D is 1.58 ± 0.07 seconds. In comparison, the average computation time for 2-D is 3377.425 ± 3138.875 seconds. Therefore, according to the observation, 2-D took the longest computation time, whereas 1-D obtained the shortest.

Discussion & conclusions

The results indicate that both inputs work comparably well in classifying emotions. This finding is demonstrated by the fact that the best valence performance was obtained using a 1-D ECG, and the best arousal performance was acquired using a 2-D ECG. Additionally, ERS with 1-D ECG was combined with dimensionality reduction, called LDA. The

presence of LDA improved the ERS performance in valence emotion but not in arousal. In terms of computational cost, 1-D ECG is better to 2-D ECG since it requires less computation time.

However, it is worth mentioning that the results obtained using 2-D ECG demonstrated potential for use as an input modality for the ERS. Additionally, 2-D ECGs are appealing because the format enables the use of a variety of image-based methods such as image augmentation to increase the data size, convolution neural networks (CNN), and the application of transfer learning from models trained using large data. To summarise, the ERS performance of the two ECG inputs is comparable since both yield a promising outcome for emotion recognition.

Data availability
Source data
The DREAMER dataset was first presented here: https://doi.org/10.1109/JBHI.2017.2688239 and can be found on Zenodo. Access is restricted and users are required to apply. The decision whether to grant/deny access is solely under the responsibility of the record owner.

Extended data
Analysis code available from: https://github.com/nr-isml/ECG-Numerical-Vs.-Image-Data-for-Emotion-Recognition-System

Archived analysis code as at time of publication: https://doi.org/10.5281/zenodo.5542739.[47]

License: Data are available under the terms of the Creative Commons Zero "No rights reserved" data waiver (CC0 1.0 Public domain dedication).

Acknowledgements
The authors would like to thank those who were involved in this experiment, either directly or indirectly.

References

1. Tayel MB, El-Bouridy ME: **ECG images classification using artificial neural network based on several feature extraction methods.** *2008 Int. Conf. Comput. Eng. Syst. ICCES 2008.* 2008; pp. 113–115.

2. Mohamed B, Issam A, Mohamed A, *et al.*: **ECG Image Classification in Real time based on the Haar-like Features and Artificial Neural Networks.** *Procedia Comput. Sci.* 2015; **73**(Awict): 32–39. **Publisher Full Text**

3. Yeh LR, *et al.*: **Integrating ECG monitoring and classification via iot and deep neural networks.** *Biosensors.* 2021; **11**(6): 1–12. **Publisher Full Text**

4. Hao P, Gao X, Li Z, *et al.*: **Multi-branch fusion network for Myocardial infarction screening from 12-lead ECG images.** *Comput. Methods Programs Biomed.* 2020; **184**: 105286. **PubMed Abstract** | **Publisher Full Text**

5. Mandal S, Mondal P, Roy AH: **Detection of Ventricular Arrhythmia by using Heart rate variability signal and ECG beat image.** *Biomed. Signal Process. Control.* 2021; **68**(May): 102692. **Publisher Full Text**

6. Du N, *et al.*: **FM-ECG: A fine-grained multi-label framework for ECG image classification.** *Inf. Sci. (Ny).* 2021; **549**: 164–177. **Publisher Full Text**

7. C. U. First Faculty of Medicine: **Electrocardiogram - WikiLectures.** 2018. [Accessed: 04-Oct-2021]. **Reference Source**

8. Soleymani M, Lichtenauer J, Pun T, *et al.*: **A multimodal database for affect recognition and implicit tagging.** *IEEE Trans. Affect. Comput.* 2012; **3**(1): 42–55. **Publisher Full Text**

9. Abadi MK, Subramanian R, Kia SM, *et al.*: **DECAF: MEG-Based Multimodal Database for Decoding Affective Physiological Responses.** *IEEE Trans. Affect. Comput.* 2015; **6**(3): 209–222. **Publisher Full Text**

10. Subramanian R, Wache J, Abadi MK, *et al.*: **ASCERTAIN: Emotion and personality recognition using commercial sensors.** *IEEE Trans. Affect. Comput.* 2018; **9**(2): 147–160. **Publisher Full Text**

11. Siddharth S, Jung T-P, Sejnowski TJ: **Utilizing Deep Learning Towards Multi-modal Bio-sensing and Vision-based Affective Computing.** *IEEE Trans. Affect. Comput.* 2019; pp. 1–1. **Publisher Full Text**

12. Hasnul MA, Ab Aziz NA, Aziz AA: **Evaluation of TEAP and AuBT as ECG's Feature Extraction Toolbox for Emotion Recognition System.** *IEEE 9th Conf. Syst. Process Control* (ICSPC 2021), no. December, 2022; pp. 52–57.

13. 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. Heal. Informatics.* 2018; **22**(1): 98–107. **PubMed Abstract** | **Publisher Full Text**

14. Koelstra S, *et al.*: **DEAP: A database for emotion analysis; Using physiological signals.** *IEEE Trans. Affect. Comput.* 2012; **3**(1): 18–31. **Publisher Full Text**

15. Fangmeng Z, Peijia L, Iwamoto M, *et al.*: **Emotional changes detection for dementia people with spectrograms from physiological signals.** *Int. J. Adv. Comput. Sci. Appl.* 2018; **9**(10): 49–54. **Publisher Full Text**

16. AliveCor: EKG Anywhere, Anytime|AliveCor. 2022. [Online]. [Accessed: 07-Jan-2022]. **Reference Source**

17. EMAY: Wireless EKG Monitor – EMAY. 2022. [Online]. [Accessed: 25-Mar-2022]. **Reference Source**

18. Liu J, *et al.*: **CRT-Net: A Generalized and Scalable Framework for the Computer-Aided Diagnosis of Electrocardiogram Signals.** *arXiv.* 2021: 1–25.

19. Wagner J: *Augsburg biosignal toolbox (aubt).*Univ. Augsbg; 2014.

20. Soleymani M, Villaro-Dixon F, Pun T, *et al.*: **Toolbox for Emotional feAture extraction from Physiological signals (TEAP).** *Front. ICT.* 2017; **4**(FEB): 1–7. **Publisher Full Text**

21. Minhad KN, Ali SHM, Reaz MBI: **Happy-anger emotions classifications from electrocardiogram signal for automobile driving safety and awareness.** *J. Transp. Heal.* 2017; **7**(November): 75–89.
Publisher Full Text

22. Tivatansakul S, Ohkura M: **Emotion Recognition using ECG Signals with Local Pattern Description Methods.** *Int. J. Affect. Eng.* 2015; **15**(2): 51–61.
Publisher Full Text

23. Song T, Zheng W, Lu C, *et al.*: **MPED: A multi-modal physiological emotion database for discrete emotion recognition.** *IEEE Access.* 2019; **7**(October): 12177–12191.
Publisher Full Text

24. He W, Ye Y, Pan T, *et al.*: **Emotion Recognition from ECG Signals Contaminated by Motion Artifacts.** *2021 Int. Conf. Intell. Technol. Embed. Syst. ICITES.* 2021; pp. 125–130.

25. Sarkar P, Etemad A: **Self-supervised ECG Representation Learning for Emotion Recognition.** *IEEE Trans. Affect. Comput.* 2020: 1–1.
Publisher Full Text

26. Hammad M, Zhang S, Wang K: **A novel two-dimensional ECG feature extraction and classification algorithm based on convolution neural network for human authentication.** *Futur. Gener. Comput. Syst.* 2019; **101**: 180–196.
Publisher Full Text

27. Bento N, Belo D, Gamboa H: **ECG Biometrics Using Spectrograms and Deep Neural Networks.** *Int. J. Mach. Learn. Comput.* 2020; **10**(2): 259–264.
Publisher Full Text

28. Kłosowski G, Rymarczyk T, Wójcik D, *et al.*: **The use of time-frequency moments as inputs of lstm network for ECG signal classification.** *Electron.* 2020; **9**(9): 1–22.
Publisher Full Text

29. Ullah A, Anwar SM, Bilal M, *et al.*: **Classification of arrhythmia by using deep learning with 2-D ECG spectral image representation.** *Remote Sens.* 2020; **12**(10): 1–14.
Publisher Full Text

30. Elalamy R, Fanourakis M, Chanel G: **Multi-modal emotion recognition using recurrence plots and transfer learning on physiological signals.** *2021 9th Int. Conf. Affect. Comput. Intell. Interact.* ACII 2021, 2021.

31. Guo X: **Study of emotion recognition based on electrocardiogram and RBF neural network.** *Procedia Eng.* 2011; **15**: 2408–2412.

32. Schmidt P, Reiss A, Duerichen R, *et al.*: **Introducing WeSAD, a multimodal dataset for wearable stress and affect detection.** *ICMI 2018 - Proc. 2018 Int. Conf. Multimodal Interact.* 2018; pp. 400–408.

33. Velliangiri S, Alagumuthukrishnan S, Thankumar Joseph SI: **A Review of Dimensionality Reduction Techniques for Efficient Computation.** *Procedia Comput. Sci.* 2019; **165**: 104–111.
Publisher Full Text

34. Rahman MA, *et al.*: **A statistical designing approach to MATLAB based functions for the ECG signal preprocessing.** *Iran J. Comput. Sci.* 2019; **2**(3): 167–178.
Publisher Full Text

35. Otsu N: **A Threshold Selection Method from Gray-Level Histograms.** *IEEE Trans. Syst. Man Cybern.* 1979; **9**(1): 62–66.
Publisher Full Text

36. Trier OD, Taxt T: **Evaluation of Binarization Methods for Document Images.** *IEEE Trans. Pattern Anal. Mach. Intell.* 1995; **17**(3): 312–315.
Publisher Full Text

37. Rublee E, Rabaud V, Konolige K, *et al.*: **ORB: An efficient alternative to SIFT or SURF.** *Proc. IEEE Int. Conf. Comput. Vis.* 2011 May: pp. 2564–2571.

38. Shi Y, Lv Z, Bi N, *et al.*: **An improved SIFT algorithm for robust emotion recognition under various face poses and illuminations.** *Neural Comput. Appl.* 2020; **32**(13): 9267–9281.
Publisher Full Text

39. Alcantarilla PF, Bartoli A, Davison AJ: **KAZE features.** *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics).* 2012; vol. **7577 LNCS**(no. PART 6): pp. 214–227.
Publisher Full Text

40. Tareen SAK, Saleem Z: **A comparative analysis of SIFT, SURF, KAZE, AKAZE, ORB, and BRISK.** *2018 Int. Conf. Comput. Math. Eng. Technol. Inven. Innov. Integr. Socioecon. Dev. iCoMET 2018 - Proc.* 2018; vol. **2018-Janua**: pp. 1–10.

41. Liu Y, Zhang H, Guo H, *et al.*: **A FAST-BRISK feature detector with depth information.** *Sensors (Switzerland).* 2018; **18**(11).
Publisher Full Text

42. Rathikarani V, Dhanalakshmi P, Vijayakumar K: **Automatic ECG Image Classification Using HOG and RPC Features by Template Matching.** 2016; pp. 117–125.

43. Bulagang AF, Weng NG, Mountstephens J, *et al.*: **A review of recent approaches for emotion classification using electrocardiography and electrodermography signals.** *Informatics Med. Unlocked.* 2020; **20**: 100363.
Publisher Full Text

44. Zhai J, Barreto A: **Stress detection in computer users based on digital signal processing of noninvasive physiological variables.** *Annu. Int. Conf. IEEE Eng. Med. Biol. - Proc.* 2006; (no. May): pp. 1355–1358.

45. Pedregosa F, Varoquaux G, Gramfort A, *et al.*: **Scikit-learn: Machine Learning in Python.** *J. Machine Learn. Res.* 2011; **12**: 2825–2830.

46. Weerts HJP, Mueller AC, Vanschoren J: **Importance of Tuning Hyperparameters of Machine Learning Algorithms.** 2020.

47. nr-isml: **nr-isml/ECG-Numerical-Vs.-Image-Data-for-Emotion-Recognition-System: First release (ECG).** *Zenodo.* 2021.
Publisher Full Text

Open Peer Review

Current Peer Review Status: ? ✔ ✔

Version 2

Reviewer Report 06 June 2022

https://doi.org/10.5256/f1000research.134293.r139365

✔ **Marios Fanourakis**
University of Geneva, Geneva, Switzerland

After evaluating the author's revisions and replies to my previous comments I would like to upgrade the status to Approved.

The authors have carefully addressed the issues previously identified with only some minor clarifications needed.

What method was used to resize/rescale the 2D ECG images (which python function)? I see in the github code that the following method was used "cv2.resize(img, dim, interpolation = cv2.INTER_AREA)", please explain what the interpolation method cv2.INTER_AREA is.

As mentioned the 2D image is 1920x620 pixels, a 60% rescaling doesn't result in either 224x224 nor 299x299. Some information is still missing here. Furthermore, why is 224x224 and 299x299 a "standard size"? Is it that these sizes are common in other ML models? Other reasons?

When splitting the data between the training and test set, was an individual participant's data also split between the train and test sets? For example, suppose there is a participant "A", would one find some data of participant "A" in the training set and some other data of the same participant "A" in the test set, or was it split such that if there was data of participant "A" in the training set there was no data from participant "A" in the test set and vice versa? This could help to further interpret the results.

From a quick glance at the github code, it seems that participants' data could be present in both training and test sets. This is not inherently bad and I don't believe it invalidates the results and conclusions presented, especially since the work's main contribution is to compare the performance of 1D ECG signals to 2D ECG images as inputs to ML models. Nonetheless, it is important to mention.

A short definition of "stratified train-test split" could be helpful for readers.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Affective computing, emotion recognition

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 31 May 2022

https://doi.org/10.5256/f1000research.134293.r139366

✔ **Md. Asadur Rahman** 🆔
Department of Biomedical Engineering, Military Institute of Science and Technology (MIST), Dhaka, Bangladesh

The revised version can be approved. I am satisfied so far.

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

- -

Version 1

Reviewer Report 16 March 2022

https://doi.org/10.5256/f1000research.76896.r121059

? **Md. Asadur Rahman** 🆔
Department of Biomedical Engineering, Military Institute of Science and Technology (MIST), Dhaka, Bangladesh

The paper titled "Evaluation of electrocardiogram: numerical vs. image data for emotion recognition system" looks interesting. Although it is a short paper, I found it interesting and am positive about this work, but it should go through some modification:
- ○ Figure 2 should be redrawn, only providing the reference is not enough to present a figure

or image in the article.

○ The ECG signals used in this work contain some baseline wandering. It should be removed before further analysis. A simple technique with Matlab code is described in Rahman *et al* (2019)[1] to remove baseline wander. I am expecting a result comparing the emotion recognition rate before and after the baseline wander removal from the ECG signal.

○ In addition to that, is it possible to use LSTM to the ECG time series to find the accuracy of the emotion recognition from ECG signals?

○ Please provide detailed information about transforming the signal to image conversion.

○ The reference-related comparison and discussion should be completed in the result and discussions. If possible, avoid referencing in the Conclusion section.

References
1. Rahman M, Milu M, Anjum A, Siddik A, et al.: A statistical designing approach to MATLAB based functions for the ECG signal preprocessing. *Iran Journal of Computer Science*. 2019; **2** (3): 167-178
Publisher Full Text

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Biomedical Signal Processing

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 20 May 2022

Sharifah Noor Masidayu Sayed Ismail, Multimedia University, Bukit Beruang,, Malaysia

The paper titled "Evaluation of electrocardiogram: numerical vs. image data for emotion recognition system" looks interesting. Although it is a short paper, I found it interesting and am positive about this work, but it should go through some modification:

We appreciate your feedback and suggestions. We have revised the manuscript as needed.

Figure 2 should be redrawn, only providing the reference is not enough to present a figure or image in the article.

Thank you for your suggestions. As per your recommendations, we have redrawn the figure and an additional note has been added in the legend. As several modifications were made to the manuscript, Figure 2 was originally changed to Figure 1. This figure can be found on page 4.

The ECG signals used in this work contain some baseline wandering. It should be removed before further analysis. A simple technique with MATLAB code is described in Rahman et al (2019)1 to remove baseline wander. I am expecting a result comparing the emotion recognition before and after the baseline wander removal from the ECG signal.

Thank you for your suggestion. We did use the method suggested as cited accordingly. However, your expectation of having a result comparing the emotion recognition before and after the baseline wander removal from the ECG signal is inappropriate for our work because the focus of this paper is to compare emotion classification performance using 1-D and 2-D ECGs to investigate the effect of the ECG input format on the ERS.

In addition to that, is it possible to use LSTM to the ECG time series to find the accuracy of the emotion recognition from ECG signals?

As far as we know, it is possible to use LSTM on the ECG time series to find the accuracy of emotion recognition from ECG signals. The work by Song et al. used LSTM for the same purpose as your concern. Below is the paper, which you can take a look at. We hope it answers your question.

T. Song, W. Zheng, C. Lu, Y. Zong, X. Zhang, and Z. Cui, "MPED: A multi-modal physiological emotion database for discrete emotion recognition," IEEE Access, vol. 7, no. October, pp. 12177–12191, 2019.

Please provide detailed information about transforming the signal to image

conversion.

Thank you for pointing this out. We have revised the 2-D ECG subsection, where the pre-processing process until the transformation from 1-D to 2-D is explained in detail. This subsection can be found on pages 11 and 12, which can be read as follows:

"The duration of the ECG recording varies according to the duration of the video (average = 199 seconds). As Katsigiannis and Ramzan proposed, this study analysed the final 60 seconds of each recording to allow time for a dominant emotion to emerge [13]. Following that, 1-D ECG was pre-processed using a simple MATLAB function by [34] to eliminate baseline wander caused by breathing, electrically charged electrodes, or muscle noise. The signal was then divided into four segments corresponding to 15 seconds each. Then, using MATLAB version 9.7, the 1-D ECG was transformed into a 2-D ECG (Figure 4). The image has a width of 1920 pixels and a height of 620 pixels.

Due to the fact that the 2-D ECG was converted to a rectangle shape, it is not easy to resize the photos to the standard input image sizes of 224×224 and 299×299. As a result, the converted 2-D ECG was resized to 60% of its original size using Python version 3.8.5. This scale percentage was chosen after considering the quality of the image, the type of feature extractor used, and the computational cost the system can afford. The coloured images were converted into greyscale images. Then, binarization of the image using an Otsu's automatic image thresholding method [35] was done. This method ascertains the optimal threshold values from pixel values of 0 to 255 by calculating and evaluating their within-class variance [36]"

The reference-related comparison and discussion should be completed in the result and discussions. If possible, avoid referencing in the Conclusion section.

Thank you for bringing this to our attention. We found your comments extremely helpful and have revised them accordingly. The revised text can be read as follows:

"The results indicate that both inputs work comparably well in classifying emotions. This finding is demonstrated by the fact that the best valence performance was obtained using a 1-D ECG, and the best arousal performance was acquired using a 2-D ECG. Additionally, ERS with 1-D ECG was combined with dimensionality reduction, called LDA. The presence of LDA improved the ERS performance in valence emotion but not in arousal. In terms of computational cost, 1-D ECG is better to 2-D ECG since it requires less computation time.

However, it is worth mentioning that the results obtained using 2-D ECG demonstrated potential for use as an input modality for the ERS. Additionally, 2-D ECGs are appealing because the format enables the use of a variety of image-based methods such as image augmentation to increase the data size, convolution neural networks (CNN), and the application of transfer learning from models trained using large data. To summarise, the ERS performance of the two ECG inputs is comparable since both yield a promising outcome for emotion recognition."

Reviewer Report 16 March 2022

https://doi.org/10.5256/f1000research.76896.r126091

❌ **Marios Fanourakis**
University of Geneva, Geneva, Switzerland

The authors use the DREAMER dataset to compare the emotion recognition performance of features extracted from the time-series ECG signal versus features extracted from images of the ECG wave signal. An SVM model is used as the classifier.

Overall, the structure of the report is not coherent, the related work is incomplete, and the motivation for this work is not convincing. Furthermore, many important details are missing about the dataset and the methods used which makes it difficult to trust the results and the comparisons they make with other works.

Detailed comments:

Structure: several improvements to be made, information seems to be scattered throughout the article. For example, information about the dataset is present in both the introduction and methods sections. It is best to keep this information in the same section. Same for the related works.

Authors should give a brief explanation on what are ECG wave images in the introduction, otherwise readers in the emotion recognition field might confuse them with spectrograms which are more widely used in the field. It may also be better to change the term "numerical data" to "time-series data" or "1D data" (2D being an image).

Typo on page 3: EGC instead of ECG.

From the references pertaining to the use of ECG images (1-6, 23,24) only half actually use ECG wave images (2,4,23,24). The rest either use time-series or convert the time-series to spectrograms. The ones that do use ECG images, mainly analyze individual beats and not the entire ECG wave in order to detect medical heart issues. For the emotion recognition use-case it is necessary to analyze significantly larger regions of the signal than individual beats.

In emotion recognition, it is common to transform the ECG signal into a spectrogram image. The authors do not cite any work mentioning this method.

Authors do not include any references to other works that use the DREAMER dataset ECG signals for emotion recognition, here are some:

Wenwen He et al. 2021 Emotion Recognition from ECG Signals Contaminated by Motion Artifacts[1]
Pritam Sarkar et al. 2020 Self-supervised ECG Representation Learning for Emotion Recognition[2]

I also came across another publication of some of the co-authors which would be advantageous to reference:

Muhammad Anas Hasnul et al. 2021 Evaluation of TEAP and AuBT as ECG's Feature Extraction Toolbox for Emotion Recognition System[3]

In the Emotion model paragraph (page 3): it is still unclear which model of emotions was used, Ekman or Russell? Authors say "the latter" (referring to Russel) but then mention binary classification which may be confusing since the arousal/valence space is continuous making it a regression problem or at least multiclass and not binary. Authors should re-word this part to make it clear.

Referring to wearable ECG devices, the authors state: "However, most of these devices store the ECG as images instead of raw numerical data". No references or other market analysis is provided to show that this is the case. The use of ECG wave images for emotion recognition is not properly motivated. I fail to see any advantage of using ECG wave images over the time-series data unless ECG time-series data is not available.

How was the data annotated in the DREAMER dataset? Were they continuous annotations or a single annotation per video clip? This is very relevant to include in the article. A few more words about the dataset are needed like a short description of the experimental protocol.

Arousal/valence rating values are ranging from 1 to 5 in the DREAMER dataset. The authors never explain how they split them into two classes (high/low) for the binary classification.

Authors do not include sufficient information on how the time-series ECG data was converted to an image (resolution, compression, windowing), or how the data was treated in general. Did the authors use the entire ECG signal for each video? Was there any windowing? Since the authors did not properly summarize the dataset (how were the videos annotated?) it is difficult to grasp or guess on how the data was processed.

None of the cited literature used any of the image feature extraction methods that the authors used, and the authors did not discuss their reasoning for why they selected those image feature extraction methods and not the ones established in the ECG image analysis literature that they cited. Some more illustrations of these features may be useful besides the one in Figure 5 in order to convince the readers.

Support vector machine: It is not clear what preprocessing steps were applied to the data. For example in the DREAMER dataset baseline they only use the last 60s of data for each film clip.

Data was divided 80:20 and 10-fold cross-validation was used. The authors do not specify exactly

how the data was split (see DREAMER dataset paper section V as an example). From reading this part, I can only assume that all data from all participants was used for each fold (general model), something that is diverging from how the data was split in the DREAMER dataset paper (they made models for each individual participant). Therefore any comparisons to the results of the DREAMER baseline are invalid.

Results: If the classes are unbalanced (as the DREAMER dataset paper indicates) accuracy is not valid on it's own, include f1 score, and/or Cohen's Kappa.

Discussion: "Ref. 11" has nothing to do with the statement in that paragraph. LDA was actually applied in the DREAMER dataset paper and they reported that there were no significant differences in performance.

References
1. He W, Ye Y, Pan T, Meng Q, et al.: Emotion Recognition from ECG Signals Contaminated by Motion Artifacts,. *International Conference on Intelligent Technology and Embedded Systems (ICITES), 2021*. 2021. 125-130 Publisher Full Text
2. Sarkar P, Etemad A: Self-supervised ECG Representation Learning for Emotion Recognition. *IEEE Transactions on Affective Computing*. 2021. Publisher Full Text
3. Hasnul MA, Ab Aziz NA, Aziz AA: Evaluation of TEAP and AuBT as ECG's Feature Extraction Toolbox for Emotion Recognition System. 2021 IEEE 9th Conference on Systems, Process and Control (ICSPC 2021), 2021. 52-57 Publisher Full Text

Is the work clearly and accurately presented and does it cite the current literature?
No

Is the study design appropriate and is the work technically sound?
Partly

Are sufficient details of methods and analysis provided to allow replication by others?
No

If applicable, is the statistical analysis and its interpretation appropriate?
Not applicable

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Affective computing, emotion recognition

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for

reasons outlined above.

Author Response 20 May 2022

Sharifah Noor Masidayu Sayed Ismail, Multimedia University, Bukit Beruang,, Malaysia

The authors use the DREAMER dataset to compare the emotion recognition performance of features extracted from the time-series ECG signal versus features extracted from images of the ECG wave signal. An SVM model is used as the classifier.

Overall, the structure of the report is not coherent, the related work is incomplete, and the motivation for this work is not convincing. Furthermore, many important details are missing about the dataset and the methods used which makes it difficult to trust the results and the comparisons they make with other works.

Thanks for taking the time to review our manuscript. We have revised the manuscript based on your comments and suggestions accordingly.

Structure: several improvements to be made, information seems to be scattered throughout the article. For example, information about the dataset is present in both the introduction and methods sections. It is best to keep this information in the same section. Same for the related works.

Thanks for pointing this out. We agree with your suggestion and have attempted to address the issues. Therefore, we have revised each section as per suggestions by revising the information written in the manuscript accordingly. The revised part can be found on pages 3 until 13.

Authors should give a brief explanation on what are ECG wave images in the introduction, otherwise readers in the emotion recognition field might confuse them with spectrograms which are more widely used in the field. It may also be better to change the term "numerical data" to "time-series data" or "1D data" (2D being an image).

Thank you for drawing our attention to this. As you mentioned, we have added an explanation of what ECG wave images are in the Introduction section that reads as follows:

"Fundamentally, ECG is used to measure electrical activity in the human heart by attaching electrodes to the human body. Due to the continual blood pumping action to the body, the electrical activity of the heart can be found in the sinoatrial node. The electrocardiogram signal is composed of three basic components: P, QRS, and T waves (Figure 1). P waves are produced during atrium depolarization, QRS complexes are produced during ventricular depolarization, and T waves are produced during ventricle recovery.

Despite this, majority of the portable devices record the ECG signal as images (2-D images) in a PDF file rather than as raw numerical data (1-D data) [16 – 18]. The example of a PDF-

based 2-D ECG is depicted in Figure 2. Due to this problem, researchers were required to convert the PDF file of the ECG into 1-D data before performing further emotion analysis, adding complexity to the pre-processing process. On this account, given the positive results obtained in monitoring and diagnosing cardiovascular-related diseases, the efficacy of 2-D ECG in emotion studies also warrants further investigation."

Additionally, we also added Figure 2 that shows the snippet of the 2-D ECG from the PDF file. This figure can be found on page 4.

Furthermore, as per suggestion, we have changed the terms "numerical data" to "1D ECG" and "wave images" to "2-D ECG".

Typo on page 3: EGC instead of ECG.

Thank you so much for catching these glaring and confusing errors, which we have now corrected.

From the references pertaining to the use of ECG images (1-6, 23,24) only half actually use ECG wave images (2,4,23,24). The rest either use time-series or convert the time-series to spectrograms. The ones that do use ECG images, mainly analyze individual beats and not the entire ECG wave in order to detect medical heart issues. For the emotion recognition use-case it is necessary to analyze significantly larger regions of the signal than individual beats.

Based on your comment, we have revised our Related Works section and corrected them accordingly. Additionally, we added the summary of existing works that use 1-D and 2-D ECG input with their purposes, tabulated in Table 1. The revision of these issues can be found on page 4 until 8.

In emotion recognition, it is common to transform the ECG signal into a spectrogram image. The authors do not cite any work mentioning this method.

Thanks for pointing this out. We agree with your comments. Therefore, we have revised our Related Works section and provided improvements through the fourth paragraph on page 6:
.
"Despite rising popularity among medical practitioners in assessing patients' cardiac disease, 2-D ECG remains inadequate compared to 1-D ECG usage as a type of input in emotion recognition studies. As a result, the number of studies employing 1-D ECG in ERS is higher than that utilizing 2-D ECG in ERS. However, rather than employing a printout-based 2-D ECG, emotion researchers classified human emotions using 2-D ECG spectral images. For example, [15] determines the R-peaks of the electrocardiogram prior to generating the R-R interval (RRI) spectrogram. Following that, CNN was used to classify the emotions, with an accuracy rate greater than 90%. Elalamy et al. [30] used ResNet-50 to extract features from a

2-D ECG spectrogram. Then, Logistic Regression (LR) was employed as a classifier and achieved an accuracy of 78.30% in classifying emotions."

Additionally, in Table 1, the ECG input was listed as either 1-D or 2-D ECGs, where 2-D was further categorised into standard 2-D ECG or spectral 2-D ECG.

Authors do not include any references to other works that use the DREAMER dataset ECG signals for emotion recognition, here are some:

Wenwen He et al. 2021 Emotion Recognition from ECG Signals Contaminated by Motion Artifacts

Pritam Sarkar et al. 2020 Self-supervised ECG Representation Learning for Emotion Recognition

I also came across another publication of some of the co-authors which would be advantageous to reference:

Muhammad Anas Hasnul et al. 2021 Evaluation of TEAP and AuBT as ECG's Feature Extraction Toolbox for Emotion Recognition System.

Thank you for the paper suggested. We have added the reference according to your recommendation, which you can find on page 6 as follows:

"Additionally, numerous other researchers also used the ECG signals from the DREAMER dataset to perform emotion recognition. For instance, 1-D ECG data from the DREAMER dataset is utilized by Wenwen He et al. [24] that suggested an approach for emotion recognition using ECG contaminated by motion artefacts. The proposed approach improved classification accuracy by 5% to 15%. Additionally, Pritam and Ali [25] also employed 1-D ECG from the DREAMER dataset to develop the self-supervised deep multi-task learning framework ERS, which consists of two stages of learning: ECG representation learning and emotion classification learning. The accuracy gained in this study was greater than 70%. Hasnul et al. [12] also used the 1-D ECG by DREAMER dataset to compare the performance of two feature extractor toolboxes. They noted that the dataset's size and the type of emotion classified might affect the suitability of the extracted features."

In the Emotion model paragraph (page 3): it is still unclear which model of emotions was used, Ekman or Russell? Authors say "the latter" (referring to Russel) but then mention binary classification which may be confusing since the arousal/valence space is continuous making it a regression problem or at least multiclass and not binary. Authors should re-word this part to make it clear.

Thank you for pointing this out. The reviewer is correct; the original phrase is confusing. Therefore, we have removed this part to avoid any further confusion. Furthermore, this concern has been changed and revised to make it more straightforward and

understandable. The revised text was located in the experimental setting subsection under the Method section on page 13:

"The scale of self-assessed emotions, which ranges from 1 (lowest) to 5 (highest), was classified using a five-point scale with middle-point thresholds (an average rate of 3.8). As a result, scales four and five were assigned to the high class, while the remaining scales were assigned to the low class."

Referring to wearable ECG devices, the authors state: "However, most of these devices store the ECG as images instead of raw numerical data". No references or other market analysis is provided to show that this is the case. The use of ECG wave images for emotion recognition is not properly motivated. I fail to see any advantage of using ECG wave images over the time-series data unless ECG time-series data is not available.

Thank you for drawing our attention to this, and we agree with the comments. Therefore, we have cited the necessary references and revised the statement as per the suggestion. The revised text can be found on page 4 as follows:

"Previous research on human emotions has primarily relied on either direct analysis of 1-D data [12–14] or the conversion of 1-D data to a 2-D spectral image [15] prior to identifying the emotions. Despite this, majority of the portable devices record the ECG signal as images (2-D images) in a PDF file rather than as raw numerical data (1-D data) [16–18]. The example of a PDF-based 2-D ECG is depicted in Figure 2. Due to this problem, researchers were required to convert the PDF file of the ECG into 1-D data before performing further emotion analysis, adding complexity to the pre-processing process. On this account, given the positive results obtained in monitoring and diagnosing cardiovascular-related diseases, the efficacy of 2-D ECG in emotion studies also warrants further investigation."

How was the data annotated in the DREAMER dataset? Were they continuous annotations or a single annotation per video clip? This is very relevant to include in the article. A few more words about the dataset are needed like a short description of the experimental protocol.

Thank you for bringing this issue to our attention. The short description of the DREAMER dataset has been improved to address this issue. The revised text was located in the Method section under a subsection called "The dataset (DREAMER")" on page 8:

"This study used ECG signals from Katsigiannis and Ramzan [13] called DREAMER. The DREAMER dataset is a freely accessible database of electroencephalogram (EEG) and electrocardiogram (ECG) signals used in emotion research. However, EEG signals were removed from this study because the primary focus is on ECG signals. The ECG was recorded using the SHIMMER ECG sensor at 256 Hz and stored in 1-D format. The DREAMER dataset contains 414 ECG recordings from 23 subjects who were exposed to 18 audio-visual stimuli designed to evoke emotion. Each participant assessed their emotions on a scale of 1

to 5 for arousal, valence, and dominance. However, because this study was primarily concerned with arousal and valence ratings, participants' evaluations of dominance were discarded."

Arousal/valence rating values are ranging from 1 to 5 in the DREAMER dataset. The authors never explain how they split them into two classes (high/low) for the binary classification.

Thank you for bringing this to our attention. We have improved this part by adding an explanation of how we categorise emotions into two classes. The explanation can be found on page 13:

"The scale of self-assessed emotions, which ranges from 1 (lowest) to 5 (highest), was classified using a five-point scale with middle-point thresholds (an average rate of 3.8). As a result, scales four and five were assigned to the high class, while the remaining scales were assigned to the low class."

Authors do not include sufficient information on how the time-series ECG data was converted to an image (resolution, compression, windowing), or how the data was treated in general. Did the authors use the entire ECG signal for each video? Was there any windowing? Since the authors did not properly summarize the dataset (how were the videos annotated?) it is difficult to grasp or guess on how the data was processed.

We agree with this comment. Therefore, we improved the explanation part of pre-processing the 2-D ECG. The improvised text can be found in the first and second paragraphs of subsection 2-D ECG on page 11.

"The duration of the ECG recording varies according to the duration of the video (average = 199 seconds). As Katsigiannis and Ramzan proposed, this study analysed the final 60 seconds of each recording to allow time for a dominant emotion to emerge [13]. Following that, 1-D ECG was pre-processed using a simple MATLAB function by [34] to eliminate baseline wander caused by breathing, electrically charged electrodes, or muscle noise. The signal was then divided into four segments corresponding to 15 seconds each. Then, using MATLAB version 9.7, the 1-D ECG was transformed into a 2-D ECG (Figure 4). The image has a width of 1920 pixels and a height of 620 pixels.

Due to the fact that the 2-D ECG was converted to a rectangle shape, it is not easy to resize the photos to the standard input image sizes of 224×224 and 299×299. As a result, the converted 2-D ECG was resized to 60% of its original size using Python version 3.8.5. This scale percentage was chosen after considering the quality of the image, the type of feature extractor used, and the computational cost the system can afford. The coloured images were converted into greyscale images. Then, binarization of the image using an Otsu's automatic image thresholding method [35] was done. This method ascertains the optimal threshold values from pixel values of 0 to 255 by calculating and evaluating their within-class variance [36]"

None of the cited literature used any of the image feature extraction methods that the authors used, and the authors did not discuss their reasoning for why they selected those image feature extraction methods and not the ones established in the ECG image analysis literature that they cited. Some more illustrations of these features may be useful besides the one in Figure 5 in order to convince the readers.

Thank you for pointing this out. Cite literature either using their own algorithm to detect peaks on PQRST waves or automatically extracting ECG features using a deep learning system. However, we had included our reason for employing these image feature extraction methods, which we believed could add a valuable contribution to the state-of-the-art. Additionally, we removed Figure 5 and replaced it with the description of each feature extractor to help convince the readers. The revised text can be found on pages 11 and 12:

"The area of interest for 2-D ECG is laying on the PQRST waves, making the peaks detector the best approach to be employed. Therefore, six different feature extractors that could extract peaks, edges, or corners were applied to extract features from 2-D ECGs using Python version 3.8.5:
1. ORB [37]: ORB features are invariant to rotation and noise because they are a combination of Features from Accelerated Segment Test (FAST) detection and Binary Robust Independent Elementary Features (BRIEF) description methods.
2. SIFT [38]: SIFT identifies feature points by searching for local maxima on the images using Difference-of-Gaussians (DoG) operators. The description approach generates a 16x16 neighbourhood around each identified feature and sub-blocks the region. SIFT is also rotation and scale invariant.
3. KAZE [39]: KAZE is based on the scale of the normalised determinant of the Hessian Matrix, with the maxima of detector responses being captured as feature points using a moving window. Additionally, KAZE makes use of non-linear space via non-linear diffusion filtering to reduce noise while keeping the borders of regions in images.
4. AKAZE [40]: AKAZE is a more sophisticated version of KAZE that is based on the Hessian Matrix determinant. Scharr filters were employed to enhance the quality of the invariance rotation, rendering AKAZE features rotation- and scale-invariant.
5. BRISK [41]: While searching for maxima in the scale-space pyramid, BRISK detects corners using the AGAST algorithm and filters them using the FAST Corner Score. Additionally, the BRISK description is based on the recognised characteristic direction of each feature, which is necessary for rotation invariance.
6. HOG [42]: HOG is a feature descriptor that is used to compute the gradient value for each pixel. The image shape denoted the edge or gradient structure derived using a high-quality local gradient intensity distribution."

Support vector machine: It is not clear what preprocessing steps were applied to the data. For example, in the DREAMER dataset baseline they only use the last 60s of data for each film clip.

Thank you for this comment. The pre-processing part can be found in the Methods section

under the 1-D ECG and 2-D ECG subsections. The pre-processing for 1-D ECG can be read as follows:

"The AUBT and TEAP feature extractors were included with the Low Pass Filter (LPF), a filter meant to reject all undesirable frequencies in a signal. The LPF was one of the most widely used filters before the computation of statistical features for physiological signals [31, 32]. As a result, automated 1-D ECG pre-processing utilizing LPF was performed in this study to reduce muscle and respiratory noise in ECG signals."

Whereas the pre-processing for 2-D ECG can be read as follows:

"The duration of the ECG recording varies according to the duration of the video (average = 199 seconds). As Katsigiannis and Ramzan proposed, this study analysed the final 60 seconds of each recording to allow time for a dominant emotion to emerge [13]. Following that, 1-D ECG was pre-processed using a simple MATLAB function by [34] to eliminate baseline wander caused by breathing, electrically charged electrodes, or muscle noise. The signal was then divided into four segments corresponding to 15 seconds each. Then, using MATLAB version 9.7, the 1-D ECG was transformed into a 2-D ECG."

Data was divided 80:20 and 10-fold cross-validation was used. The authors do not specify exactly how the data was split (see DREAMER dataset paper section V as an example). From reading this part, I can only assume that all data from all participants was used for each fold (general model), something that is diverging from how the data was split in the DREAMER dataset paper (they made models for each individual participant). Therefore, any comparisons to the results of the DREAMER baseline are invalid.

Thank you for pointing this out. We have addressed this issue by adding a new subsection under Method, namely, Experimental Setting. This subsection is located on pages 13 and 14, and can be read as follows:

"The hyperparameters for SVM were tuned using an exhaustive parameter search tool, GridSearchCV, from Scikit-learn that automates the tuning procedure [45]. This study tuned only the parameters with a high and relative tuning risk and left the remainder at their default values because they are the least sensitive to the hyperparameter tuning process, as suggested by Weerts, Mueller, and Vanschoren [46].

The dataset was split into a reasonable proportion of training and testing sets to evaluate the model's performance on new unseen data. This study used a stratified train-test split of 80:20 for the training and testing sets. This strategy guarantees the dataset's exact proportions of samples in each class are preserved.

Additionally, as we had a small dataset size, this study applied KFold Cross-Validation, with the number of folds set to 10, the most commonly used number in prior research to improve ERS performance."

Additionally, the summary of the experimental setting values has been tabulated in Table 5 on page 14.

Results: If the classes are unbalanced (as the DREAMER dataset paper indicates) accuracy is not valid on its own, include f1 score, and/or Cohen's Kappa..

Thank you for your suggestions. As per your recommendations, we have included the F1-score along with the accuracy to address the unbalanced class distribution issue. This result can be found in Table 6 on page 14.

Discussion: "Ref. 11" has nothing to do with the statement in that paragraph. LDA was actually applied in the DREAMER dataset paper and they reported that there were no significant differences in performance.

Thank you for your great observations. We have revised that part as needed. Regarding the LDA, we agree with you. However, as far as we know, LDA was used for classification purposes in the DREAMER dataset paper but not for dimension reduction.

Competing Interests: The authors declare that they have no conflict of interest.

Reviewer Report 21 February 2022

https://doi.org/10.5256/f1000research.76896.r121060

? **Umesh Chandra Pati** (iD)

Department of Electronics and Communication Engineering, National Institute of Technology, Rourkela, Odisha, India

The paper presents a comparison of the Emotion Recognition System (ERS) performance using ECG image and ECG numerical data. DREAMER dataset containing 23 ECG recordings has been used for experimentation. ECG numerical data has been converted to ECG images for comparison. The proposed method consists of feature extraction, dimension reduction, and emotion classification. Arousal and Valence accuracy have been compared using different feature extractors on ECG image and numerical data. ORB, SIFT, KAZE, AKAZE, BRISK, and HOG have been used to extract features from ECG image data whereas TEAP and AUBT have been used to extract features from ECG numerical data. It has been concluded that ECG numerical data provides better performance of emotion classification in comparison to ECG image data.

The authors have done a good job in addressing a typical issue. However, there are many aspects

that need attention. Hence, the following suggestions should be addressed to enhance the quality of the manuscript.

1. Literature survey is poor. More number of related state-of-the-art works should be cited. There is no citation of the work taking into account both ECG image and ECG numerical data.

2. It has been mentioned that the converted ECG images have been resized to 60% of the original size to reduce the computational time. What is the reason for choosing 60%?

3. There is no analysis of the computational complexity of the proposed method.

4. In Table 4, emotion classification accuracies for both ECG image and ECG numerical data have been provided. Can these accuracy values be accepted in practical applications?

5. There is no comparison of the proposed method with state-of-the-art methods.

6. References should be complete in all respect and in a uniform style. A few exceptions are 1, 32, 39, 40.

Is the work clearly and accurately presented and does it cite the current literature?
Partly

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Computer Vision, Image/Video Processing, IoT, Artificial Intelligence.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 20 May 2022

Sharifah Noor Masidayu Sayed Ismail, Multimedia University, Bukit Beruang,, Malaysia

The paper presents a comparison of the Emotion Recognition System (ERS) performance using ECG image and ECG numerical data. DREAMER dataset containing 23 ECG recordings has been used for experimentation. ECG numerical data has been converted to ECG images for comparison. The proposed method consists of feature extraction, dimension reduction, and emotion classification. Arousal and Valence accuracy have been compared using different feature extractors on ECG image and numerical data. ORB, SIFT, KAZE, AKAZE, BRISK, and HOG have been used to extract features from ECG image data whereas TEAP and AUBT have been used to extract features from ECG numerical data. It has been concluded that ECG numerical data provides better performance of emotion classification in comparison to ECG image data.

The authors have done a good job in addressing a typical issue. However, there are many aspects that need attention. Hence, the following suggestions should be addressed to enhance the quality of the manuscript.

Thank you for taking the time to review our manuscript. We reviewed the comments and made improvements based on the comments provided accordingly.

Literature survey is poor. More number of related state-of-the-art works should be cited. There is no citation of the work taking into account both ECG image and ECG numerical data.

We have revised Section Related Works as per the suggestion by adding more related state-of-the-art work citations. The revised part is on pages 5 until 8. We also include the summary information about the existing works that employed 1-D and 2-D ECG input, which is tabulated in Table 2 on pages 6 through 8. Furthermore, the work that considers both ECG images and ECG numerical data is also included in this section. This work can be found on page 6, which reads as follows:

"Additionally, Mandal et al. [5] published a study comparing 1-D and 2-D ECGs for the diagnosis of Ventricular Arrhythmia. They concluded that both ECG inputs are effective at detecting the disease."

It has been mentioned that the converted ECG images have been resized to 60% of the original size to reduce the computational time. What is the reason for choosing 60%?

Thank you for drawing our attention to this. We have improved the explanation of this matter in the manuscript. You can find the text on page 11 that reads as follows:

"Due to the fact that the 2-D ECG was converted to a rectangle shape, it is not easy to resize the photos to the standard input image sizes of 224×224 and 299×299. As a result, the converted 2-D ECG was resized to 60% of its original size using Python version 3.8.5. This

scale percentage was chosen after considering the quality of the image, the type of feature extractor used, and the computational cost the system can afford."

There is no analysis of the computational complexity of the proposed method.

Thank you for pointing out this matter. We have included the analysis of the computational complexity under the result section, tabulated in Table 7, which discussed the time analysis of both inputs used to train and test the ERS model. The analysis was located on page 14:

"For comparison purposes, the computation time for both ECG inputs was recorded and reported in Table 7. The average time required to compute 1-D is 1.58 ± 0.07 seconds. In comparison, the average computation time for 2-D is 3377.425 ± 3138.875 seconds. Therefore, according to the observation, 2-D took the longest computation time, whereas 1-D obtained the shortest."

Additionally, the analysis of the computational time was touched on in the Discussion and Conclusion section as follows:

"In terms of computational cost, 1-D ECG is better to 2-D ECG since it requires less computation time."

In Table 4, emotion classification accuracies for both ECG image and ECG numerical data have been provided. Can these accuracy values be accepted in practical applications?

Thank you for pointing this out. For your information, our results have been updated according to the result of the latest experiment. This result is tabulated in Table 6 on page 13. Based on this result, the accuracy and F1-score achieved for this study are on par with the existing works, which shows that our ERS model can be accepted in practical applications. However, this does not rule out the possibility of improving this result, as there is much more room to improve the ERS performance to develop a more robust ERS in the future.

There is no comparison of the proposed method with state-of-the-art methods.

A comparison with existing work (DREAMER) has been provided for 1-D ECG. This comparison can be found in the result section on page 15, tabulated in Table 6. However, no existing work using the 2-D ECG of the DREAMER dataset has been reported. Therefore, no comparison can be made.

References should be complete in all respect and in a uniform style. A few exceptions are 1, 32, 39, 40.

Thank you for bringing our attention to this. We have revised the references and citation part accordingly and completed it in all respects and in a uniform style as per suggestions.

Competing Interests: The authors declare that they have no conflict of interest.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias

- You can publish traditional articles, null/negative results, case reports, data notes and more

- The peer review process is transparent and collaborative

- Your article is indexed in PubMed after passing peer review

- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com   F1000 Research