End-to-End Emotion Recognition using Peripheral Physiological Signals

M. Pidgeon
Software Research institute
Technological University of the Shannon
Athlone, Ireland
mary.pidgeon@tus.ie

N. Kanwal
School of Computing and Mathematics
Keele University
Newcastle, UK
n.kanwal@keele.ac.uk

N. Murray
Department of Electronics and Informatics
Technological University of the Shannon
Athlone, Ireland
niall.murray@tus.ie

M. Asghar
Insight Centre for Data Analytics
National University of Ireland, Galway
Galway, Ireland
mamoona.asghar@nuigalway.ie

The study of emotion recognition from physiological signals has seen a huge growth in recent decades. Studies initially used traditional machine learning classification to estimate either discrete emotions or combinations of arousal and valence. However, different feature engineering techniques such as Ensemble Empirical Mode Decomposition (EEMD) Analysis for electroencephalography (EEG) signals and statistical calculations for peripheral signals are used prior to machine learning. Also, several sensors needed to be worn by participants to predict the emotions. This study aims to investigate whether arousal and valence can be predicted from a single peripheral signal using deep learning. The Galvanic Skin Response (GSR), Respiration (RSP), Blood Volume Pulse (BVP) and Temperature (Temp) signals from the DEAP dataset are used. The signals are downsampled to approximately three hertz (Hz) and input to a convolutional network (CNN) to predict arousal and valence. GSR, RSP and BVP had similar F1 and accuracy results. BVP had an F1 result of 0.673 and 0.632 and accuracies of 63.5% and 61.1% respectively for arousal and valence. RSP’s F1 results were 0.677 and 0.669 and accuracies were 61.3% and 64.2% for arousal and valence respectively. GSR had F1 results of 0.699 and 0.663 and accuracies of 62.5% and 60.2% respectively for arousal and valence. Using raw signals and examining the peripheral signals individually, we were able to identify which sensors showed the best potential for further research to bring emotion classification into a real-world scenario using non-invasive sensors.

1. INTRODUCTION

With the advancement of artificial intelligence (AI) machine learning and deep learning techniques, the study of emotion recognition from physiological signals has seen a huge growth in recent decades. Emotion recognition research involves the selection of participants and the triggering of different emotions using stimuli such as video clips, music, or images whereby the participant experiences the stimulus. The participants have sensors attached to them that collect data that captures the participant’s response to the emotion’s stimulus. Basic inputs such as ElectroCardioGraphy (ECG), ElectroEncephaloGraphy (EEG), Blood Volume Pulse (BVP), Galvanic Skin Response (GSR), and Body Temperature (BT), can be measured and powerful processors can be used to calculate parameters such as Heart Rate (HR), Heart Rate Variability (HRV) and interbeat intervals (IBI) from these basic inputs. Studies initially collected EEG signal data in a laboratory environment while more recently wearable sensors are being used. In emotion recognition research, AI models are applied to the sensor outputs. What to measure as an emotion varies. Affect is the general sense of feeling that a person experiences. It has two features: valence (V) which is how pleasant/positive or unpleasant/negative a person feels and arousal (A) which is how active or passive a person feels. Discrete emotions such as the six widely used basic emotions (happiness, fear, surprise, anger, sadness, and disgust) can be mapped to a V-A quadrants (LVLA, LVHA, HVLA, HVHA) as shown in Fig. 1 where low is negative valence and high is positive valence. In research, emotion
measurements have been presented as combinations of A and V, the four quadrants of emotion (HVHA, LVHA, LVLA, HVLA) as well as an attempt to identify discrete emotions. To classify the emotions, features are engineered from the sensor outputs before being input to the AI models. However, while feature engineering is necessary when using traditional classifiers such as k-nearest neighbour (kNN), linear discriminant analysis (LDA), logistic regression (LR), and support vector machines (SVM), deep learning networks such as convolutional neural network (CNN) and long short-term memory (LSTM) should derive features thus removing the need for manual feature engineering. Table 1 shows feature engineering divided into three terms: extracted which are those features engineered from EEG signals; statistical which are those features engineered from non-EEG peripheral signals, and raw which are signals which are not feature engineered. The table also shows the signal and model used and what was measured as the emotion. This study aims to further research by removing the feature engineering step and classifying A and V from individual peripheral signals only. The main contributions of this work are as follows:

- Automatic feature extraction from down sampled raw peripheral physiological signal data for the prediction of emotion;
- Exploration of the possibility of using a single peripheral physiological signal for the prediction of arousal or valence.

2. RELATED WORK

Shu et al. (2018) identified several challenges in emotion recognition based on physiological signals. Before discussing traditional machine learning and deep learning approaches, they identified three stages of data preparation that influence emotion classification: emotion elicitation, data acquisition, and data preprocessing.

2.1. Emotion Elicitation

In a lab setting, emotion elicitation is the triggering of emotions. Participants experience something usually video and/or different types of music/loudness, to trigger these emotions as done by Koelstra et al. (2011), Katsigiannis and Ramzan (2017), Miranda-Correa et al. (2018) and Subramanian et al. (2016) to create the corresponding emotion recognition databases DEAP, DREAMER, AMIGOS and ASCERTAIN. Saganowski et al. (2020) in their systematic review of emotion recognition using wearables (2424 papers with 2 percent used in the final analysis), identify six stages for emotion recognition in the real-life environment, the first being the emotional model used. They note that as the different studies have different emotion models, it is impossible to compare results. Barrett (2017) proposes that emotion, at a particular moment, is predicted on experiences learned over a lifetime. Siegel et al. (2018) finds that there is no 1-1 mapping between an emotion category and a specific autonomic nervous system (ANS) response pattern, ANS being a component of the peripheral nervous system that regulates involuntary physiologic processes including heart rate, blood pressure and respiration. Focussing on measuring brain activity Saarimäki et al. (2016) and Clark-Polner et al. (2017) disagree on the fundamental understanding of the neural basis of emotion and whether brain activity patterns can be classified into e.g. six basic emotions (disgust, fear, happiness, sadness, anger, and surprise). Other studies of predicting affect from physiological signals score high and low values for arousal and valence (HA, HV, LA, LV) as done by Koelstra et al. (2011) or a combination thereof to give four quadrants of emotion HVHA, LVHA, LVLA, HVLA and as done by Katsigiannis and Ramzan (2017) and Miranda-Correa et al. (2018), or both as done by Alazrai et al. (2018). However, it is widely accepted that the ANS responds to different emotion stimuli. Goshvarpour et al. (2017) found in their investigation of emotion recognition from ECG and GSR that valence was more detectable than arousal. Vijayakumar et al. (2020) showed that arousal out-performed valence for all models for all physiological signals including GSR, RSP, BVP and skin temperature. In this study, both arousal and valence will be evaluated for each of these four physiological signals.
Table 1: Feature Engineering for Emotion Recognition

| Reference                  | Signal                | Feature   | Model       | Affective Rating         |
|----------------------------|-----------------------|-----------|-------------|--------------------------|
| This study                 | GSR, Resp, BVP, Temp  | Raw       | CNN         | Arousal, Valence         |
| Nakisa et al. (2020)       | EEG                   | Raw       | LSTM        | Quadrants                |
| Kanjo et al. (2019)        | HR,GSR,Temp           | Raw       | CNN-LSTM    | Valence                  |
| Maier et al. (2019)        | BVP, EDA              | Raw       | CNN         | Discrete                 |
| Martinez et al. (2013)     | BVP, SC, both         | Extracted | SVM         | Discrete                 |
| Riaz et al. (2021)         | EEG                   | Extracted | kNN, RF     | Arousal, Liking          |
| Galvão et al. (2021)       | EEG                   | Extracted | CNN(2D)     | Quadrants                |
| Liu et al. (2020)          | EEG                   | Extracted | SVM         | Arousal, Valence         |
| Gao et al. (2019)          | EEG                   | Extracted | DBN         | Discrete                 |
| Chao et al. (2018)         | EEG                   | Extracted | SDAE, DBN   | Arousal, Valence, Liking |
| Xu and Plataniotis (2016)  | EEG                   | Statistical | LR, SVM, NN | Arousal, Valence         |
| Vijayakumar et al. (2020)  | Peripheral EEG, ECG, GSR, RSP | Statistical | LSTM, SVM, kNN | Discrete                 |
| Song et al. (2019)         | HR, GSR, O2           | Statistical | RF          | Quadrants                |
| Wen et al. (2014)          |                       |           |             |                          |

2.2. Data Acquisition and Datasets

Data acquisition involves connecting different types of sensors to human participants to measure physiological activities. Invasive sensors, such as EEG electrodes, 12 or more, to measure brain activity, conventional 12-lead ECG to measure heart rate and a finger clip to measure pulse, were used in the creation of the DEAP dataset by Koelstra et al. (2011). These sensors are being replaced with low-cost off-the-shelf sensors. The Emotive EPOC wireless EEG Headset and the wireless SHIMMER ECG sensor were used to compile the DREAMER dataset by Katsigiannis and Ramzan (2017) and the AMIGOS EEG, ECG and GSR based datasets by Miranda-Correa et al. (2018). Other off-the-shelf sensors were used to measure EEG, ECG, GSR in the ASCERTAIN dataset by Subramanian et al. (2016). The Empatica wristband was used to detect optimal user experience from BVP and Electrodermal Activity (EDA), an umbrella term for GSR, by Maier et al. (2019). To further enhance the non-invasiveness of sensors, although not used in studies to detect emotion, sensors embedded in textiles have been used to measure respiration by Piuzzi et al. (2020) and to measure HR, skin temperature and respiration as highlighted in a review by Angelucci et al. (2021). Datasets from several of these studies are available for further research. The datasets can be briefly described as follows:

- DEAP: Koelstra et al. (2011) collected data from 40 sensors on 32 participants who scored arousal, valence, liking and dominance on a likert-like scale of 1 to 9 while watching 40 60-second videos;
- DREAMER: Katsigiannis and Ramzan (2017) collected data from EEG and ECG sensors on 23 participants who scored valence, arousal, and dominance;
- ASCERTAIN: Subramanian et al. (2016) collected data from EEG, ECG, GSR and Facial expression using off-the-shelf sensors on 58 participants who scored emotional ratings and personality scales;
- K-Emocon: Kim and André (2008) had 32 subjects who participated in paired live debates while wearing off-the-shelf sensor devices: the E4 Empatica for pulse rate, 3 axis, body temp, EDA and derived (HR, IBI); the Polar H7 Bluetooth heart rate sensor and the Neurosky mindwave with 2 dry sensors for EEG signals. There were 16 live 10 minute debates. Arousal and valence labels were assigned by observers at 5 second intervals.

As this study is more interested in peripheral psychological signals, the DEAP dataset was chosen for two main reasons: it has four peripheral physiological responses compared to less than four in other studies and both labels arousal and valence are used.

2.3. Data Preprocessing

The output from the sensors are analog signals which vary in both amplitude and frequency for the different physiological measurements. Following sampling, varying degrees of feature engineering is done to extract features according to the learning techniques used. For traditional ML methods using classification algorithms such as Random
End-to-End Emotion Recognition using Peripheral Physiological Signals

As early as 2013, Martinez et al. (2013) introduced the idea of using deep learning (DL) for affective modelling from physiological signals. They showed the potential of DL as a mechanism for eliminating manual feature extraction. CNN is a deep learning network that uses convolution to enable the model itself to extract the features. Therefore, the raw signal itself can be input to a machine learning model. Fahimi et al. (2019) in their study of classification into attentive and non-attentive states, showed that using CNN to extract features from a raw signal EEG signal outperformed a band pass filtered EEG signal. Lee et al. (2020) focussing on emotion recognition using a Photoplethysmogram (PPG) signal, have extracted features using CNN. They extracted 200 features from the PPG signal itself using a CNN network in parallel to extracting 30 features from the NN interval calculated from the PPG signal and then fused these 230 features with 10 statistically calculated features prior to classifying the output into high and low valence. Kanjo et al. (2019) compared three models (MLP, CNN and CNN-LSTM) using only raw signals gathered from the body (HR, GSR, body temperature, motion data), environment (noise, UV, air pressure) and location data (GPS). The CNN-LSTM model, with all the raw signals as input, outperformed the other models (MLP and CNN) and outperformed the experiments that used subsets of the signals such as the body only sensors. Maier et al. (2019) used raw BVP and EDA signals collected from wrist worn devices with CNN to predict affective states (boredom, flow, stress). Goshvarpour et al. (2017) used a matching pursuit algorithm (MP) and wavelet dictionary to extract MP coefficients on seven subjects and then statistically analysed for significance concluding that on average GSR is a better predictor of valence than of arousal and that the classification of both valence and arousal using ECG was better than GSR. In this study, four

For peripheral signals used for emotion recognition, the features extracted are statistical calculations on the signals with different statistical features for different signal type. For example, DEAP’s Koelstra et al. (2011) extracted features for GSR are average skin resistance, average of derivative, average of derivative for negative values only (average decrease rate during decay time), proportion of negative samples in the derivative versus all samples, number of local minima in the GSR signal, average rising time of the GSR signal, 10 spectral power in the [0-2.4]Hz bands, zero crossing rate of Skin conductance slow response (SCSR) [0-0.2]Hz, zero crossing rate of Skin conductance very slow response (SCVSR) [0-0.08]Hz, SCSR and SCVSR mean of peaks magnitude. Emotions are then classified using a selection of statistical features as input to machine learning models. Feature engineering has been used with both traditional machine learning algorithms SVM, SDAE, KN, RF as well as with deep learning models neural networks (NN), DBN CNN and CNN-LSTM. However, Oh et al. (2020) do not use statistical features. They used a low sampling rate of 32 Hz of the raw signal and then automatically calculated five HRV parameters (heart rate, HRV amplitude, LF, HF, LF/HF) and two RSP parameters (RSP value and RSP rate) using Biotrace and software processing. They then created four CNN models with different input shapes for combinations of the parameters calculated to classify six discrete emotions finding that a combination of all seven parameters was the best classifier.
of the peripheral signals from the DEAP dataset will be used. The GSR signal data was collected by measuring galvanic skin response on the left middle and ring finger. A respiration belt was used to measure RSP. A plethysmograph on the left thumb was used to measure the BVP and a temperature sensor on the left pinky was used to measure temperature. The signals were then preprocessed by downsampling the data to 128Hz, segmenting it into 60 second trials and removing a 3 second pre-trial baseline, resulting in the data preprocessed python zip file provided by DEAP. These preprocessed signals form the raw signals that will be further down sampled and input to a CNN, which will itself, extract the features prior to classification of both arousal and valence.

3. METHODOLOGY

Design of the CNN involved outlining a process with the steps data preparation, data modelling, classification and validation. Manual hyperparameter optimisation was performed. First, initial parameters were set for the model that were similar to those used by Maier et al. (2019) as follows: four convolutional layers (32 filters, kernel size 3) with ReLU activations and alternate max pooling layers, a final dense layer set to one for one class and a softmax activation. Then, tuning of the hyperparameters for the model involved running several tests until a model was found that did not overfit. Fig.2 shows the accuracy and loss for each of the four sampled raw signals GSR, respiration, BVP, temperature for the final chosen hyperparameters which are shown in Fig.3.

Cross-validation was used to assess the generalisation ability of the predictive model. The k-fold cross validation method with k set to 10 was used. Analysis of the dataset found that the arousal and valence samples per binary class in the dataset were not balanced as shown in Fig. 4. Therefore, a stratified K-Folds cross-validator was used. This stratified validator performs the sampling in such a way that maintains the class proportions in the fold subsets that reflects the proportions in the training set.

3.1. Data Preparation

The preprocessed python version of DEAP dataset has been used. It has 32 participant .dat files which each have a data array of shape 40x40x8064 for video/trialxchannelxdata and a labels array of shape 40x4 for video/trialxlabel. In the DEAP dataset, the data channels have been downsampling from the original 512 Hz signals to 128 Hz. The channels comprise 32 Electroencephalogram (EEG) channels, horizontal and vertical Electrooculography (EOG) channels, horizontal and vertical Electromyography (EMG) channels and a channel each for GSR, RSP, BVP and Temperature. The labels comprise valence, arousal, dominance and liking each with floating point numbers in the range of 1 to 9. Data preparation was executed in the Google’s collaborative hosted Jupyter notebook service. To prepare the data channels for input to the CNN conv1d model, the 32 dat files were read into the notebook. The EEG, EoG and EMG channels were removed from the data array resulting in a data array shape of 1028 x 4 x 8064 (samples x channels x timesteps) derived from 32 participants in 40 video trials (1028), 4 channels and 63 seconds signals sampled at 128 Hz (8064) readings. Each of the four remaining channels (GSR, Respiration, BVP and Temperature) were rescaled to the range [0,1]. The 8064 timesteps represent a 63 minute sensed signal which has been sampled at the rate of 128 Hz (samples per second).

By further sampling the 8064 timestep readings, taking every 40th reading, the signal was reduced to 201 timesteps. This is in effect equivalent to a sampling frequency of 3 Hz of the original signal. As shown in Fig. 5 for one sample on each of the four channels, the integrity is maintained after down sampling the signals. Finally, for each of the four channels, the array was reshaped to (none,201,1) for the conv1d format (batch size, timesteps, feature) whereby with batch size set to none, it can be varied in the model. To prepare the label channels, the labels dominance and liking were removed and the values for the remaining channels, arousal and valence which ranged from 1 to 9, were processed into binary values 0 and 1 using a threshold of 4.5 for binary classification.

3.2. Proposed Model

The CNN is a deep learning class of Artificial Neural Network (ANN) which has been used successfully for visual imagery where the data has two dimensions for images or three dimensions for video. The physiological signal data is one dimensional which can be used with CNN conv1d. The input shape, as prepared in the previous section, for CNN conv1d time series problem, is a 3-dimensional array comprising samples, timesteps and features. There are 1028 samples comprised of the 40 trials done by the 32 participants. There are 210 timesteps for each signal. There is one feature per signal. The CNN model is sequential. There are three convolutional layers all with a kernel size of three and each respectively having 8, 16 and 32 filters. These are alternated with maxpooling by 3. After the convolutions and maxpooling, the features are flattened to one fully connected layer which is followed by a dense layer with one class as arousal and valence were classified separately. The Adam optimisation algorithm was used with the default
Figure 2: Accuracy and Loss for Valence and Arousal of each Peripheral Signal
learning rate and loss was binary cross entropy. To estimate the generalisation performance, we conducted a 10-fold cross validation using a stratified k-fold validator to preserve the class proportions in the dataset.

4. RESULTS AND DISCUSSION

We set out to predict arousal and valence from peripheral physiological signals. The GSR, Respiration, BVP and Temperature channels from the DEAP dataset were downsampled and input to a CNN which itself extracted features. As seen in table 2, a selection of results from studies of emotion recognition, valence does not outperform arousal or vice versa in all studies. For a study with more than one model, the result for the best performing model was selected. Salari et al. (2018) found that their CNN model outperformed the DNN model achieving an accuracy of 68.59% and 62.51% respectively for arousal and valence. However, they used EEG signals and extracted features which were input to the models. Xu and Plataniotis (2016) achieved much higher classification accuracies for arousal at 88.33% and valence at 88.59% using a deep belief network (DBN). These scores were achieved using the EEG signals but they removed the first 20s of the 60s signal in a preprocessing step. Riaz et al. (2021) did their own experiment using the commercially available Emotiv Insight headset to sense EEG signal and achieved accuracies of 80.5% for arousal and 70.3% for valence. They extracted statistical features and classified arousal and valence using the traditional SVM classifier. Lee et al. (2020) using only a PPG signal achieved accuracies of 80.9% for arousal and 82.1% for valence. They extracted 200 features from the raw signal using a CNN model. But before classification they fused these 200 features with a further 30 features extracted using a second CNN model from intervals derived from the PPG signal and another 10 statistically derived features. Katsigiannis and Ramzan (2017), using low-cost off-the-shelf sensors to sense EEG and ECG signals and classification using traditional SVM with RBF achieved accuracies of 62.57% using only the ECG device and 62.49% for arousal using only EEG signals. In our methodology, four of the peripheral signals were used from the DEAP dataset. These raw signals were downsampled and the CNN itself extracted the features for classification of both arousal and valence. As shown in table 3, the best accuracy was achieved for RSP at 64.2% for valence followed by an accuracy of 63.5% for arousal using the BVP signal. GSR, at 62.5%, was second best for arousal followed by RSP at 61.3%. For valence, BVP was second best at 61.1% following closely by RSP at 60.2%. Finally, the temperature signal had the lowest accuracy for both arousal (56.5%) and valence (53.2%).

5. CONCLUSION

In this paper, the use of raw peripheral physiological signals for emotion recognition has been investigated. Using the DEAP dataset, each of the GSR, RSP, BVP and temperature signals were down sampled to approximately 3 Hz. A CNN then automatically extracted features from each signal separately, each of which were then used to classify arousal and valence. The accuracies and F1 scores, when compared to other studies, indicate that a single raw peripheral physiological signal is not useful for predicting arousal or valence using our current CNN model. Moreover, we found that the GSR, RSP and
End-to-End Emotion Recognition using Peripheral Physiological Signals

Figure 5: Physiological signals before and after sampling

Table 2: Affective State Scores

| Ref.                  | Dataset | Physiological Response | Arousal Acc | Valence Acc | Model   |
|-----------------------|---------|------------------------|-------------|-------------|---------|
| This work             | DEAP    | RSP                    | 61.25       | 64.24       | CNN     |
| This work             | DEAP    | GSR                    | 62.50       | 60.18       | CNN     |
| This work             | DEAP    | BVP                    | 63.49       | 61.13       | CNN     |
| Riaz et al. (2021)    | Own Experiment | EEG | 80.5 | 70.3       | SVM    |
| Lee et al. (2020)     | DEAP    | PPG                    | 80.9        | 82.1        | CNN     |
| Barradas Chacon et al. (2019) | DEAP | HRV,BPM,Pulse | 57.0 | 60.6       | FNN    |
| Salari et al. (2018)  | DEAP    | EEG                    | 68.59       | 62.51       | CNN     |
| Salari et al. (2018)  | DEAP    | EEG                    | 56.36       | 60.72       | DNN     |
| Katsigiannis and Ramzan (2017) | DREAMER | EEG | 62.17 | 62.49 | SVM with RBF |
| Katsigiannis and Ramzan (2017) | DREAMER | ECG | 62.37 | 62.37 | SVM with RBF |
| Katsigiannis and Ramzan (2017) | DREAMER | Fusion of EEG and ECG | 62.32 | 61.84 | SVM with RBF |
| Xu and Platanitis (2016) | DEAP | EEG | 88.33 | 88.59 | DBN     |

Table 3: Results

| Signal     | Accuracy   | A   | V   | F1-Score |
|------------|------------|-----|-----|----------|
| GSR        | 62.5       | 60.2| 0.699| 0.663    |
| Respiration| 61.3       | 64.2| 0.677| 0.669    |
| BVP        | 63.5       | 61.1| 0.673| 0.632    |
| Temperature| 56.5       | 53.2| 0.626| 0.558    |

BVP signals outperformed the temperature signal. Our future work will further fine tune the hyperparameters and include regularisation methods to improve the network’s performance. It will investigate classification of each of the four peripheral signals into the four quadrants of emotion HVHA, LVHA, LVLA and HVLA. Further, the GSR, RSP and BVP peripheral signals will be combined to classify arousal and valance separately as well as classifying the four emotion quadrants.

6. ACKNOWLEDGEMENT

This work is funded by the Technological University of the Shannon: Midlands President’s Doctoral Scholarship 2020.
REFERENCES

Bahareh Nakisa, Mohammad Naim Rastgoo, Andry Raktoneira, Frederic Maire, and Vinod Chandran. Automatic emotion recognition using temporal multimodal deep learning. IEEE Access, 8:225463–225474, 2020.

Eiman Kanjo, Eman MG Younis, and Chee Siang Ang. Deep learning analysis of mobile physiological, environmental and location sensor data for emotion detection. Information Fusion, 49:46–56, 2019.

Marco Maier, Daniel Elsner, Chadly Marouane, Meike Zehnle, and Christoph Fuchs. Deepflow: Detecting optimal user experience from physiological data using deep neural networks. In AAMAS, pages 2108–2110, 2019.

Hector P Martinez, Yoshua Bengio, and Georgios N Yannakakis. Learning deep physiological models of affect. IEEE Computational intelligence magazine, 8(2):20–33, 2013.

Majid Riaz, Muhammad Majid, and Junaid Mir. Emotion recognition using electroencephalography in response to high dynamic range videos. In 2021 International Conference on Information Technology (ICIT), pages 565–570. IEEE, 2021.

Filipe Galvão, Soraia M Alarcão, and Manuel J Fonseca. Predicting exact valence and arousal values from eeg. Sensors, 21(10):3414, 2021.

Junxiu Liu, Guopei Wu, Yuling Luo, Senhui Qiu, Su Yang, Wei Li, and Yifei Bi. Eeg-based emotion classification using a deep neural network and sparse autoencoder. Frontiers in Systems Neuroscience, page 43, 2020.

Zhilin Gao, Xingran Cui, Wang Wan, and Zhongze Gu. Recognition of emotional states using multiscale information analysis of high frequency eeg oscillations. Entropy, 21(6):609, 2019.

Hao Chao, Huilai Zhi, Liang Dong, and Yongli Liu. Recognition of emotions using multichannel eeg data and dbn-gc-based ensemble deep learning framework. Computational intelligence and neuroscience, 2018, 2018.

Haiyan Xu and Konstantinos N Plataniotis. Affective states classification using eeg and semi-supervised deep learning approaches. In 2016 IEEE 18th International Workshop on Multimedia Signal Processing (MMSP), pages 1–6. IEEE, 2016.

Sowmya Vijayakumar, Ronan Flynn, and Niall Murray. A comparative study of machine learning techniques for emotion recognition from peripheral physiological signals. In 2020 31st Irish Signals and Systems Conference (ISSC), pages 1–6. IEEE, 2020.

Tengfei Song, Wenming Zheng, Cheng Lu, Yuan Zong, Xilei Zhang, and Zhen Cui. Mped: A multi-modal physiological emotion database for discrete emotion recognition. IEEE Access, 7:12177–12191, 2019.

Wanhui Wen, Guanyuan Liu, Nanpu Cheng, Jie Wei, Pengchao Shangguan, and Wenjin Huang. Emotion recognition based on multi-variant correlation of physiological signals. IEEE Transactions on Affective Computing, 5(2):126–140, 2014.

Lin Shu, Jinyan Xie, Mingyue Yang, Ziyi Li, Zhenqi Li, Dan Liao, Xiangmin Xu, and Xinji Yang. A review of emotion recognition using physiological signals. Sensors, 18(7):2074, 2018.

Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. Deap: A database for emotion analysis; using physiological signals. IEEE transactions on affective computing, 3(1):18–31, 2011.

Stamos Katsigiannis and Naeem Ramzan. Dreamer: A database for emotion recognition through eeg and ecg signals from wireless low-cost off-the-shelf devices. IEEE journal of biomedical and health informatics, 22(1):98–107, 2017.

Juan Abdon Miranda-Correa, Mojtaha Khomami Abadi, Nicu Sebe, and Ioannis Patras. Amigos: A dataset for affect, personality and mood research on individuals and groups. IEEE Transactions on Affective Computing, 12(2):479–493, 2018.

Ramanathan Subramanian, Julia Wache, Mojtaha Khomami Abadi, Radu L Vieriu, Stefan Winkler, and Nicu Sebe. Ascertain: Emotion and personality recognition using commercial sensors. IEEE Transactions on Affective Computing, 9(2):147–160, 2016.

Stanislaw Saganowski, Anna Dutkowiak, Adam Dziadek, Maciej Dzieycz, Joanna Komoszyńska, Weronika Michalska, Adam Polak, Michal Ujma, and Przemyslaw Kazienko. Emotion recognition using wearables: A systematic literature review-work-in-progress. In 2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), pages 1–6. IEEE, 2020.

Lisa Feldman Barrett. How emotions are made: The secret life of the brain. Pan Macmillan, 2017.

Erika H Siegel, Molly K Sands, Wim Van den Noortgate, Paul Condon, Yale Chang, Jennifer Dy, Karen S Quigley, and Lisa Feldman Barrett. Emotion fingerprints or emotion populations? a meta-analytic investigation of autonomic features of emotion categories. Psychologica Bulletin, 144(4):343, 2018.

Heini Saarimäki, Athanasios Gotsopoulos, liro P Jääskeläinen, Jouko Lampinen, Patrik Vuilleumier, Riitta Hari, Mikko Sams, and Lauri Nummenmaa. Discrete neural signatures of basic emotions. Cerebral cortex, 26(6):2563–2573, 2016.

Elizabeth Clark-Polner, Timothy D Johnson, and Lisa Feldman Barrett. Multivoxel pattern analysis does not provide evidence to support the existence of basic emotions. Cerebral Cortex, 27(3):1944–1948, 2017.
Rami Alazrai, Rasha Homoud, Hisham Alwanni, and Mohammad I Daoud. Eeg-based emotion recognition using quadratic time-frequency distribution. *Sensors*, 18(8):2739, 2018.

Atefeh Goshvarpour, Ataollah Abbasi, and Ateke Goshvarpour. An accurate emotion recognition system using ecg and gsr signals and matching pursuit method. *Biomedical journal*, 40(6):355–368, 2017.

Emanuele Piuazzi, Stefano Pisa, Erika Pittella, Luca Podestà, and Silvia Sangiovanni. Wearable belt with built-in textile electrodes for cardio—respiratory monitoring. *Sensors*, 20(16):4500, 2020.

Alessandra Angelucci, Matteo Cavicchioli, Ilaria A Cintorino, Giuseppe Lauricella, Chiara Rossi, Sara Strati, and Andrea Aliverti. Smart textiles and sensorized garments for physiological monitoring: A review of available solutions and techniques. *Sensors*, 21(3):814, 2021.

Jonghwa Kim and Elisabeth André. Emotion recognition based on physiological changes in music listening. *IEEE transactions on pattern analysis and machine intelligence*, 30(12):2067–2083, 2008.

Maciej Dzieżyk, Martin Gjoreski, Przemysław Kazienko, Stanisław Saganowski, and Matjaž Gams. Can we ditch feature engineering? end-to-end deep learning for affect recognition from physiological sensor data. *Sensors*, 20(22):6535, 2020.

Xinzheng Xu, Tianming Liang, Jiong Zhu, Dong Zheng, and Tongfeng Sun. Review of classical dimensionality reduction and sample selection methods for large-scale data processing. *Neurocomputing*, 328:5–15, 2019.

Fateme Fahimi, Zhuo Zhang, Wooi Boon Goh, Tih-Shi Lee, Kai Keng Ang, and Cuntai Guan. Inter-subject transfer learning with an end-to-end deep convolutional neural network for eeg-based bci. *Journal of neural engineering*, 16(2):026007, 2019.

SeungJun Oh, Jun-Young Lee, and Dong Keun Kim. The design of cnn architectures for optimal six basic emotion classification using multiple physiological signals. *Sensors*, 20(3):866, 2020.

MinSeop Lee, Yun Kyu Lee, Myo-Taeg Lim, and Tae-Koo Kang. Emotion recognition using convolutional neural network with selected statistical photoplethysmogram features. *Applied Sciences*, 10(10):3501, 2020.

Luis Alberto Barradas Chacon, Artem Fedoskin, Ekaterina Shcheglakova, Sutthida Neamsup, and Ahmed Rashed. Emotion analysis using heart rate data. In *International Conference on Database and Expert Systems Applications*, pages 147–154. Springer, 2019.

Soorena Salari, Amin Ansarian, and Hajar Atrianfar. Robust emotion classification using neural network models. In *2018 6th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS)*, pages 190–194. IEEE, 2018.