

# A Comparative Study of Physiological Feature Selection Methods for Emotion Recognition

Andreas De Lille

**Abstract**—An emerging topic of research is emotion recognition based on physiological signals and machine learning. Emotion recognition is the process of recognizing a person's emotional state. In this work the emotion recognition was done using a combination of physiological signals and machine learning. The flow of this approach is to record physiological signals from a person, extract features and feed them to a machine learning algorithm. This algorithm will then predict the user's emotional state. Even though a lot of research has been done, there is no agreement on what features are important. This work tries to overcome this problem by comparing a wide range of features with several feature selection methods.

**Index Terms**—Emotion recognition, physiological signals, machine learning, feature selection methods

## 1 INTRODUCTION

Emotion recognition is the process of recognizing a person's emotion. Observing and recognizing emotion can be done in several ways. Psychology makes a clear distinction between physiological behaviour and a person's expression of emotion. The expression is often prone to social masking, the process of hiding emotion to conform to social standards and ideas, making it less reliable. The physiological behaviour on the other hand is much harder to control, making it more reliable.

This work will focus on emotion recognition using physiological signals. Next the classification of emotions will be explained in Section 2, before an introduction of physiological signals is given in Section 3. What follows in Section 4 is an overview of the used features. The problem statement and goal of this work is given in 5. Sections 6 and 7 give the used approach, results and conclusion.

## 2 CLASSIFICATION OF EMOTIONS

Before emotion can be recognized, different emotions need to be defined. One way to do this is to use several distinct emotions, e.g.

anger, joy, sad and pleasure. The advantage of this approach is that all emotions have a clear label. The disadvantage is that this model is often not complex enough to fully represent an emotion state. To solve this problem, the bipolar valence-arousal model was introduced. This model puts each emotion in a two dimensional space. The first dimension, ranges from inactive to active and indicates how active a person is feeling. The next dimension is valence. This dimension indicates how pleasant or unpleasant the emotion is perceived.

The valence-arousal model has the advantage that an emotion can be defined, without needing to explicitly labelling it. All discrete emotions can be mapped to the valence-arousal space. For example, excitement corresponds to an active feeling with a pleasant experience, meaning that it will be in the high valence, high arousal quadrant of the space. A depressed feeling, on the other hand, will have a low valence and low arousal. As a result depressed is mapped in the low valence, low arousal quadrant. The mapping of other emotions can be done similarly and is shown in Figure 1.

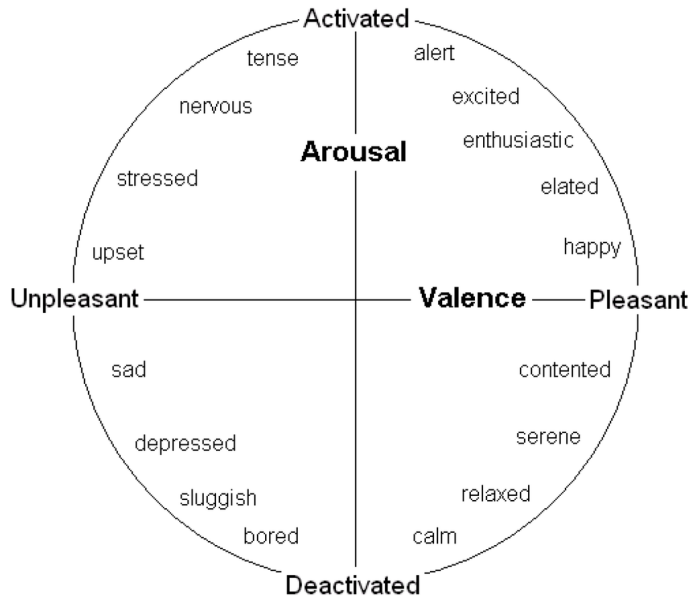


Fig. 1. The arousal - valence model maps emotions in a two dimensional plane.

### 3 PHYSIOLOGICAL SIGNALS

In case machine learning is used for emotion recognition, some physiological signals are taken as input. The machine learning will then output a valence or arousal score. In short, machine learning can be defined as an input output model that predicts output values for different samples based on the inputs. The inputs are features of the input samples, e.g. the frequency or amplitude of a signal.

To do emotion recognition with machine learning, good features are required. This work focusses on physiological signals that can be split into two groups of features. The first group is are the peripheral signals, a.o. heart rate, blood pressure, respiration rate, perspiration, etc.

The second group of features originate from the brain. These signals are recordings of brain activity using electroencephalography (EEG). EEG is a technique that measures electrical activity of the brain, by placing electrodes on the scalp. Even though EEG is very noisy by nature as the signal is distorted by the bone between the cortex and the electrodes it still provides significant insight in the brain activity. The electrodes are placed according to the 10/20 system, that labels each location. The locations

and the corresponding channel names used in this work are visible in Figure 2.

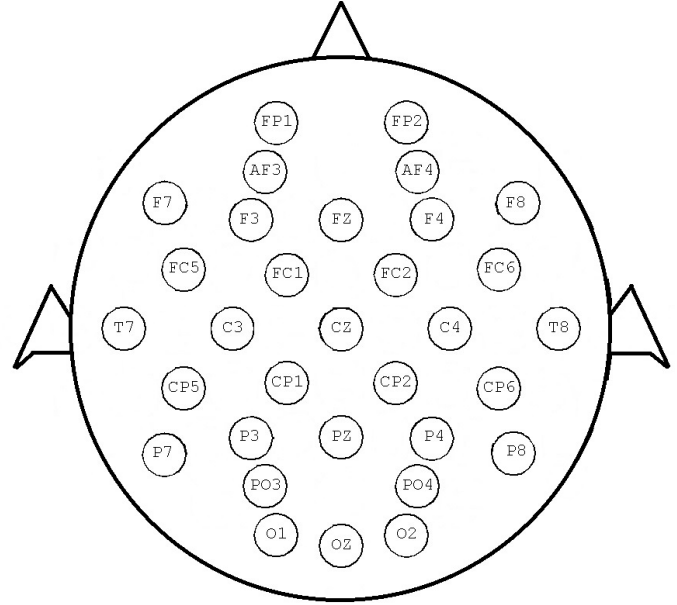


Fig. 2. Placement of the 32 electrodes in this thesis.

EEG measures electrical activity at each channel. Each measurement can be split in different frequency bands, with medical relevance. The frequency bands are:

- **Alpha:** 8-13Hz, indicate how relaxed and/or inactive the brain is.
- **Beta:** 13-30HZ, indicate a more active and focused state of mind.
- **Gamma:** 30-50Hz, relate to simultaneous processing of information from different brain areas.
- **Delta:** 0-4hz, these waves are generated during dreamless sleep and meditation.
- **Theta:** 4-8Hz, occurs during dreaming.

### 4 FEATURES

In this work the Dataset for Emotion Analysis using Physiological Signals (DEAP) was used. This dataset consists of recordings of a physiological experiment. This experiment recorded emotional reactions of 32 subjects. The reactions were triggered using music video excerpts. Each subject watched 40 one-minute videos, while several physiological signals were recorded. These physiological signals

consist of 32 channel, 512Hz EEG signals combined with the following peripheral signals:

- galvanic skin response (GSR), which measures perspiration
- respiration belt, which measures the respiration rate
- plethysmograph, which measures the blood pressure
- skin temperature

For the peripheral signals, statistical values of each channel are used. These statistical values are minimum, maximum, variation, standard deviation, average and median of each channel.

EEG features use a different approach, here the power spectral density (PSD) of each EEG signal is calculated. The PSD gives the distribution of the signal's energy in the frequency domain. Another power feature is the differential entropy (DE), it is proven that the differential entropy of a certain band is equivalent to the logarithmic power spectral density for a fixed length EEG sequence.

$$DE_{ch} = \log(PSD_{ch})$$

The most used features for valence classification are asymmetry features that measure asymmetry between two channels. This can be done in two ways. The first way is the differential asymmetry (DASM) which is defined as:

$$DASM = DE_{left} - DE_{right}$$

Another way to measure the asymmetry is by division. The Rational Asymmetry (RASM) does exactly this and is given by:

$$RASM = \frac{DE_{left}}{DE_{right}}$$

The asymmetry features can also be used in the other direction. Instead of looking at the asymmetry between left and right, one can also compare the frontal power with the posterior power. This is known as the caudality. The differential caudality (DCAU) and rational caudality (RCAU) are defined as:

$$DCAU = DE_{front} - DE_{post} \quad RCAU = \frac{DE_{front}}{DE_{post}}$$

Another feature are the power fractions. These fractions give the power distribution for each channel and are defined as:

$$frac_{band,channel} = \frac{power_{band,channel}}{power_{total,channel}}$$

## 5 PROBLEM STATEMENT

Looking at different features in Table 1, one can see that the features counts quickly adds up. Having a total of 894 features will increase the risk for overfitting significantly. Additionally, not all 897 features are important. There is no agreement on the important features in literature.

TABLE 1

An overview of the different features that were compared in this thesis.

Name	Type	Channels	Freq bands	Total
PSD	EEG	32	6	192
DE	EEG	32	6	192
DASM	EEG	13	6	78
RASM	EEG	13	6	78
DCAU	EEG	11	6	66
RCAU	EEG	11	6	66
Frac	EEG	32	5	160
Alpha / Beta	EEG	32	1	32
EEG Total				864
Name	Type	Total		
HR	non-EEG	6		
Plethysmograph	non-EEG	6		
GSR	non-EEG	6		
ST	non-EEG	6		
RSP	non-EEG	6		
non-EEG Total		30		
Overall Total		894		

The goal of this work is twofold, first important features are needed for person specific emotion recognition. Second, feature are needed that enable the machine learning algorithm to work in a cross-person setting. In the second goal the machine learning algorithm is trained on several person and tested on other people. Working in a cross-subject setting is more challenging, because physiological signals are personal by nature.

## 6 USED APPROACH AND RESULTS

Several feature selection methods were compared. A feature selection method is a method that takes as input a large set of features, and returns a (smaller) set of important features. There exists three categories of feature selection methods, filter, wrapper and embedded.

Filter methods are the less complex method that simply use a statistical test to rank the features. An example would be to use the Pearson correlation between a feature and the valence. Features with low or zero importance would then be filtered. This method is fast and simple, but is not capable to find groups of relevant features, as each feature is reviewed on its own. Groups of relevant features are features that are not important on their own, but give insight when combined with other features.

Wrapper methods are somewhat more advanced, they use an arbitrary machine learning technique and look at the assigned weights. Features with large weights have more influence on the output than features with low weights. Absolute values are used here; large positive and negative values have the same amount of influence on the result. The performance of this methods are often dependent on the underlying machine learning algorithms.

The last category of feature selection methods are the embedded methods. Embedded methods also use an underlying machine learning technique. This time the machine learning algorithm has a build-in feature selection functionality though. Random forest are a good example of this category. Random forest use multiple decision trees that each have a random subset of the samples and features.

Building a decision tree starts with all samples in one node. This node will therefore be impure, meaning that it has samples of multiple classes. In the next step, the data will be split iteratively. In each iteration, a feature is picked at random and the data is split according to the different values of features. This split is done so that it lowers the impurity of the child nodes. After several iterations, the node will eventually be pure. A pure node contains only samples of one class.

Each feature corresponds to a drop in impurity. By averaging the drop in impurity over all trees, one can estimate the importance of the different features quite well. The advantage is that this method is capable of finding groups

of relevant features. When a group of random features is selected at different nodes in a decision tree, the impurity will drop significantly. This will result in a higher feature ranking.

The following approach was used, first the DEAP dataset is splitted in a train and test set (30 and 10 samples respectively). The test set is kept separate, while the random forest feature selection is performed on the train set. This step returns the top 30 features. With these 30 features a model is built by iteratively repeating the following steps:

- 1) add a new feature to the feature set
- 2) determine the cross validation error and standard deviation (std)
- 3) if the validation error and std are better than the previous best, the feature is kept. Otherwise the feature is neglected.

Note that this model is built using a support vector machine (SVM) as SVM's have proven themselves in emotion recognition studies. The cross validation step ensures that the chosen features also work well on unseen samples. The final result is evaluated on the test set.

This was done for each person, the resulting performance was averaged over all persons giving a performance of 70 and 73.75 % with std 13.9 and 13.1 % for arousal and valence respectively.

To review which features were selected, groups of similar EEG features were created. These groups are:

- Power features: the PSD and DE features
- Asymmetry features: DASM, RASM, DCAU, RCAU
- Fraction features: the fractions

The non-EEG features are:

- heart rate
- GSR
- respiration rate
- blood pressure
- skin temperature

The distribution of the selected features is given in Figure 3 for arousal and Figure 4 for valence, the legend is given in Figure 5.

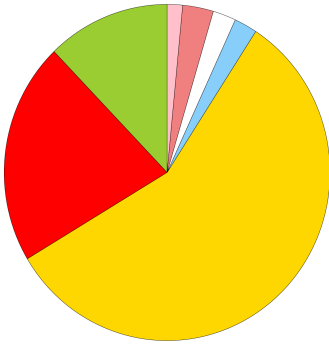


Fig. 3. Selected features for arousal

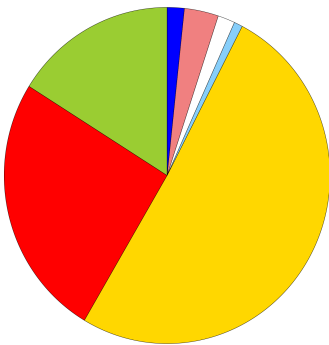


Fig. 4. Selected features for valence



Fig. 5. Each color and its corresponding feature category.

It is clear that for both arousal and valence, the asymmetry features are the most important. The second most important category are the power features. The non-EEG features seem to have very little influence on the result. To verify that the non-EEG feature are worse than the EEG features, the setup was run two additional times. One time with only EEG features and

one time with only non-EEG features. The results are visible in Figure 6 and Figure 7 for arousal and valence respectively.

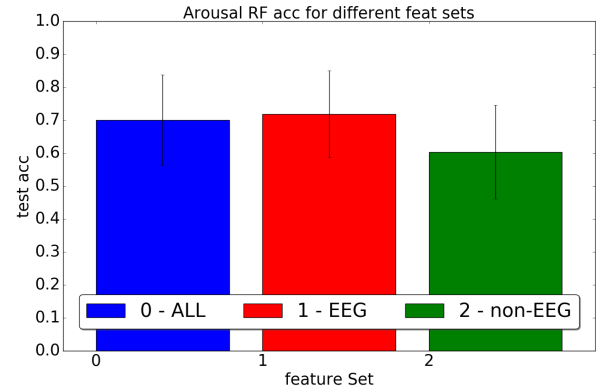


Fig. 6. The performance of arousal prediction for all, EEG and non-EEG features.

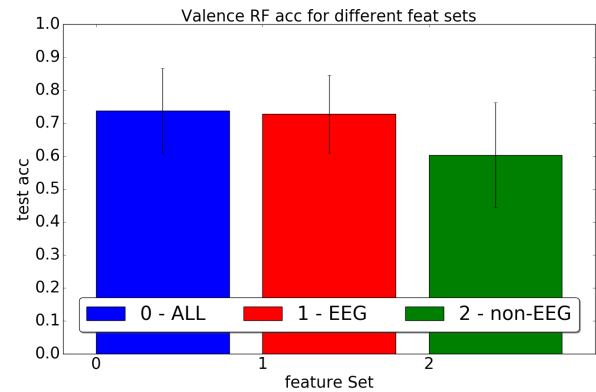


Fig. 7. The performance of valence prediction for all, EEG and non-EEG features.

Statistical tests conclude that the EEG and ALL feature sets have better performance than the non-EEG sets, albeit with a p-value of only 80%. Repeating this with more data might give a more definitive results.

## 7 CONCLUSION AND FUTURE WORK

### REFERENCES

- [1] G.D. Abowd, A.K. Dey and P.J. Brown, N. Davies, M. Smith and P. Steggles. Towards a Better Understanding of Context and Context-Awareness. In *Proceedings of the 1st International Symposium on Handheld and Ubiquitous Computing (HUC)*, 1999, pages 304–307.