Taking it at face value: confound or defining feature?

July 21, 2020

Utrecht University

Thesis by Heleen Kerstholt
Thesis advisor: Sjoerd Stuit
Second reader: Micheal De

Bachelor artificial intelligence at Utrecht University
Course code: KI3V12011
Subject: 2019-2020 JAAR Bacheloreindwerkstuk CKI for 7,5 ECTS

1 Introduction

In the human eye, faces are not particularly extraordinary images. However, this is surely not the case for other observers. For instance, in the field of computer vision faces have been remarkably difficult stimuli. In 2000, a review (Pantic & Rothkrantz, 2000) was written on the possibilities of an automatic interpreter of faces and facial expressions; a then still far too difficult problem to solve. In the past 20 years, the field of computer vision has made applaudable progress in expression recognition. However, in present day still little is known on the mechanism behind these facial expressions and what defines them.

Facial expressions likely originated from the evolutionary advantages accompanying them. The widening of the eye, for example in the fear expression, causes greater perceptibility for sensory input, giving actors of this facial expression an evolutionary benefit in threatful situations (Lee, Susskind, & Anderson, 2013). Frith (2009) substantiated the theory that these facial expressions were first perceived by observers as public information on the environment, which later evolved into a complex communication system. The research on how humans understand and process these facial expressions is in agreement that emotional expressions are differently processed and capture more attention than neutral faces (Vuilleumier & Schwartz, 2001). However, which emotion affects observers the most is still a topic of debate. The hypothesis that happy faces are more efficient in capturing attention is called the happiness superiority effect (HSE) while the hypothesis that angry faces capture attention more effectively is called the angry superiority effect (ASE). Specifically, some studies have reported a HSE, while others reported an ASE (Becker, Anderson, Mortensen, Neufeld, & Neel, 2011; Savage, Lipp, Craig, Becker, & Horstmann, 2013).

In reviews on the existing literature, researchers concluded that in most (if not all) studies little unconfounded evidence exists for either ASE or HSE (Becker et al., 2011; Frischen, Eastwood, and Smilek, 2008). The promising results on ASE are likely to have resulted from experiment design issues and low level visual information. Studies which recorded an HSE also have been criticised due to the happy stimuli in these studies that often display an open mouth. This makes the stimulus more salient due to the high contrast of the mouth area. In fact, Savage et al. (2013) described these confounds as the shortcuts of the visual search experiment overshadowing the actual effect of the different expressions in the tasks. However, not all visual features are confounds. The visual features described above can either be a confound or a part of the mechanism defining an expression, making them defining features of the expression. To find if the visual features causing these attentional effects are defining features or confounds, they need to be evaluated. Seeing that any variety between conditions is a valid possibility to explain the behavioural differences between these conditions, visual features are in need of being thoroughly investigated. Here we might find that some visual features like the V-shape of the eyebrows do contain useful information on the expression, while others, for example the visibility of teeth, are correctly classified as confounds. To summarize: specific visual features may give a more sufficient explanation for the attentional effect found in studies on the superiority effect when compared to the emotional label. Therefore the facial expression can be described with the use of defining features.

The question that comes to mind is: What are the different defining stimuli of happy and angry faces? Stimulus properties that could be responsible for the attentional effects of emotional faces might be the spatial frequency (SF) content of the image and edge detection (HOG; histogram oriented gradients) (Chen, Chen, Chi, & Fu, 2014; Déniz, Bueno, Salido, & De la Torre, 2011; Halit, De Haan, Schyns, & Johnson, 2006). Therefore, this paper focuses on HOG and SF features for both happy and angry faces. Section 2 contains literature studies into the stimuli and visual features used in this field, because they are at the base of the argument. Next, section 3 and 4 explain the current approach and methods used for the experiments. Section 5 and 6 give an overview of the visual features found and whether they were generalizable. Finally, in the conclusion this paper defines the happy and angry faces with visual features in the different datasets and explains possible implications of this research for the discussion on the superiority effect and the field of artificial intelligence in general.

2 Visual features

Most of the papers in this field use a visual search task[1] to test their hypotheses. However, some difficulties arise when one is not careful with the use of face stimuli in these visual search tasks. Two important reviews of this literature are Frischen et al. (2008) and Becker et al. (2011). Both of these papers reviewed visual search tasks and concluded that previous research into the superiority effect was often flawed in its methodology. Subsequently, Frischen constructed three criteria for further research: set size, content of the distractors controlled for, and vigilance for the influence of search strategies. Becker et al. (2011) added two criteria to this list: eliminating low level visual features and different identities for the face stimuli. While reviewing these problems in the existing literature, Becker et al. (2011) came to the conclusion that there

[1] See Frischen et al. (2008) and Becker et al. (2011) for elaborate explanations of the visual search tasks.

was little unconfounded evidence for either ASE or HSE. As explained above, Becker et al. (2011) claimed that the issues in previous papers are due to stimulus problems and flaws in the design. By thoroughly describing the set criteria they could design experiments that where valid, avoid these pitfalls, and make reasonable claims on the superiority effect.

After conducting these experiments, Becker et al. (2011) concluded from their evidence that HSE was more prominent than ASE in the visual search tasks. If his approach was correct this found HSE should have been a reasonable claim on the superiority effect. However, by recreating Becker's experiment Savage, Becker, and Lipp (2016) found that these effects could be traced back to only three faces and when these faces were replaced, an ASE could also be discovered. Moreover, from these results it is still unclear if happy or angry expressions are the most salient. The reason for this is that the visual search tasks on ASE and HSE stimulus selection appears to plays a critical role for which effect will emerge. Even with faces in the same test-set this can lead to contradicting results (Savage et al., 2016). These results seem to indicate that, even though researchers focus specifically on trying to control for confounds in the stimuli and flaws in the design, they still appear. This intuition is supported by Becker et al. (2011) themselves by noting that visual search tasks like these are prone to a host of potential confounds which lead to misleading conclusions.

Although this thesis does agree with Becker et al. (2011) that visual features cause problems for the interpretation of the face stimuli, controlling the experiments for all visual features seems reductive. Removing the confounds from the data is nearly impossible and it could also lead to the loss of important visual data. The criterium described by Becker et al. (2011) to eliminate low level visual features does not appear to be useful. Instead of filtering these visual features from the stimuli, it might be interesting to look at how these visual features play a part in expression recognition. To this end it is useful to look at another review of the superiority effect also done by Savage et al. (2013). According to these authors, the visual features described in the papers above can be traced back to the complex face stimuli used in those papers. Instead of measuring the mechanism behind the expression, the papers researching the superiority effect sometimes measure the visual features in the face stimuli that provide shortcuts for the visual search task. Thus, instead of excluding the visual features that cause attentional effects, these features have to be studied in order to differentiate the confounds from the defining features.

A face can be represented by different attributes each consisting of specific visual features. Some of the visual features can help people and algorithms more efficiently distinguish the expression (label) of a face. These features can include the "V"-shape of eyebrows or the presence of teeth but also luminance in areas of the face can contain specific visual features. The mentioned visual features are either part of the mechanism defining the expression or can help predict the expression but do not define it. Note that in the papers researching the superiority effect, this second category of visual features caused attentional effects, which confounded the evidence. For example, in the paper of Hansen and Hansen (1988) where the contrast areas caused misleading conclusions in the research (Purcell, Stewart, & Skov, 1996). This was reinforced by the literature reviews discussed earlier and by a different study into this debate which found that HSE and ASE could be caused by emotional related and emotional unrelated visual features (Savage et al., 2013). The visual features in these studies that confounded the results often arise from differences in contrast information. However, contrast information could also contain defining features.

Spatial frequency is an attribute of visual information containing contrast information of an image. Numerous studies into the relation of SF and facial recognition have been done (Gao & Maurer, 2011; Goren & Wilson, 2006; Jeantet, Caharel, Schwan, Lighezzolo-Alnot, & Laprevote, 2018). However, in review on this literature Jeantet et al. (2018) found several limitations within these studies, partially due to design and the complexity of the face stimuli used. Therefore, though indications are present hinting at the importance of SF features, more research has to be done to generalise conclusions on SF to faces and specific expressions. Due to the sheer interest and research into SF content regarding facial and expression recognition, this might be an interesting visual attribute to investigate at a high level of detail. Another promising visual attribute for expression recognition is HOG information, which reflects the structural properties of an image. HOG information has been found particularly useful in human expression recognition due to its focus on object orientation (Dalal & Triggs, 2005) and received much attention in the field of computer science due to its success in expression recognition (Carcagnì, Del Coco, Leo, & Distante, 2015). Therefore, in this thesis the focus is specifically on HOG and SF.

Histograms of oriented gradients are feature descriptors successfully used in face recognition. HOG consists of a histogram which tracks the counts of gradients with certain orientations for a part of an image. In the HOG feature extraction process, an image is first divided into $N \times N$ pixel blocks. In these blocks, the orientation of each pixel is recorded and combined into a histogram. Because of this, HOG features are highly spatially specific. Furthermore, an image can be described by its changes in light and dark across space. Specifically, spatial frequency refers to the number of dark-light cycles in a given unit of space. Each spatial frequency of an image also has an orientation: the angle across which the dark-light alternation occurs within the image. Crucially, the SF

features used in this paper only consist of the magnitude of the frequency component. No phase information is used in the analysis. As such, the SF features used here only convey global contrast information. Furthermore, to make sure no residual local information remains in the SF data, this data has been down-sampled by taking the sums of the magnitudes within a given frequency and orientation range. After having inferred which attributes of the face can be useful for defining happy and angry faces, namely HOG and SF, it is now possible to re-evaluate the research question and divide it in four smaller questions which are easier to answer:

1. What HOG features are defining for happy faces?
2. What HOG features are defining for angry faces?
3. What SF features are defining for happy faces?
4. What SF features are defining for angry faces?

3 Machine learning

The machine learning algorithm explained here will be used in the experiment in order to answer the four subquestions mentioned in the previous section. By using machine learning models trained on these features it is possible to test within the mentioned attributes (HOG and SF) which specific features are helpful for identifying the expressions in the datasets. To map specific features directly to the success of the model, two preconditions are required. First, the model has to be reasonably good at predicting the correct expression. Second, the link between the features and how the model uses those features to identify the expression is fairly straightforward. In this process it is important to constrain the machine learning models to a smaller number of features to make sure it is possible to easily decode the features that are important. Through using attributes that are linked to expression recognition in humans (SF) and in computer vison (HOG) and using machine learning to find specific features that achieve good results at expression recognition in the face stimuli, we made a model for recognising happy and angry faces using only the most relevant features within the attribute. To summarize: through the use of machine learning models the aim of this paper is to better understand and aid the research into human expression recognition, specifically happy and angry expressions. This is attempted by constructing a model of expression recognition using these features.

This machine learning model can lead to new insights into human expression recognition. Here an AI-technique is used to perform a narrow but intelligent task when done by humans. This kind of use of machine learning models falls under the branch of Weak AI. In the past, this approach has given new insight into different cognitive processes like facial recognition. For example, by using Hidden Markov Models (HMM) Chuk, Chan, and Hsiao (2014) found promising information on eye movement during facial recognition, hinting to the significance of different types of processing in facial recognition. Since the process of expression recognition is fairly similar, this thesis is optimistic about the possibility of translating the result found in this paper to models of human expression recognition in further research.

To make such a machine learning model we have to follow certain steps: feature extraction, data splitting, feature selection and cross-validation. For all the datasets, features are extracted from each stimulus. The two attributes of features researched in this thesis are SF and HOG. The SF features are represented by the Fourier Magnitude Spectrum. Here, the Fourier Magnitude Spectrum is divided into 16 orientations and 24 spatial frequencies. This entails that for each of the 16 orientations, the stimulus is divided into 24 spatial frequencies where the information of the contrast energy is saved. In total, this amounts to 384 features. For the HOG features, the stimuli are divided into 400 10×10 px blocks. For each block, 9 orientations are evaluated. Then, the gradient information is saved for each orientation.

All further steps have to be repeated for each combination of dataset (Nimstim, Karolinska or Radboud), emotion (happy or angry) and attribute (HOG or SF). This will result in 12 models. To illustrate the process, we will only describe one of these combinations: Nimstim HOG happy. The process for the other combinations is identical. In the methods section, the procedure is explained in more detail. After extracting the HOG features from the Nimstim set, the data was divided into 10 partitions of exactly the same balance between neutral and happy or neutral and angry faces, see Stuit, Paffen, and van der Stigchel (Under review). These partitions are combined into 10 folds. Within each fold, 9 partitions are make up the feature selection set and the remaining partition is a test set. The folds differ in which partitions fall in the feature selection set and which partition is the validation set. Each feature selection set is divided into 70% training set and 30% validation set.

In the first step of the feature selection, the features are sorted by chi-squared analysis of variance score, indicating their suspected importance. In other words, the features are ranked on the difference between the two classes which are tested (neutral vs happy). These chi-squared values are also used to determine the maximum amount of features to be included in the model: 25% of the sum of the total chi-squared values. Second, based on this ranking, the best cluster of features is selected for a first iteration of the wrapper model. The wrapper model entails that combinations of features are tested on their joint ability to decode classes in the dataset. The selected cluster of features is trained on the train set and tested on the validation set. Then, the features within the cluster are reordered on the classification performance

of this iteration (F1 macro score). Subsequently, the best features are selected for inclusion in the intermediate wrapper model. The features that were not included in this selection are added back into the initial ranking with the ordering based on their added value still intact. Next, all these steps are repeated until the predetermined maximum amount of features is included and the wrapper model of this fold is complete. The final step of feature selection is that all features in the intermediate model are re-evaluated on their usefulness in the model. The features that are not contributing to the model are deleted from it. Because 10 fold cross-validation is used in this thesis, the feature selection is performed for each of the 10 folds. This results in 10 complete wrapper models. Each model is then tested on the test set of their own fold resulting in 10 performances. In the remainder of this thesis the average of these 10 folds are taken to represent the model of the given emotion, attribute, and dataset, in this example the Nimstim HOG happy model.

4 Methods

The experiments in this thesis are all without participants and partly use the same methods as in Stuit et al. (Under review). The apparatus was identical to this article and will therefore not be discussed.

4.1 Stimuli

The stimuli consist of photographs of faces with differing expressions: angry, happy and neutral. Three datasets with such faces were used in this thesis. First, the Nimstim (N) dataset containing 81 faces. Secondly, the Karolinska (K) dataset containing 140 faces. Finally, the Radboud (R) dataset containing 114 faces with frontal gaze. From the Nimstim dataset, exuberant stimuli and stimuli with an open mouth were excluded due to the intenseness of these expressions. This decision was made to keep the same level of expression intensity between different expressions and different datasets (to avoid comparing happy with very happy).

4.2 Measurement

The measures of the experiments are the performance of the models on the different datasets. This is measured in accuracy (percentage of faces classified correctly) per fold. The error bars used are the standard error of the mean. A two-sided one sample t-test was used for the cross-validation matrices. All effects were considered significant above $\alpha = 0,05$.

4.3 Procedure

The algorithm described in the machine learning chapter of this thesis was used for each combination of dataset, emotion label and attribute. This resulted in 12 models that were numbered as displayed below. The models had to perform a classification task, either classifying neutral and happy faces or angry and neutral faces. Each of these models was first trained and tested on its respective dataset leading to the base performance of the models. From these models the most important features were extracted and visualised. Then, for each combination of attribute and emotional label the models within this group were tested on the two datasets they were not trained on. This is the cross validation. For example, for the combination Happy HOG, model 1 was tested on dataset R and K, model 5 was tested on the N and K datasets, and model 9 was tested on dataset N and R. This resulted into four cross-validation matrices (figure 6). See the table below for the complete overview of the models.

Dataset	**Attribute & emotional label**			
	Ha HOG	Ha SF	An HOG	An SF
N	1	2	3	4
R	5	6	7	8
K	9	10	11	12

5 Results

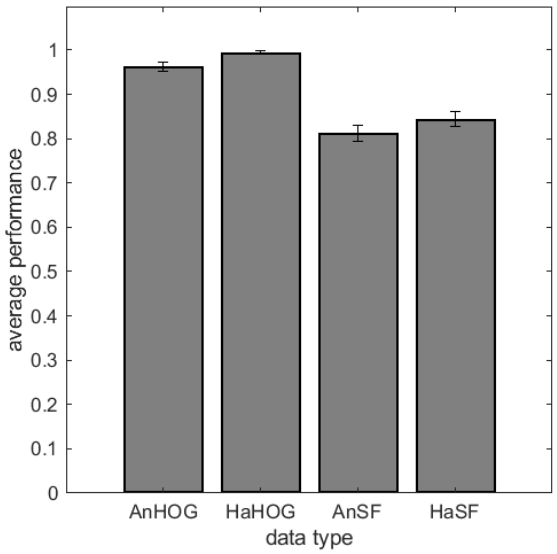

Figure 1: Visualisation of the average performance of the different data types, here the data types (x-axis) represent a combination of attribute and expression. Note that all averages are above the 0.5 mark. The error-bars are the standard error of the mean (SEM).

All 12 trained models achieved above chance (0.5) performance on the test set when the test set was from the same dataset the models were trained on.

When looking at the HOG models, the R HOG happy model stands out because of its relatively small error bar compared to the other models. *Figure 2* shows that all models perform above chance, corresponding with the average values for the HOG happy and HOG angry models in *figure 1*. This is also the case for the SF angry and SF happy models in *figure 3*. Lastly, these figures show that all the models did not just perform above chance, but most even achieved an accuracy above 0.8, which can be seen in the averages per emotion in *figure 1*.

Note that all the HOG happy expression models resulted in features around the mouth to be found important, see *figure 4*. For the HOG angry expression models the space between the eyes resulted in important features for all datasets. However, in the Nimstim and Karolinska datasets the eyes also held important features for the HOG angry models. This was not the case in the Radboud dataset. Interestingly, the HOG happy model trained on the Radboud dataset selected no features in the area of the teeth or the area of the lips and the teeth, even though historically this has been an important feature for the recognition of happy faces. The Karolinska dataset did have important features located on and surrounding the teeth. The mean amount of features used for each model are: N An = 16.7600, N Ha = 12.1200, K An = 16.2000, K Ha = 12.0000, R An = 12.0400 and R Ha = 11.2800.

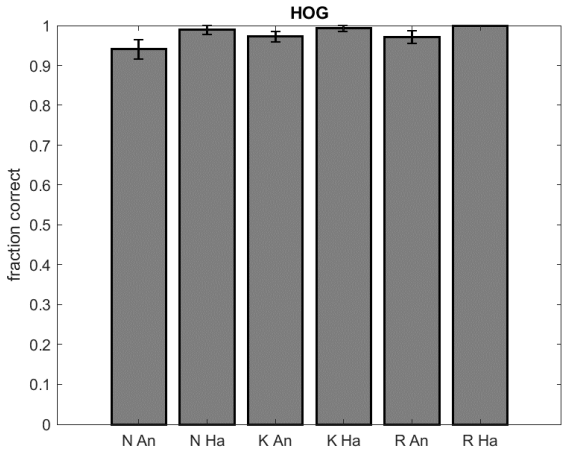

Figure 2: Performance of the different models trained on HOG features. On the x-axis the data type which is a combination of dataset and expression and on the y-axis the performance. Note that all models perform above chance. The error-bars are the SEM.

In the SF happy models located at the right of *figure 5*, the features in these models seem to have a slight horizontal orientation, most evident in the R SF happy model. In the SF angry models shown on the left, the features seem to have a diagonal orientation. In the N SF angry model the features seem to have an orientation diagonal to the left while in the R SF angry model the features have an opposite orientation to this (diagonal to the right). Within the features of the K SF angry model, both orientations are present (diagonal left and diagonal right). The mean amount of features used for each model are: N An = 4.3600, N Ha = 4.0000, K An = 3.0000, K Ha = 3.0000, R An = 3.9600, R Ha = 3.0800.

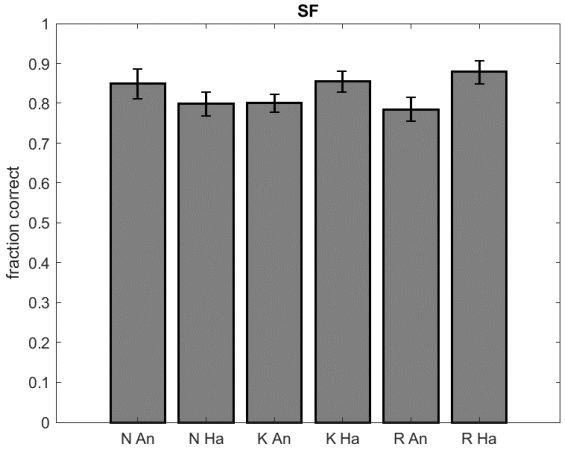

Figure 3: The models trained on SF features, performance is shown on the y-axis and data type (data set plus expression) is shown on the x-axis. The error-bars represent the SEM and all models perform above chance.

To test if the models generalise well across datasets, all models were tested on the two other datasets. Note that generalisation of the models implies the presence of defining features, while no generalisation implies the presence of confounds. *Figures 4* and *5* indicate that in the HOG features patterns are present while the SF features do not result in clear patterns. Similar effects can be observed in the confusion matrices visualised in *figure 6*. In the HOG angry models (top left), the model trained on the K dataset performed significantly well on the N and R datasets. It seems that a model for angry faces based on HOG features can be made that generalises well to other datasets. In the HOG happy models (top right), the N HOG happy model performed significantly well on both the R and K datasets. Here again, it seem that HOG features can generalise well to other datasets, in this case for happy faces. The SF models were less successful for both happy and angry faces: no model was present that could classify the other two datasets significantly better than chance. This suggests that SF models do not generalise well across different datasets. Another noteworthy result from the cross-validation can

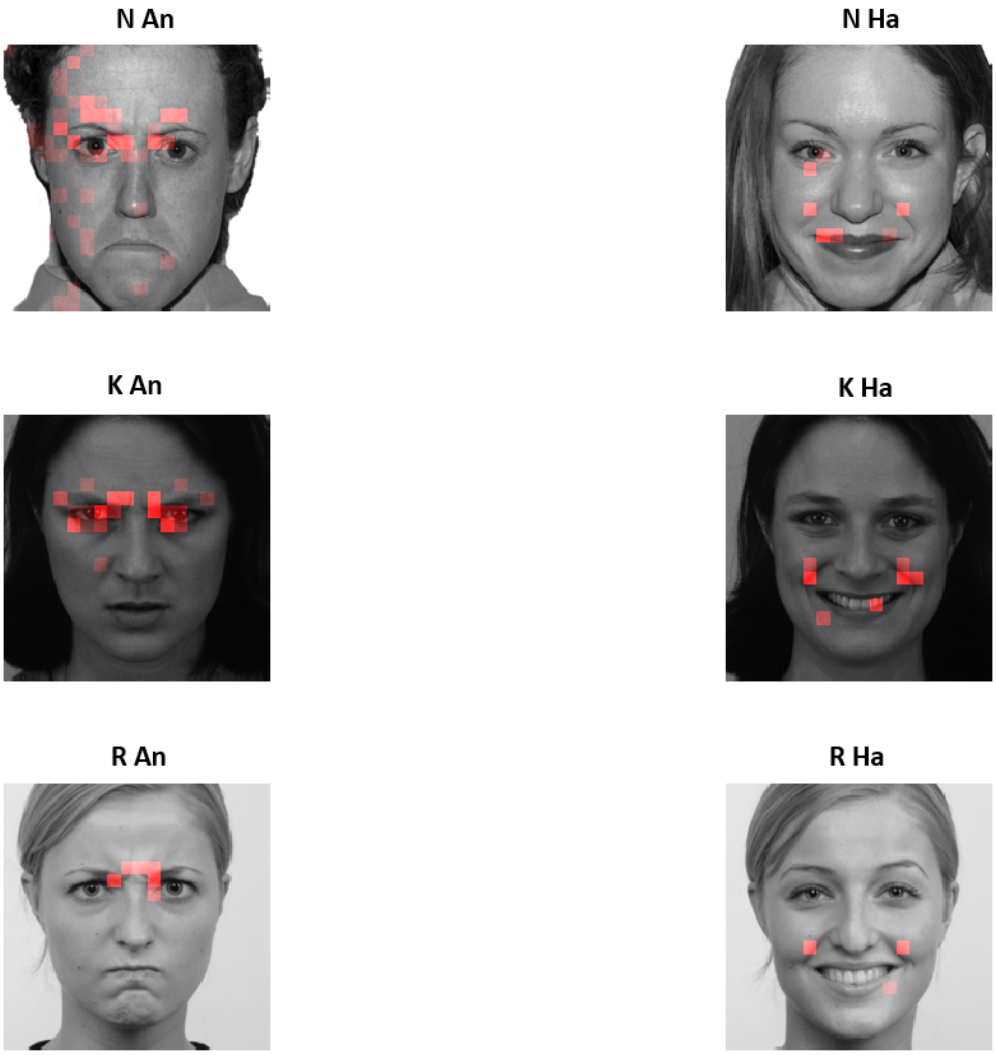

Figure 4: Here the most important HOG features per model are shown mapped onto faces. On the right side the angry expressions are shown and on the left side the happy. Note that the angry expression models seem to use more features than the happy expression models.

be found in the R HOG angry model. This model had significant results on both other datasets, but while the model performed well on the K dataset, it performed significantly worse than chance at classifying the N dataset. This effect is seen even stronger for the R HOG happy model which was significantly worse at classifying both the N and the K datasets. Lastly, only one dataset from the entire cross-validation was significantly well classified by both models not trained on this dataset. This was the R dataset for the HOG angry models. To summarise: the main result found in *figure 6* is that HOG features seem to generalise well, hinting at the presence of defining features, while SF features do not generalise well, hinting to the presence of confounds.

6 Discussion

The research question of this thesis is: what different features define happy and angry faces? To answer this question, it is important to first investigate how difficult classifying happy or angry faces is, based on HOG and SF features within each dataset. For this purpose, 12 different models were trained on the visual features to classify these expressions. All of the 12 models performed substantially above chance, see *figure 2* and *figure 3*. Thus, the conclusion can be drawn from the averages in *figure 1* that both HOG and SF features can be used to accurately classify and therefore define angry and happy faces within their own datasets.

Secondly, if the features within an attribute are part of the mechanism defining an expression, similar best per-

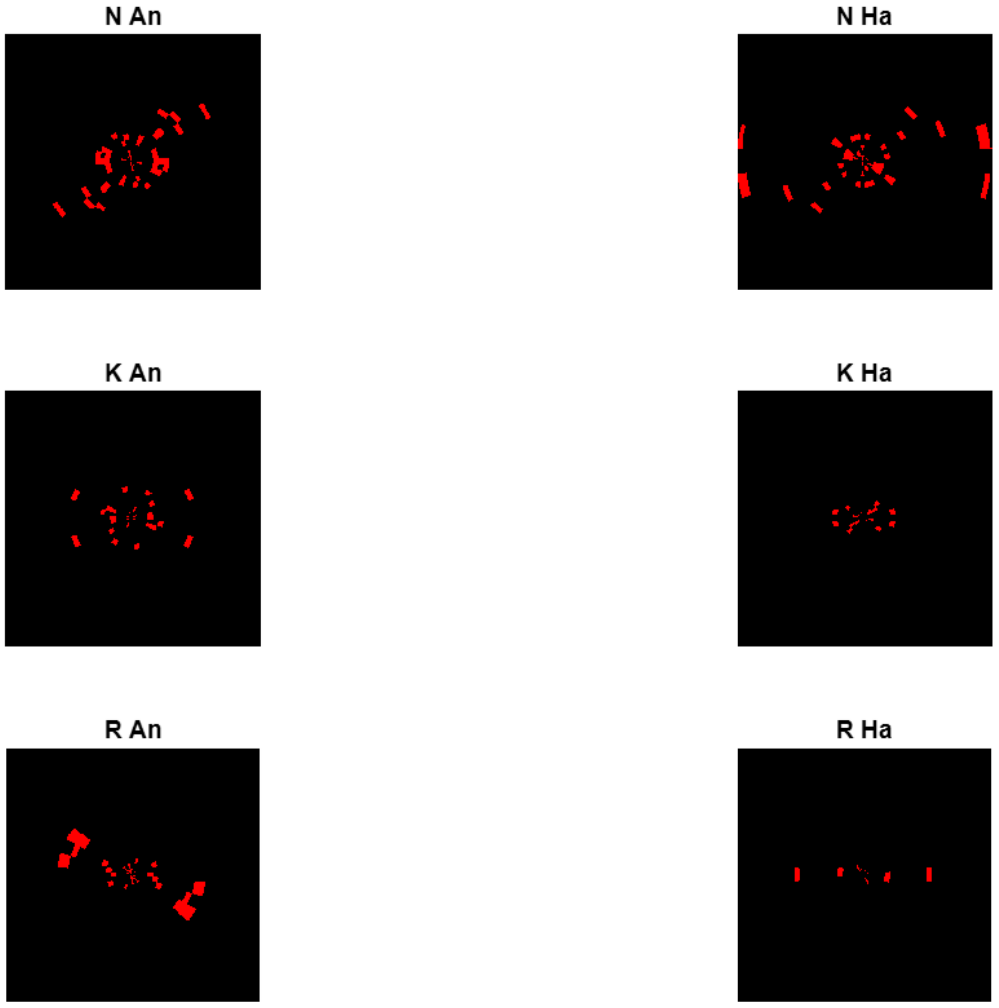

Figure 5: : These pictures are the SF feature maps. Here the most important SF features are visualised. On the right side are the angry expression models and on the left side are the happy expression models. Note that there is no clear pattern in either the angry models or the happy models.

forming features should be chosen for the different models. Otherwise, these features are likely to be confounds. The best performing features of the models were visualised in feature maps for both HOG and SF, see *figure 4* and *figure 5*. Analysis of the HOG features showed that in both the models for angry faces and the models for happy faces, familiar patterns were found. For the HOG angry models the features that decoded information between the eyebrows, the upper nasal area, was found important in all models. However, HOG information found on the eyes seems only important for the models trained on the N and K datasets, see *figure 4*. This implies that the information found in these datasets might differ from the R dataset, where the same HOG information on the eyes was not found to be among the most important features. Furthermore, from these patterns the prediction can be made that specific HOG features (between the eyebrows) can help define angry faces. Similar effects were found for the HOG happy models; here also a specific pattern of features was found. HOG information around the mouth area was important in all three models. For the HOG happy models, the conclusion can be drawn that specific HOG features (around the mouth) can help define a happy face in these datasets. Different results were found by analysing the SF feature maps: no clear patterns were visible for both the SF angry models and the SF happy models. Due to the lack of patterns in the SF models, it is unlikely that the features found in the different SF models hold information that could help define happy or angry faces. Therefore the intuition in this paper is that the SF features without spatial information are confounds. If the SF features are confounds, then these features should not generalise well across different datasets.

Finally, the question has to be answered if the fea-

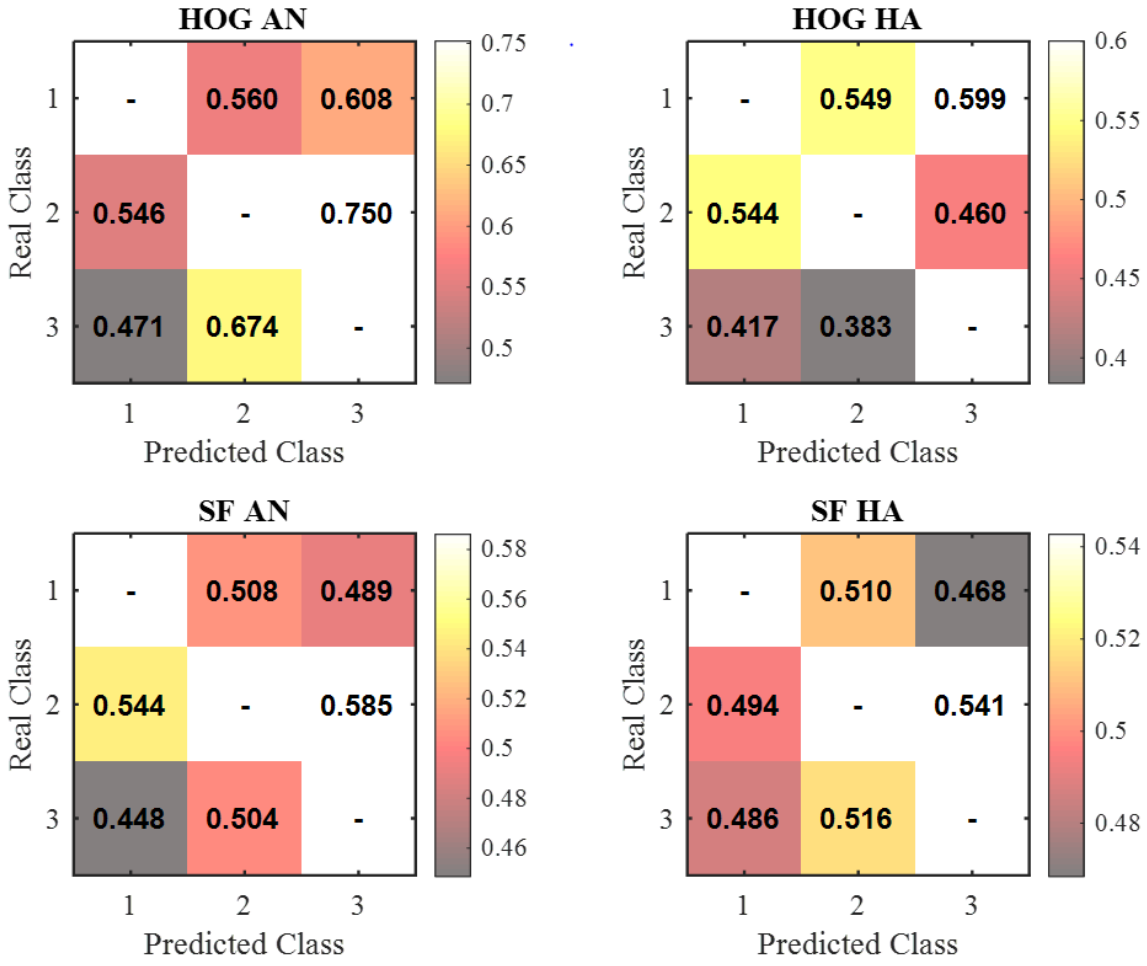

Figure 6: Confusion matrices of the cross-validation of the models. The y-axis are the labels for the dataset that the model was trained on and the x-axis the label of the dataset the model was tested on. The results of the models tested on the same dataset that they were trained on were excluded. The labels used here are: 1 = Nimstim (N), 2 = Karolinska (K) and 3= Radboud (R).

tures in these models can be generalised and thus if the features are confounds or part of the mechanism defining the expressions. For this purpose the cross-validation[2] was performed. From the results visualised in *figure 6*, the conclusion can be drawn that HOG features can generalise well for happy and angry faces. Also, as described above, patterns were found in the HOG feature maps per emotion. From these results the conclusion can be drawn that HOG features are useful as a mechanism for defining happy and angry faces. On the other hand, the SF features did not generalise well for either happy or angry faces. Also, in the SF feature maps no patterns were found. Therefore, these SF features without spatial information are not defining features but instead they are confounding variables.

The results mentioned above are the main results found in the experiment. Nevertheless, some other important results were found as well. First the feature maps in *figure 4* and *figure 6* will be discussed. Interesting was that in the HOG happy models, faces with an exposed open mouth did not lead to many important features being identified in and around the teeth, see *figure 4*. The presence of teeth has been found important for the recognition of happy faces due to its luminance content and therefore we could expect that the edge information on teeth and the place around the teeth would lead to interesting HOG features. However, this was not found by the models. Another surprising aspect found in the feature maps is the horizontal orientation of the SF happy models. This orientation is best seen in the features of the R SF happy model in *figure 5*. SF features with a horizontal orientation have been found useful for emotion recognition specifically in happy faces (Huynh & Balas, 2014). However, this preference has been found relative

[2]More details on the analyses of the cross-validation can be found in appendix A.

to other aspects of the face and thus can be dataset specific. The SF angry models also indicate an orientational preference; these models have diagonal orientations. The N SF angry model has a diagonal orientation to the left while the R SF angry model has a diagonal orientation to the right, opposite to the N SF angry model. For the features found in the K SF angry model, both orientations are present. These differences could be explained by the luminance information of the stimuli. For the different datasets the light source used by the photographer could be differently oriented. This can influence the luminance information in the stimuli. SF is highly sensitive to this kind of information and the presence of more light in the picture might influence this and could explain the differences.

The cross-validation also resulted in some additional interesting results, see *figure 6*. Firstly: the significantly worse than chance performance of some of the HOG feature models. Both the R HOG happy model and the R HOG angry model were significantly worse than chance on at least one other dataset. The R HOG happy model performed significantly badly at both the K and the N dataset, while R HOG angry performed significantly badly on the N dataset, but did generalise well to the K dataset. Thus, while this is not true for the other datasets, all HOG models trained on the R dataset were significant. This implies that models trained on the R dataset have a higher chance of finding features important for classification, however these models also have a higher chance that these features are disadvantages. This seems to indicate that something interesting is decoded within the R dataset. When looking at the HOG feature maps (*figure 4*) it is evident that the models trained on the R dataset use a smaller amount of features compared to the other HOG models with the R HOG happy model consisting of the lowest amount of features. A smaller number of features means that individual features are likely to have a stronger influence on the classification. Possibly this low amount of features is responsible for the significant but often bad performance of the R HOG models. Secondly, only one dataset seemed to be significantly well performed on by both models trained on the other datasets. The dataset in question was the R dataset for the HOG angry models. Both the N HOG angry and K HOG angry models had similar features selected as the model trained on the R dataset. These results are not surprising since all the important features for the model trained on this dataset, R HOG angry, are also found important for the other HOG angry models. This could be caused by the R HOG angry model only using a small amount of features. These features were al located between the eyebrows implying that the most useful HOG information is only located between the eyebrows for the R dataset. This provides further evidence for the hypothesis that the low amount of features selected by the HOG models might be related to the interesting effects found in the Radboud dataset.

7 Conclusion

In this thesis, the aim was to find defining visual features for happy and angry faces. For this purpose, first a literature study was done to target specific visual features that could be of use. The conclusion was that contrast information, especially HOG and SF, could be useful attributes for the search of specific features. With this information the research question was divided into 4 smaller questions:

1. What HOG features are defining for happy faces?
2. What HOG features are defining for angry faces?
3. What SF features are defining for happy faces?
4. What SF features are defining for angry faces?

In order to answer these questions a machine learning algorithm was used to train models on these features with the task of classifying either angry (angry vs neutral) or happy (happy vs neutral) faces. The experiment that followed contained 12 models trained on either HOG or SF features tested across different datasets. The results from this experiment show that HOG features can be defining features for happy and angry faces across our datasets. These features can be used to optimise the classification process for these emotions and better understand the essence of happy and angry faces. However, SF features without spatial content did not generalise well for happy and angry faces. This information appeared to be very dataset specific and therefore is a confounding variable for defining happy or angry faces. Now the research questions can be answered. A HOG features model was found that can define happy faces across different datasets, namely the Nimstim HOG happy model. A generalizable HOG feature model for angry faces was also found, namely the Karolinska HOG angry model. No SF feature models were found that can define happy or angry faces in different datasets.

The goal of artificial intelligence is to achieve human-level intelligence or even surpass it. Intelligent tasks often are difficult for computers but come naturally to humans. For example, verbal communication has been a cornerstone in the field of AI for decades because it is a typical example of an intelligent task. Using and recognising facial expressions are intelligent tasks as well because they have a similar function as a communication system and in social interaction. The importance of this research for the field of AI is that it can help in understanding human expression recognition. Finding the defining features and confounds supports this research. For example, the confounding variables that cause difficulties in the debate on the superiority effect can be studied to see which effects are caused by confounds and which are caused by the ex-

pressions themselves. Generally speaking, the goal of this thesis is to apply machine learning, an AI technique, to aid research that attempts to gain a better understanding of how a specific part of human intelligence works.

Future research could look into the comparison of eye-tracking data and the models used in this study. If similarities are found between eye-tracking data and the defining features from this thesis, the theory that these features define an expression will have additional support. One study already showing interesting data is Schurgin et al. (2014) which found the importance of the upper nasal area for the anger expression. This area was found to be part of the defining features for angry faces in this thesis. Furthermore, it would be interesting to look at more attributes that may be important for defining angry and happy faces and to investigate if these features are confounds or defining features. Finally, additional expressions such as fear or disgust could be studied to investigate if HOG can also successfully define those expressions.

References

Becker, D. V., Anderson, U. S., Mortensen, C. R., Neufeld, S. L., & Neel, R. (2011). The face in the crowd effect unconfounded: Happy faces, not angry faces, are more efficiently detected in single-and multiple-target visual search tasks. *Journal of Experimental Psychology: General*, *140*(4), 637.

Carcagnì, P., Del Coco, M., Leo, M., & Distante, C. (2015). Facial expression recognition and histograms of oriented gradients: A comprehensive study. *SpringerPlus*, *4*(1), 645.

Chen, J., Chen, Z., Chi, Z., & Fu, H. (2014). Facial expression recognition based on facial components detection and hog features, In *International workshops on electrical and computer engineering subfields*.

Chuk, T., Chan, A. B., & Hsiao, J. H. (2014). Understanding eye movements in face recognition using hidden markov models. *Journal of vision*, *14*(11), 8–8.

Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection, In *2005 ieee computer society conference on computer vision and pattern recognition (cvpr'05)*. IEEE.

Déniz, O., Bueno, G., Salido, J., & De la Torre, F. (2011). Face recognition using histograms of oriented gradients. *Pattern recognition letters*, *32*(12), 1598–1603.

Frischen, A., Eastwood, J. D., & Smilek, D. (2008). Visual search for faces with emotional expressions. *Psychological bulletin*, *134*(5), 662.

Frith, C. (2009). Role of facial expressions in social interactions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1535), 3453–3458.

Gao, X., & Maurer, D. (2011). A comparison of spatial frequency tuning for the recognition of facial identity and facial expressions in adults and children. *Vision Research*, *51*(5), 508–519.

Goren, D., & Wilson, H. R. (2006). Quantifying facial expression recognition across viewing conditions. *Vision Research*, *46*(8-9), 1253–1262.

Halit, H., De Haan, M., Schyns, P., & Johnson, M. (2006). Is high-spatial frequency information used in the early stages of face detection? *Brain Research*, *1117*(1), 154–161.

Hansen, C. H., & Hansen, R. D. (1988). Finding the face in the crowd: An anger superiority effect. *Journal of personality and social psychology*, *54*(6), 917.

Huynh, C. M., & Balas, B. (2014). Emotion recognition (sometimes) depends on horizontal orientations. *Attention, Perception, & Psychophysics*, *76*(5), 1381–1392.

Jeantet, C., Caharel, S., Schwan, R., Lighezzolo-Alnot, J., & Laprevote, V. (2018). Factors influencing spatial frequency extraction in faces: A review. *Neuroscience & Biobehavioral Reviews*, *93*, 123–138.

Lee, D. H., Susskind, J. M., & Anderson, A. K. (2013). Social transmission of the sensory benefits of eye widening in fear expressions. *Psychological science*, *24*(6), 957–965.

Pantic, M., & Rothkrantz, L. J. M. (2000). Automatic analysis of facial expressions: The state of the art. *IEEE Transactions on pattern analysis and machine intelligence*, *22*(12), 1424–1445.

Purcell, D. G., Stewart, A. L., & Skov, R. B. (1996). It takes a confounded face to pop out of a crowd. *Perception*, *25*(9), 1091–1108.

Savage, R. A., Becker, S. I., & Lipp, O. V. (2016). Visual search for emotional expressions: Effect of stimulus set on anger and happiness superiority. *Cognition and Emotion*, *30*(4), 713–730.

Savage, R. A., Lipp, O. V., Craig, B. M., Becker, S. I., & Horstmann, G. (2013). In search of the emotional face: Anger versus happiness superiority in visual search. *Emotion*, *13*(4), 758.

Schurgin, M., Nelson, J., Iida, S., Ohira, H., Chiao, J., & Franconeri, S. (2014). Eye movements during emotion recognition in faces. *Journal of vision*, *14*(13), 14–14.

Stuit, S., Paffen, C. L. E., & van der Stigchel, S. (Under review). Introducing the prototypical stimulus characteristics toolbox: Protosc.

Vuilleumier, P., & Schwartz, S. (2001). Emotional facial expressions capture attention. *Neurology*, *56*(2), 153–158.

Appendix A Additonal statistics

A.1 Significance

HOG AN		
	0,1127	0,0022
0,0120		0,0000
0,0268	0,0000	

HOG HA		
	0,0024	0,0000
0,0345		0,4580
0,0027	0,0000	

SF AN		
	0,0410	0,2238
0,1378		0,0132
0,0629	0,2472	

SF HA		
	0,3316	0,8682
0,7204		0,0309
0,4026	0,2393	

Table 1: p-values

HOG AN		
	2,1499	3,4693
2,906		15,5938
-2,0194	13,7858	

HOG HA		
	4,4624	9,5667
2,1468		-2,8244
-4,9249	-19,6362	

SF AN		
	0,3046	-0,6774
1,5905		4,1467
-2,6374	0,1881	

SF HA		
	0,3504	-1,5266
-0,3572		1,8541
-0,5699	1,1901	

Table 2: t-values

A.2 Interpretation

In this appendix the results that lead to the interpretation of figure 6 are explained in more detail. This paragraph contains in text the information found in appendix A.1. For the HOG Angry models cross-validation, see figure 6 upper left corner, this resulted in 4 values which were significant above chance (the model trained on N tested on R [x=3, y=1], p = 0.0022, t = 3.4693, the model trained on K tested on N [x=1, y=2], p = 0.0120, t = 2.9060, the model trained on K tested on R [x=3, y=2], p < 0.0001, t = 15.5938, and the model trained on R tested on K [x=2, y=3], p < 0.0001, t = 13.7858), one value that was significant below chance (the model trained on R tested on N [x=1, y=3], p = 0.0268, t = -2.0194) and one value which was not significant (the model trained on N tested on K [x=2, y=1]). Secondly, the HOG Happy models cross-validation, see figure 6 upper right corner, resulted in 3 values which were significant above chance (the model trained on N tested on K [x=2, y=1], p = 0.024, t = 4.4624, the model trained on R tested on N [x=3, y=1], p < 0.0001, t = 9.5667, and the model trained on K tested on N [x=1, y=2], p = 0.0345, t = 2.1468), two values that were significant below chance (the model trained on R tested on N [x=1, y=3], p = 0.0027, t = -4.9249 and the model trained on R tested on K [x=2, y=3], p < 0.0001, t = -19.6362) and one value which was not significant (the model trained on K tested on R [x=3, y=2]). Thirdly, the SF Angry models cross-validation, figure 6 lower left corner, resulted in 2 significant values above chance (the model trained on N tested on K [x=2, y=1], p = 0.0410, t = 0.3046 and the model trained on K tested on R [x=3, y=2], p = 0.0132, t = 4.1467). The other values were not significant. Lastly, the SF Happy models cross-validation, see figure 6 right below, had only one significant value. This value was significant above chance ([x=3, y=2], p = 0.0309, t = 1.8541).