Classification of Emotional States with Facial Action Units

Ajani Blackwood
Electrical and Computer Engineering
University of Connecticut
Storrs, CT 06269
ajani.blackwood@uconn.edu

Atharva Manjrekar
Electrical and Computer Engineering
University of Connecticut
Storrs, CT 06269
atharva.manjrekar@uconn.edu

Abstract

The Covid-19 pandemic has necessitated the wearing of face masks in public which undermines the ability to discern emotional cues from the entire face. Face masks typically cover regions around the nose, cheeks, and mouth, which undermines the visual dimension of face-to-face communication. Current studies use a variety of approaches to quantify the importance of facial action units in determining emotional expression, but none specifically focus exclusively on the top portion of the face. Here we present a method that uses the Facial Action Coding System (FACs) to quantify the performance of muscle movements in the top portion of the face using time averaged facial action units (AUs) for classifying happiness and sadness. Results show that for both men and women, high activation of the muscles *depressor glabellae, depressor supercilii, currugator supercilii* (together denoted as AU4) is a strong predictor of sadness. High activation of the muscles *orbicularis oculi and pars orbitalis* (together denoted as AU6) is the strongest predictor of happiness.

1 Introduction

The motivation behind our work stems from the desire to understand how certain facial action units predominantly found over the nasal tip can be used to classify emotions. To solve this, we will use the Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) Facial Landmark Tracking dataset. There are 8 emotions with two labeled intensity levels (normal and strong). In total there are 64 recognized facial action units, but our dataset only considers 17 of them. Specifically, we are interested in the facial action unit variables (AUs) which represent different facial muscle movements with intensities ranging from 0 to 5. An intensity value of 0 represents no muscle contraction whereas 5 represents full muscle contraction. We will consider all AUs that are over the nasal tip as top face AUs and the remaining as bottom face AUs. We aim to classify two emotions (happiness and sadness) for both sexes as gender differences were not emphasized in Ekman and Friesen's original work in developing the FACs database [1].

2 Related Work

Current research surrounding FACs employs HMM models for predicting emotional composition of a facial expression using AUs [2]. Here, all AUs that represent the entire face are used. Other research involves automatic lip tracking and AU classification using two step active contours and probabilistic neural networks [3]. Our work differs in that we use a Gaussian Naive Bayes classifier with features as top face AUs only. This will contribute to the field by focusing on how well the top portion of the face separates emotions.

3 Methodology

3.1 Data Acquisition and Processing

We decided to consider normal intensity and strong intensity for one emotion as the same label (happy or sad). The RAVDESS dataset has speech trials from 12 men and 12 women, acting out a specific emotion per trial. Per subject, 4 sad speech trials and 4 happy trials were used, giving us 192 total trials with even class distributions. Each speech trial has data for the intensities of specific AUs recorded every 0.033 seconds. In a given speech trial the AUs for all time stamps were averaged. This will be referred to as a speech event and individual ones will serve as observations in a new dataset.

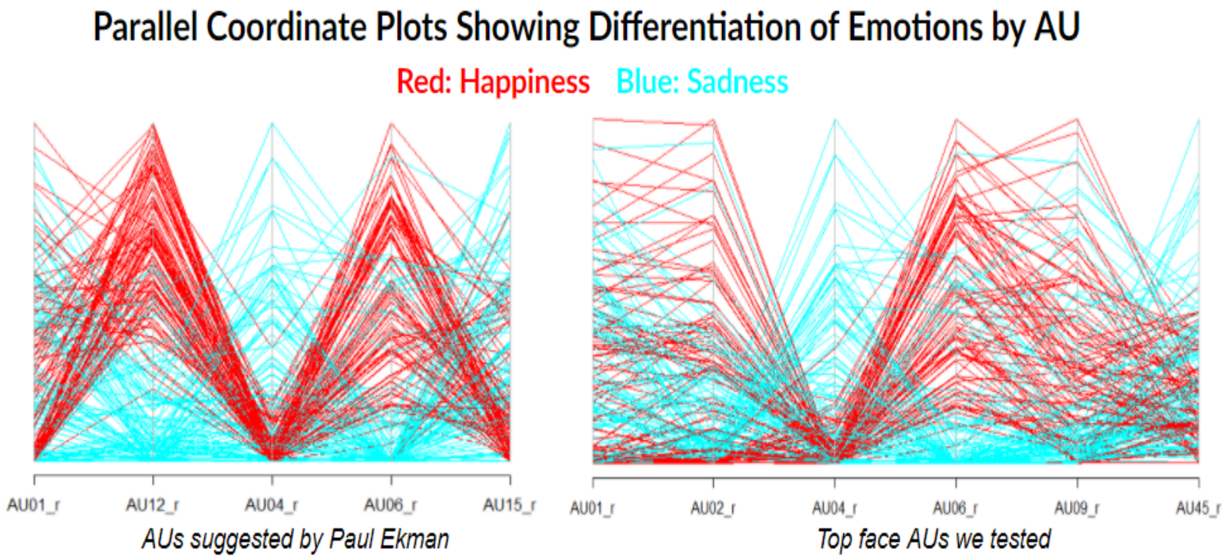


Figure 3.1A: Parallel Coordinate plots showing differentiation between happiness and sadness (left: variables suggested by Ekman, right: variables we tested against the literature)

For happiness and sadness, the most relevant AUs according to Ekman are (1, 4 & 15 for sad) and (6 &12 for happiness) which show strong differentiation for emotions (Figure 3.1A). We chose to test these 5 against (1, 4 & 45 for sadness) and (2, 6 & 9 for happiness). The 6 AU's chosen by us denote top face muscle movements and have a mean intensity value greater than 0.5 across all subjects for the respective emotions. All other features were discarded.

Here, E indicates the specific speech event and t represents the talk time for said speech event. I is the interval. T_E is the number of timestamps for speech event E is defined as,

$$T_E = \frac{Tt}{I} = \frac{Tt}{0.033s} \tag{1}$$

Similarly, for the set of all AU values, measurements are made for every timestamp. Let i denote the i^{th} speech event and j be the j^{th} AU in the set of all 64 AUs going from $1,2,3...j$. An example can be seen below of a speech event (time averaged speech trial) given the AUs we are testing.

$$AU = \{ \overline{AU}_{i,1}, \overline{AU}_{i,4}, \overline{AU}_{i,45}, \overline{AU}_{i,2}, \overline{AU}_{i,6}, \overline{AU}_{i,9} \} \tag{2}$$

3.2 Algorithms used for Analysis

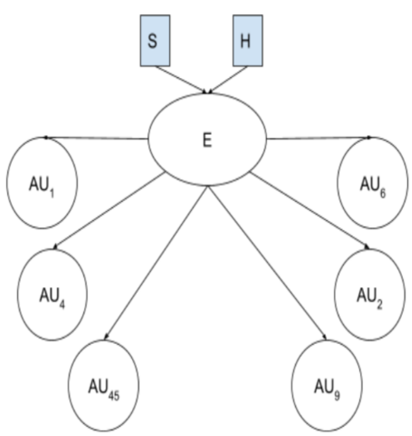
Classification was tested with a Gaussian Naive Bayes model and compared with KNN, SVM (radial, polynomial, and linear kernels), logistic regression, and Random Forest. Testing it across 5 models allows us to be more confident in our results. A Python library, scikit-learn v0.23.1 was used for all of these algorithms and the classification accuracy results were cross-checked in R, except for the Naive Bayes classifier. Accuracy scores from both Python and R deviated by no more than 5%. Both implementations reserved 20% of data for testing. Stratified sampling was performed to create the training and test data to reduce any biases. Five fold cross-validation was used for KNN using R's Class library v7.3-17 and SVM using R's e1071 library v1.7-6. For KNN, 21 neighbors were used and a cost parameter of 10 was used for SVM based on cross validation error. For Random Forest, we used the R library randomForest v4.6-14 with 1000 trees and randomly selected 2 variables from the set of tested AUs (1, 2, 4, 6, 9 & 45) as candidates for each node split. Logistic regression was performed with R library nnet v7.3-14. This process was done for the 6 variables we chose to test, then repeated for the variables suggested by Ekman using the modelled parameters in Figure 3.2A.

The Maximum Likelihood Estimation (MLE) is used to estimate prior probabilities for the Gaussian Naive Bayes classifier. Here, we let N be the number of speech events and Y be the outcome such that,

$$P(E_k) = P(Y = E_k) = \frac{\sum_{i=1}^{N} I(Y_i = E_k)}{N}. \tag{3}$$

The Naive Bayes classifier will use the following formula to make a prediction, K = 1 or 0 for happy and sad respectively.

$$y = argmax\, P(Y = E_k) \prod P(AU \mid Y = E_k) \tag{4}$$



$$\mathrm{AU} = \begin{bmatrix} AU_1 \\ AU_4 \\ AU_{45} \\ AU_2 \\ AU_6 \\ AU_9 \end{bmatrix} \quad \mathrm{E(AU)} = \begin{bmatrix} \mu_{AU_1} \\ \mu_{AU_4} \\ \mu_{AU_{45}} \\ \mu_{AU_2} \\ \mu_{AU_6} \\ \mu_{AU_9} \end{bmatrix} \quad \mathrm{Cov\,(AU)} = \Sigma = \begin{bmatrix} \delta_{AU_1} & 0 & 0 & 0 & 0 & 0 \\ 0 & \delta_{AU_4} & 0 & 0 & 0 & 0 \\ 0 & 0 & \delta_{AU_{45}} & 0 & 0 & 0 \\ 0 & 0 & 0 & \delta_{AU_2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \delta_{AU_6} & 0 \\ 0 & 0 & 0 & 0 & 0 & \delta_{AU_9} \end{bmatrix}$$

The Generative Process:

$$\pi \sim \mathrm{Beta}(\gamma_\pi)$$

$$\mathrm{E} \sim \mathrm{Bern}(\pi)$$

$$\mathrm{AU_j} \sim \mathrm{MV}N(\mu_{AU_i}, \Sigma)$$

$$\mathrm{P(E, AU_1, AU_4, AU_{45}, AU_2, AU_6, AU_9) = P(C)} \prod_{j \in \{AU\}} P(AU_j | E)$$

Figure 3.2A: Naive Bayes graphical model, parameters and generative process

4 Results

We may answer the question concerning the performance of top face AUs for emotional classification by comparing variable importance as well as seeing agreement across classification algorithms. ROC plots were also generated to see how likely an algorithm would generate false positives for class labels (happy and sad). For AUs chosen by us, we noticed that AU6 and AU4 are the best top face features for predicting happiness and sadness compared to 2,9,45, and 1 (Figure 4A).

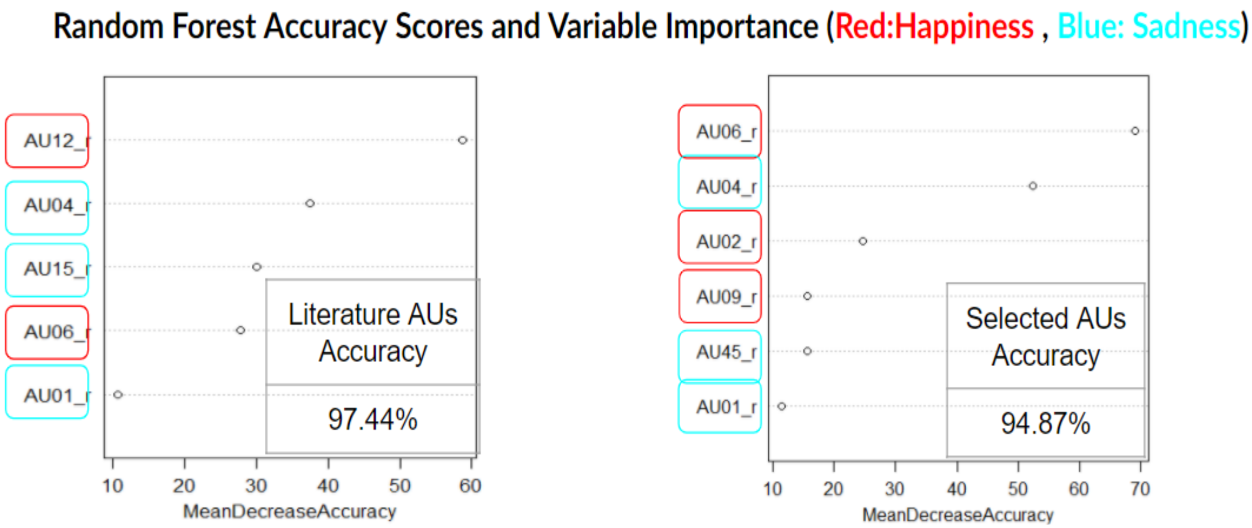


Figure 4A: Mean Decrease Accuracy scores for AU values
(left: suggested AUs by Ekman, right: top face AUs chosen by us)

The accuracy scores of AUs we chose to test show overall similar performances to the literature by Ekman. The radial kernel was the best for support vector machines (SVM). K-NN showed a lower cross validation error with 21 neighbors and provided for a simpler model. More than 1000 trees did not yield better results for Random Forest. Naive Bayes performed poorly for the variables we tested in comparison to those suggested by Ekman (Tables 1 and 2).

Table 1: Test set accuracy scores for AUs provided by Ekman (AUs 1, 4, 6, 12, 15)

Algorithms	Gaussian Naive Bayes	Logistic Regression	SVM (kernel: rbf)	K-NN (n = 21)	Random Forest (n = 1000)
Accuracy	92.3%	94.87%	97.43%	92.31%	97.44%

Table 2: Test set accuracy scores for top face AUs we chose to test (AUs 1, 2, 4, 6, 9, 45)

Algorithms	Gaussian Naive Bayes	Logistic Regression	SVM (kernel: rbf)	K-NN (n = 21)	Random Forest (n = 1000)
Accuracy	82.05%	89.74%	97.4%	89.74%	94.87%

An ROC curve (receiver operating characteristic curve) helps us inspect the performance of a binary classification algorithm where we compare the true positive rate (TPR) to the false positive rate (FPR) for proper detection of the happiness label.All classifiers yielded an excellent trade-off between TPR and FPR, but SVM (kernel: (radial basis function) rbf and polynomial 3rd degree) can be considered the better classifier since it had a TPR of 1 and a FPR of 0 (Figure 4B).

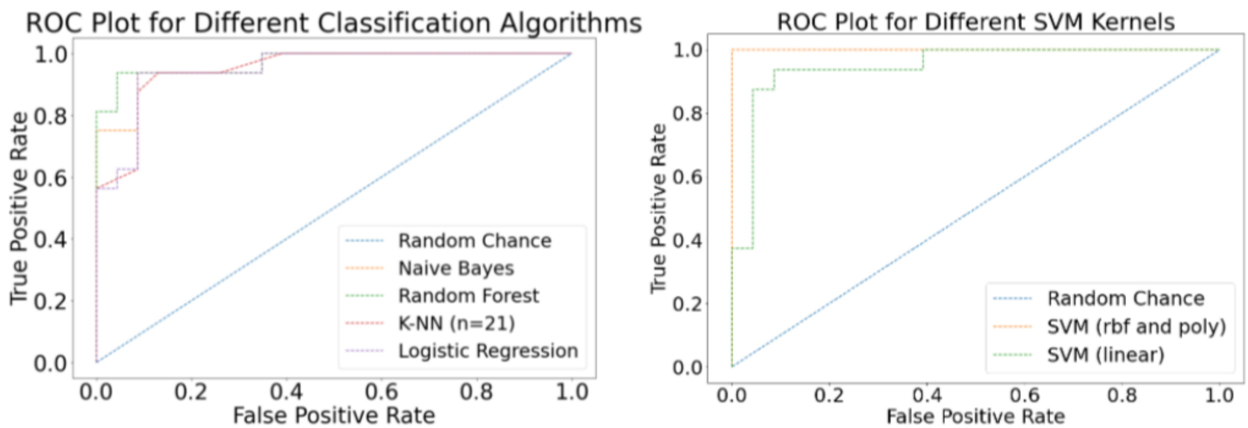


Figure 4B: ROC plots for various classification algorithms to the left and different SVM kernels to the right

The area under the ROC curve (AUC) for various classification algorithms was compared numerically. We noticed SVM had the highest AUC which is as expected since the accuracy for SVM (rbf/poly) was the highest at 97.4% (Table 3).

Table 3: Area under the ROC curve for various classification algorithms

Algorithms	Gaussian Naive Bayes	Logistic Regression	SVM (kernel: rbf and poly)	SVM (kernel: linear)	K-NN (n = 21)	Random Forest (n = 1000)
AUC	0.962	0.948	1.00	0.948	0.948	0.973

5 Discussion

Our results show that AU4 and AU6 are strong predictors of happiness and sadness respectively. Furthermore, since we pooled our data between sexes, we can also say that top face AUs alone are very good predictors of an emotional state. Our results can benefit the fields of psychotherapy, interrogation and general communication since facial features are of focus in such fields. Further research may suggest using different sets of AUs than the ones we choose. For further insight on other AUs that we did not choose to test, please see the supplementary appendix.

References

[1] Ekman, P., & Friesen, W. (1978): Facial Action Coding System: A Technique for the Measurement of Facial Movement. Palo Alto, CA: Consulting Psychologists Press.

[2] Suvashis Das and Koichi Yamada. Article: An HMM based Model for Prediction of Emotional Composition of a Facial Expression using both Significant and Insignificant Action Units and Associated Gender Differences. International Journal of Computer Applications 45(11):11-18, May 2012.

[3] Seyedarabi, Hadi & Lee, Won-Sook & Aghagolzadeh, Ali. (2006). Automatic Lip Tracking and Action Units Classification using Two-Step Active Contours and Probabilistic Neural Networks. 2021 - 2024. 10.1109/CCECE.2006.277379.

[4] Bryn Farnsworth, Ph.D, Facial Action Coding System (FACS) – A Visual Guidebook. August 18, 2019

[5] Raut, Nitisha, "Facial Emotion Recognition Using Machine Learning" (2018). Master's Projects. 632.

[6] Swanson, R., Livingstone, SR., & Russo, FA. (2019). RAVDESS Facial Landmark Tracking (Version 1.0.0) [Data set]. Zenodo. http://doi.org/10.5281/zenodo.3255102

[7] Livingstone SR, Russo FA (2018) The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English. PLoS ONE 13(5): e0196391. https://doi.org/10.1371/journal.pone.0196391.

[8] Chuong B. Do, The Multivariate Gaussian Distribution. October 8, 2010

[9] Richard Zemel, Raquel Urtasun and Sanja Fidler, Lecture 9: Naive Bayes, October 12, 2016

Appendix A - A Comparison for All AUs in the RAVDESS Dataset

Table A-1 AU combinations and corresponding emotions

Emotion	Action Units
Happiness	6 & 12
Sadness	1, 4 & 15
Surprise	1, 2, 5B, & 26
Fear	1, 2, 4, 5, 7, 20 & 26
Anger	4, 5, 7 & 23
Disgust	9, 15 & 17
Contempt	R12A & R14A

Certain combinations of AUs according to the literature by Ekman et al. correspond to specific emotions. In reality, emotions are more nuanced, but the AUs listed should be activated much more than others. We chose to focus specifically on happiness (AUs 6 and 12) and sadness (AUs 1, 4 and 15). The parallel coordinate plot shows that some individual AUs make it very easy to differentiate between happiness and sadness, namely AU4 (for sadness) and AU6 (for happiness) were the most important top face AUs (Figure A-1). This result may not be surprising for happiness, but for sadness it may be more important given the other contenders shown by Ekman's work are AUs 1 and 15.

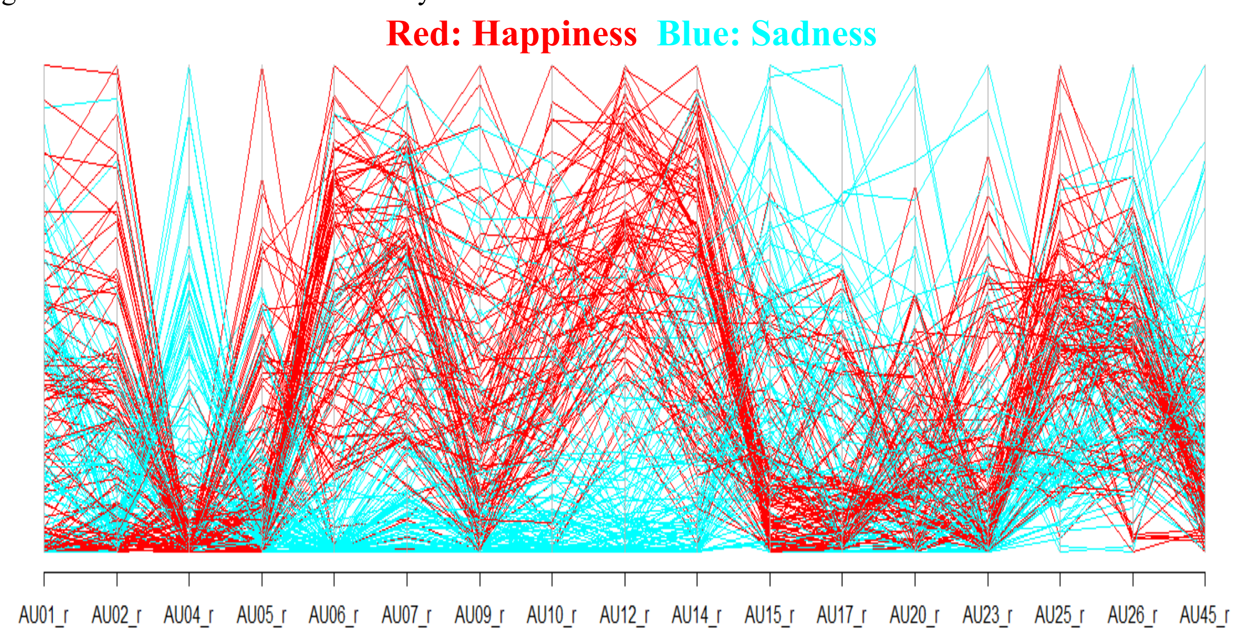


Figure A-1 How all given AUs differentiate happiness and sadness

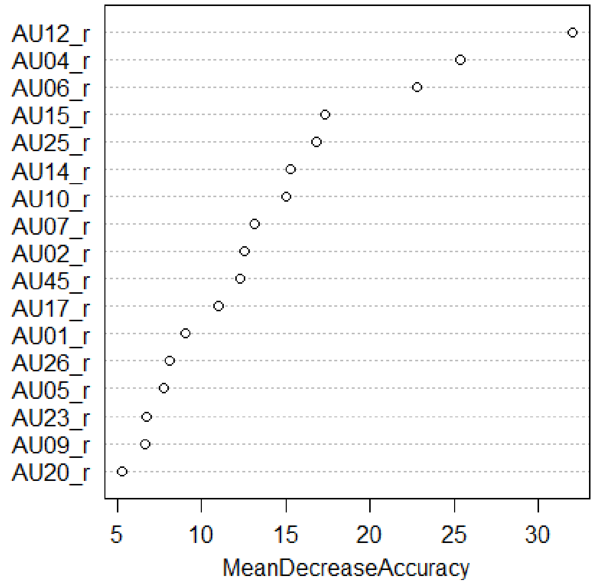


Figure A-2 Variable Importance for All Given AU values

Using all AU values provided in the dataset led to less favorable results for classification with random forest (Figure A-2). Accuracy for classification dropped to around 70% which is why we decided using all variables may not be the best for classification purposes. AUs 4, 6, 12 and 15 are at the top in terms of importance for node splitting which is mostly expected (Figure A-2). AU1 however is quite low even though Ekman's literature suggested high activation of the corresponding muscle movement for sadness. For AUs we chose to test, the pairwise plot shows adequate differentiation of classes (Figure A-3).

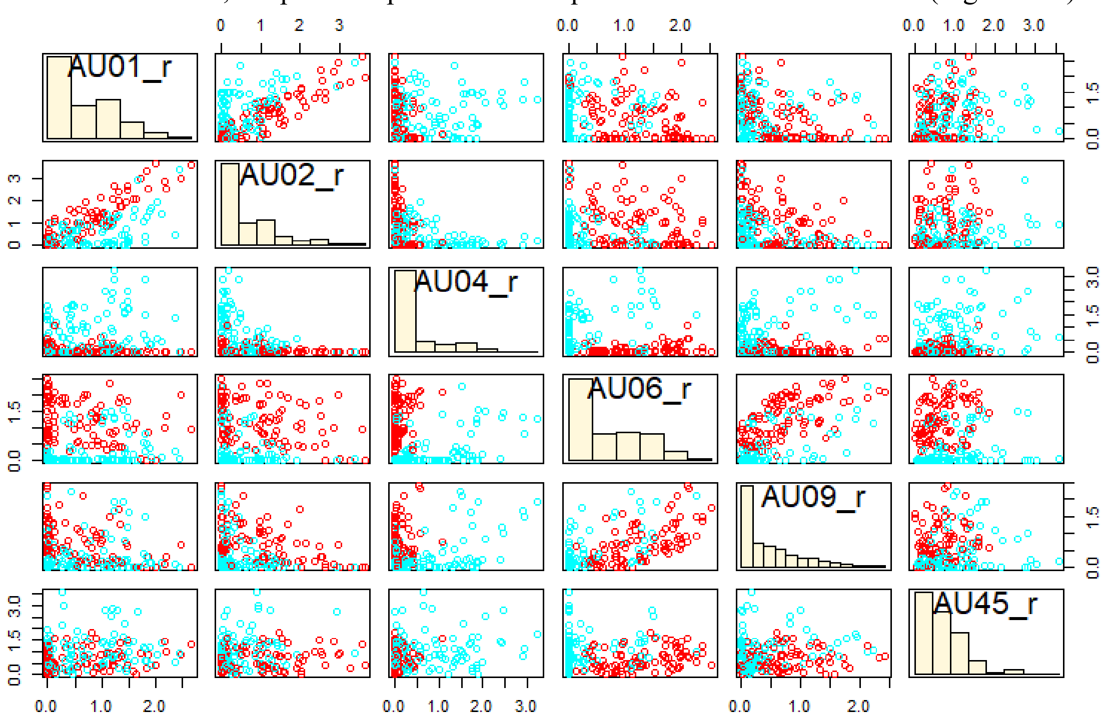


Figure A-3 Pairwise plot for AUs we chose to test