Automatic detection of a driver's complex mental states

Zhiyi Ma[1], Marwa Mahmoud[2], and Peter Robinson[2]

[1] Department of Engineering, University of Cambridge, Cambridge, UK
[2] Computer Laboratory, University of Cambridge, Cambridge, UK

Abstract. Automatic classification of drivers' mental states is an important yet relatively unexplored topic. In this paper, we define a taxonomy of a set of complex mental states that are relevant to driving, namely: *Happy, Bothered, Concentrated* and *Confused*. We present our video segmentation and annotation methodology of a spontaneous dataset of natural driving videos from 10 different drivers. We also present our real-time annotation tool used for labelling the dataset via an emotion perception experiment and discuss the challenges faced in obtaining the ground truth labels. Finally, we present a methodology for automatic classification of drivers' mental states. We compare SVM models trained on our dataset with an existing nearest neighbour model pre-trained on posed dataset, using facial Action Units as input features. We demonstrate that our temporal SVM approach yields better results. The dataset's extracted features and validated emotion labels, together with the annotation tool, will be made available to the research community.

1 INTRODUCTION

Complex mental states occur often in real-world driving, and drivers' mental states can have an impact on their driving behaviours, such as speed, acceleration and traffic violations [26]. Therefore, it would be desirable to identify drivers' complex mental states automatically. This can also be very useful for car manufacturers, such as to make cars smarter.

Although there have been many studies looking at stress level [16, 12], drowsiness [13, 19], and basic emotions [23] of a driver, there is not much work on detection of complex mental states, which were actually found to occur more often in real life [27]. Also, speech recognition [14], physiological signals [16, 22] and grip strength [23] have been investigated in classifying drivers' mental states. However, analysing emotions from drivers' facial expressions are not fully explored, partially due to the challenge of lighting conditions during driving and the relatively subtle expressions exhibited while driving.

In this paper, we present our work on automatic detection of drivers' complex mental states based on facial analysis and real-world driving videos. We looked into four complex mental states that appear frequently in driving scenarios, and

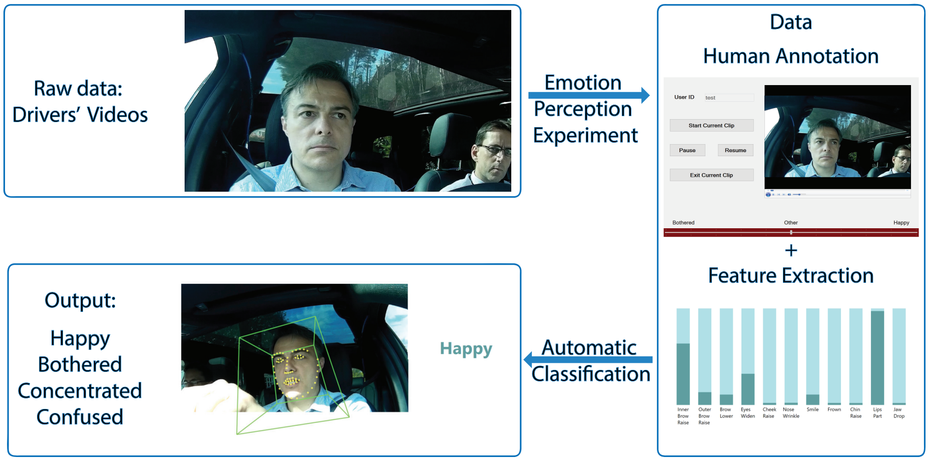


Fig. 1: An overview of our work: The mental states were labeled per frame by human annotators. The features (facial Action Units) were extracted by OpenFace [2]. The long video segments were then processed to small clips and subsampling was used to balance the dataset. Two models were evaluated: per video nearest neighbour model based on an existing classifier and per frame SVM classification.

performed our automatic classification on natural data. Fig. 1 gives an overview of our emotion labeling and classification approach.

The main contributions of our work can be summarised as follows:

1. We present a spontaneous dataset collected in a natural driving scenario and define a set of four categories of complex mental states that are relevant to driving: *Happy*, *Bothered*, *Concentrated*, and *Confused*. We obtained our ground truth labels through emotion perception experiment. The features and validated labels, together with the annotation tool, will be be publicly available to the research community.
2. Using the dataset, we present a methodology for automatic complex mental states classification based on extracting per frame facial Action Units (AU) and using Support Vector Machine (SVM).
3. We compare the SVM models with a nearest neighbour automatic complex mental states classifier pre-trained on posed data [1], and demonstrate that the temporal SVM model achieves significantly better classification results.

1.1 Related Work

Lots of studies have explored automatic emotion recognition, such as [4, 5, 30]. However, most of these studies were based on the basic emotions [7]. In automotive domain specifically, Katsis et al [16] studied the automatic stress level recognition in car-racing drivers. Hu and Zhang [13], Lisetti and Nasoz [19] both investigated drivers' drowsiness. Basic emotions such as *happiness* and *anger* were also investigated [23]. A few studies have explored more complex mental states [1, 9] but not related to driving.

Lots of work on automatic emotion classification in automotive domain is based on simulated racing conditions [15, 16, 12]. Although the validity of driving simulator has been proved [18], racing conditions can be different from normal driving.

Meanwhile, there have been some studies in automatic emotion recognitions in cars using various features, such as speech recognition [14], physiological signals like galvanic skin response and heart beat [16, 22], and grip strength [23]. While speech recognition relies on the driver to talk, which can be hard when the driver is alone, obtaining physiological signals can sometimes be invasive and distractive. In contrast, facial analysis has the advantage of non-invasive and non-distractive.

2 Relevant mental states

There have been two most common models on constructing computational models of emotions: categorical and dimensional model [11]. Categorical model divides affective states into discrete emotion categories, assuming that there exists such emotions that are hard-wired in our brain. Dimensional model conceptualises emotions as points in a continuous space on the chosen dimensions. So far both theories have shown their merits and demerits [11]. We adopted the categorical approach in our work as it is more intuitive to non-expert labellers.

We chose our taxonomy based on the work of Baron-Cohen et al [28]. This work contained an exhaustive list of 24 complex mental states groups, with 412 emotion concepts associated. Based on an initial investigation of the videos, eight complex mental states from this list were chosen, i.e. *angry, bothered, disgusted, excited, happy, interested, thinking* and *unsure*, which are believed to be relevant in a driving scenario.

Because the facial expressions of the drivers were mostly subtle, we combined similar groups to form four big categories of complex mental states. *Excited, happy* and *interested* were combined to *Happy*, while *angry, disgusted* and *bothered* to *Bothered*. Hence we defined our taxonomy to constitute of four categories, namely *Happy, Bothered, Concentrated* and *Confused* (see Fig. 2 for some sample frames).

3 Dataset

This section describes five stages of data collection: collecting driving videos, designing annotation tool, conducting emotion perception experiment, validating the collected labels and preparing data for classification.

3.1 Video Collection

The original dataset is composed of 30 video segments of 10 participants driving in a natural environment. The lengths of segments vary from 8 to 20 minutes.

All segments were recorded using a frontal webcam placed on the dashboard in front of the driver, with a frame rate of 30 fps.

Among the 10 drivers, five were driving alone following the instructions on a sat-nav, while the rest were following the instructions given by a co-pilot (passenger). This setup was meant to get varied amount of mental states expressed.

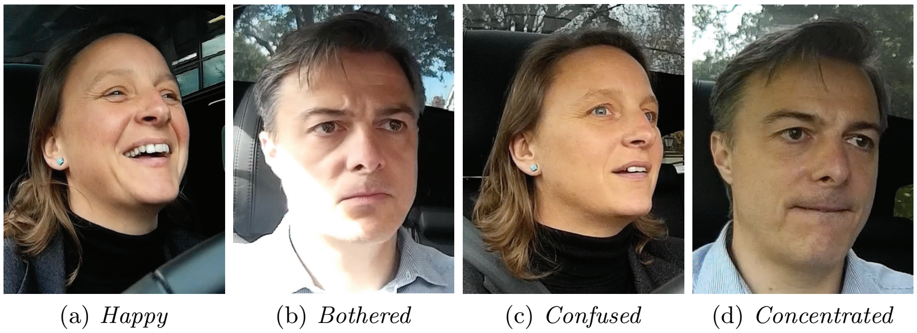

(a) *Happy* (b) *Bothered* (c) *Confused* (d) *Concentrated*

Fig. 2: Samples from videos for each emotion category: *Happy, Bothered, Confused, Concentrated*. The emotion concepts included in each category are: *Happy: Comfortable, Happy*; *Bothered: Frustrated, Nervous, Bothered*; *Confused: Unsure, Confused*; *Concentrated: Thinking, Concentrated*.

3.2 Annotation Tool

In order to collect emotion labels, we developed an annotation tool for real-time annotation. Unlike some available open-source tools such as FEELTRACE [6] and ANNEMO [25], our tool has three features that makes it suitable for our labelling task: 1) It is a stand-alone offline tool. Since the original videos were confidential, this was important for our task. 2) It supports real-time annotation. Given the length of the segments, it was impossible to describe a segment with only one emotion label. 3) As a real-time annotation tool, the interface should be as easy as possible to use, i.e. fewer states with straightforward layout to be chosen from, in order to reduce the reaction time and avoid the need to rewind the video.

We used two annotation modes to further simplify the task of real-time labeling. In every mode, two categories are labelled. To explain each category to non-expert annotators, we explained the meaning of each mental state with their associated emotion concepts [28]. Fig. 2 shows sample frames from our videos for each emotion category, accompanied by the emotion concepts included in each category.

In each annotation mode, a continuous scale of 5 levels for each mental state was used, with the higher level indicating a higher intensity. Since the emotional expressions can be subtle in the driver's face, the lower levels encourages annotators to capture a weak mental state appearance. An *Other* option representing level 0 is also included to avoid a forced choice. Fig. 3 shows a snapshot.

Labels are logged by the annotation tool in real-time, and converted to per-frame annotation according the frame rate of the video. The tool is open-source and will be available to the research community.

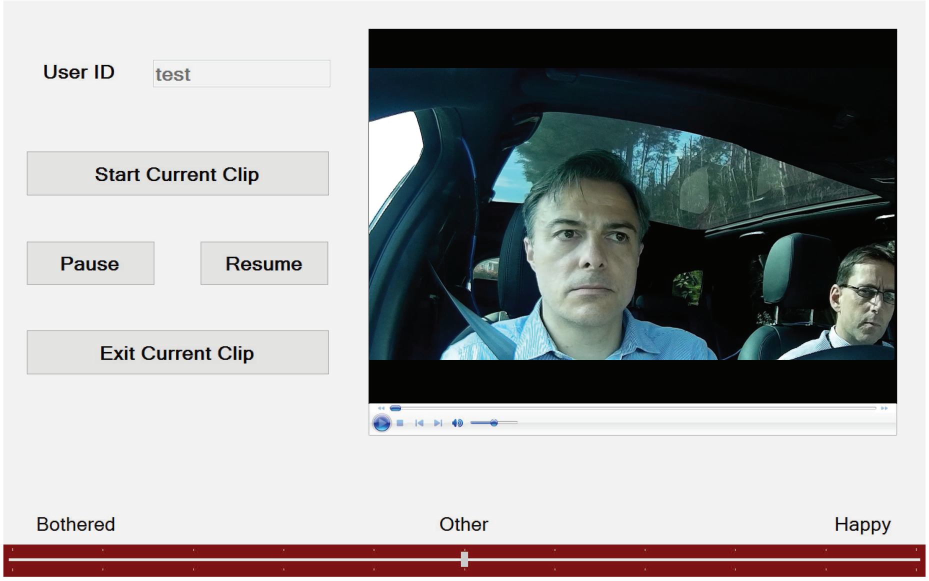
3.3 Emotion Perception Experiment

Annotators were recruited to our laboratory to label the videos via an emotion perception experiment. All audio in the original video segments were muted to exclude verbal cues.

A total number of 24 segments from 10 drivers were annotated by 48 annotators. Each segment was fully annotated by 6 annotators, half in annotation mode A and half in mode B. Each annotator was rewarded with a £10 Amazon Voucher after finishing annotating three randomly pre-selected segments. To eliminate learning effect, the segments for each annotator were chosen from three different drivers, and the annotation mode was alternate (i.e. ABA or BAB).

3.4 Inter-Rater Agreement Analysis

To evaluate the reliability of the collected labels, we used Krippendorff's Alpha to measure multiple annotators' agreement [17]. It was calculated per video segment for both original continuous scale labels and categorical labels of each



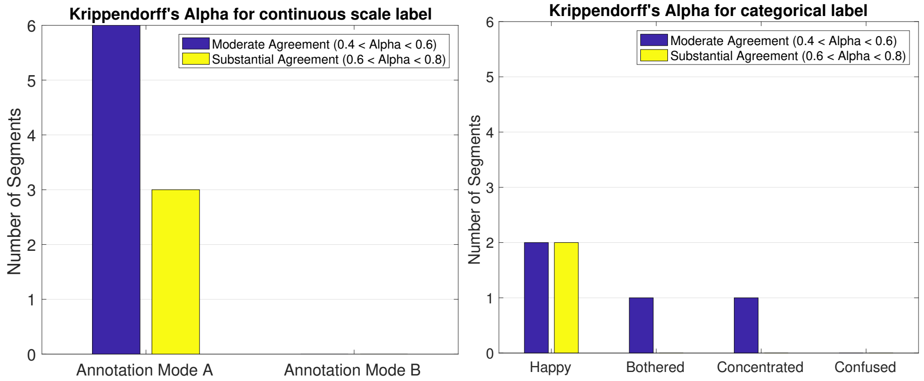
User ID test

Fig. 3: A snapshot of the video annotation tool interface

mental state category (i.e. converting to binary labels for each mental state without considering the intensity). All level 0's were treated as empty labels in continuous scale labels since it is systematically different from intensity values.

Fig. 4 shows Alpha values distributions. For continuous scale labels in annotation mode A, a quarter of the segments have moderate agreement (alpha value 0.4∼0.6) while 12.5% have substantial agreement (alpha value 0.6∼0.8). Only 21% have fair agreement (0.2∼0.4) in mode B (not shown in Fig. 4). Considering each mental state category, *Happy* has the best agreement, indicating it being the most reliably labelled category. The fairly smaller number of segments with moderate agreement from other categories explains the relatively low agreement for mode B annotations.

It can be a challenge to achieve good agreement in collecting large naturalistic datasets [20], especially with mostly non-expert annotators. To deal with this challenge, we adopted a short clip approach in data preparation, which is discussed in Section 4.2.



(a) Continuous scale label

(b) Categorical label

Fig. 4: Krippendorff's Alpha distribution for: (a) continuous scale label of the two annotation modes; (b) categorical label of the each mental state category.

4 Methodology

This section describes the main stages of our automatic mental state estimation approach: feature extraction, data pre-processing (including filtering and balancing), and automatic classification of mental states.

4.1 Feature extraction

Input features, i.e. facial Action Units (AU) [8] intensities, are extracted by OpenFace [2] from the segments. A continuous scale from 0 to 5 indicating the intensities of 14 AUs are returned respectively per frame. OpenFace is based on the state-of-art AU recognition framework [3, 29] with slight modifications tailored for natural video dataset [2]. It outperformed similar softwares DL [10], BA and BG [29] in most AUs on the public dataset SEMAINE [21] and BP4D [31], achieving an average F1-score of 0.48 [2]. Therefore, it is reliable and well-suited to our work.

4.2 Data Pre-processing

Data filtering To avoid different annotator's bias when annotating in the continuous scale, categorical labels were used throughout classification.

We adopted a perfect agreement small clip approach to deal with the long segments and low agreement in this experiment. We defined "Perfect agreement" as all three annotators agreeing on the mental state category of the same frame. There are two variables in choosing the small clips: 1) A minimum length of continuous perfect agreement on one of the four mental state categories. This ensures a stable and reliable perceived existence of the mental states. 2) A minimum threshold on the confidence level of the face tracking. Because OpenFace feature extraction depends on reliable facial Action Unit tracking, specifying this threshold is crucial to guarantee the reliability of the feature vector, especially when the natural driving inevitably includes some extreme head motions that affects face tracking.

There are two rationales for not using the whole segment directly: 1) Labels were collected per frame. For frames where people fail to agree on the group of mental states, we cannot produce a reliable single ground truth label. 2) There is an "Other" option in labelling, which is a placeholder of other mental states that might occur but were not included in our taxonomy. The small clip approach excluded this label altogether. Thus, we could ensure the reliability of the labels of those short segments.

To guarantee data quality, frames from perfect agreement small clips, with the minimum length set to 1 second and tracking successful rate of 100%, were used for SVM training.

Data Balancing One challenge in our work was to balance the dataset. Table 1 shows the total number of frames available for each mental state category. Clearly not many were obtained for *Confused*. The videos show that the associated head pose change and hand-over-face occlusion causes face tracking to fail from time to time. Therefore, the 100% tracking successful rate requirement was waived on *Confused* clips, but only frames that were successfully tracked were used for classification.

We then used random subsampling to balance the number of frames from each mental state category. For each driver, the number of frames from the mental state that has the minimum number of frames was set as a baseline, and frames were subsampled from the other mental states of this driver to its baseline. When this baseline was 0, we combined a driver's data with another driver into a new fold. Specifically, drivers 6, 7, 10 and drivers 2,3 were combined to be two "drivers", resulting in a total of 7 "drivers", all containing frames from mutually different faces (see Table 1). Subsampling guaranteed having a balanced dataset for the subsequent automatic classification steps.

Table 1: The number of frames of each mental state from different drivers before subsampling. Numbers in bold represent the baseline adopted for each driver (or sum up for a group of drivers). Subsampling guaranteed having a balanced dataset for the subsequent automatic classification steps.

Driver	Mental States Categories			
	Happy	Bothered	Concentrated	Confused
1	702	2086	9937	**100**
2	**90**	384	9292	101
3	**130**	0	1361	469
4	9655	2520	17647	**432**
5	3500	3365	7745	**1616**
6	39	781	997	**456**
7	0	1186	2416	**119**
10	6308	0	1452	**18**
8	6179	2462	8893	**2261**
9	9758	3737	15603	**866**
Total	36361	16521	75343	6438
Subsampled Total	6088	6088	6088	6088

4.3 Automatic classification of mental states

We trained Support Vector Machine (SVM) models with our spontaneous dataset to automatically classify the labelled mental states. A per frame approach was adopted, to match the collected label format.

We first trained two binary classifiers according to the two annotation modes, and then performed a four-class classification. Two approaches for each type of classifiers were used: 1) SVM per frame approach, where the 14 AU intensities were directly used as the feature vector for each frame. 2) Temporal per frame approach, where the context of the video is taken into consideration, and the AU intensities from the previous frame was appended to the current frame to form a 28-dimension feature vector. We experimented with linear and radial basis function (RBF) kernels SVM approaches. To optimise the parameters, we used a leave-one-driver-out cross-validation, with gamma and C varied in the range [0.01, 0.1, 1, 10, 100].

Pre-trained Nearest Neighbour Classification For comparison purposes, using our spontaneous dataset, we evaluated a nearest neighbour (NN) approach based on a complex mental states classifier [1], pre-trained on posed data [24]. We chose this system as it is based on the same set of complex mental states we use in our experiments. In this approach, a single feature vector is obtained per video, by dividing the AUs and speed of AUs uniformly into different bins. And the three emotion categories whose overall feature vector is the closest to the video's are returned as the classification results.

We used perfect agreement small clips with fixed length of five seconds and 80% successful rate, since previous studies showed that two seconds is the minimum time for human to reliably detect a mental state [9]. There are a total of 18 emotions available in the classifier. We discarded those that are not relevant to our taxonomy, and grouped the rest to match ours according the emotion grouping discussed in Section 2. Details are shown in Table 2. An extra emotion called "average" was used in this classifier to normalise the feature vectors. We kept this emotion but did not assign it to any mental states. Overall a total of 10 emotions were used.

Table 2: Emotions grouping for per video NN classifier

Mental State Category	Emotions
Happy	Excited, Happy, Interested
Bothered	Angry, Disappointed, Disgusted, Frustrated
Concentrated	Neutral
Confused	Worried
Discarded Emotions	Afraid, Ashamed, Bored, Hurt, Joking Pround, Sad, Sneaky, Surprised

5 Experimental Results

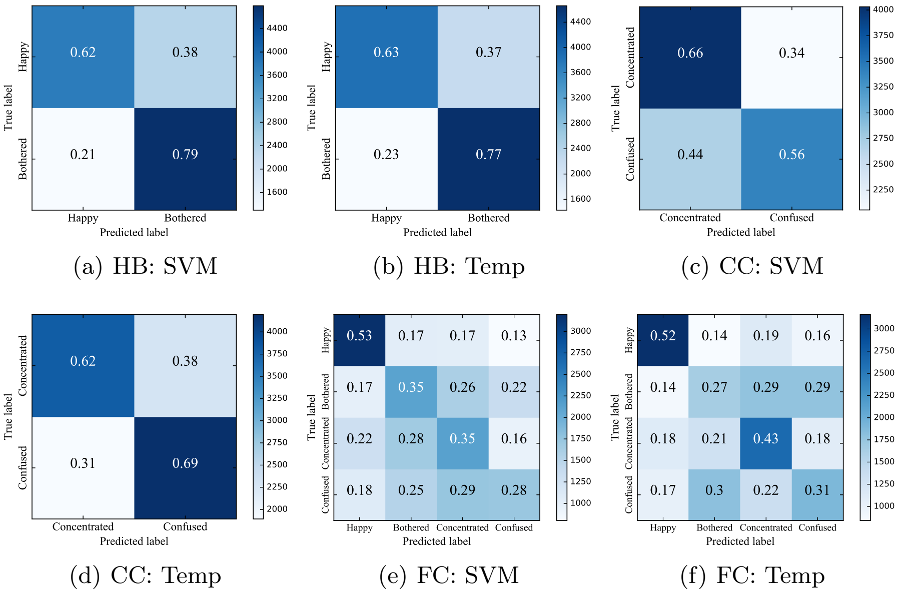
This section presents the classification results of our experiments. We use F1-score to report the performance of each system.

5.1 Per Frame Approaches using SVM

In all of these experiments, we used a user-independent 7 fold cross validation method. The normalised confusion matrices for binary and four-class classifiers are shown in Fig. 5. Table 3 tabulates the F1-score for each mental state and overall for each classifier. The statistical significances were calculated using F1-score for each mental state with a one-tail t-test, which indicates the statistically significant improvements from random choice F1-score (0.5 and 0.25 for binary and four-class, respectively).

All SVM training methods achieved F1-scores significantly higher than chance except for *Concentrated-Confused* SVM per frame classifier. In general, *Happy-Bothered* classifiers perform better than their counterparts *Concentrated-Confused* classifiers, which agrees with the inter-rater agreement results that showed that *Happy* and *Bothered* are more reliably detected by human annotators. Moreover, the SVM temporal approach had the best performance.

Overall, we can see that our per frame approaches using SVM managed to classify the four mental states categories. The four-class classifiers achieved significantly better results than chance. Concurrent states might have affected the classification results. This will be discussed in Section 6.



(a) HB: SVM

(b) HB: Temp

(c) CC: SVM

(d) CC: Temp

(e) FC: SVM

(f) FC: Temp

Fig. 5: Normalised confusion matrices for per frame approaches. "HB" stands for "*Happy-Bothered* binary classifier", "CC" for "*Concentrated-Confused* binary classifier", and "FC" for "Four-Class classifier". "SVM" means "SVM per frame approach", and "Temp" is short for "temporal per frame approach".

5.2 Per Video Nearest Neighbour Classification

The per video classifier returns three nearest labels/mental states by default, and we only consider the nearest one in order to compare with our four-class classifiers.

Table 4 tabulates the confusion matrix and F1-score for each mental state category. The five-second requirement means extremely limited number of *confused* clips, thus we eliminate the effect of varied number of clips for each mental state by taking a weighted average.

When we performed a one-tail t-test, p value was found to be around 0.4, indicating that there is no significant improvement (or difference) compared to

Table 3: The F1-scores of binary and four-class SVM classifiers

Classifier	F1-score					Chance F1-score
	Overall	Happy	Bothered	Concentrated	Confused	
HB SVM*	0.70	0.67	0.73	-	-	0.50
HB Temp*	0.70	0.68	0.72	-	-	
CC SVM	0.61	-	-	0.63	0.59	
CC Temp*	0.66	-	-	0.64	0.67	
FC SVM*	0.37	0.50	0.34	0.34	0.31	0.25
FC Temp*	0.38	0.51	0.28	0.41	0.32	

* Indicates a statistically significant difference of p < 0.05 compared with chance F1-score

chance. The main drawback is the poor performance on classifying *Concentrated* and *Confused*. This suggests that the nearest neighoubr classifier, pre-trained on posed data, does not perform as well on our spontaneous driving dataset.

We could conclude that, for four-class classifiers, the temporal per frame approach using SVM outperforms the pre-trained nearest neighbour classifier, and can classify each mental state with an accuracy significantly higher than chance.

Table 4: Per video nearest neighbour classification, $p < 0.4$

		Predicted label				F1-score	Chance F1-score
		Happy	Bothered	Concentrated	Confused		
True label	Happy	111	34	9	3	0.51	0.24
	Bothered	20	40	3	10	0.27	0.11
	Concentrated	146	145	104	11	0.40	0.63
	Confused	5	5	3	0	0	0.02
Overall	-	-	-	-	-	0.40	0.46

6 Discussion and Challenges

One big challenge in mental states classification is the existence of concurrent states. From the four-class confusion matrix in Fig. 5, we can see that *Bothered* and *Confused* are generally confused with each other, while *Happy* and *Concentrated* are less confused and usually well-identified. This explains the relatively low accuracy of the former two compared with the latter.

Another challenge is the unbalanced dataset. As Table. 1 shows, *Confused* has the least amount of data, which might have affected the results despite the efforts to balance the dataset using subsampling. However, this is common because we do not expect *Confused* to happen too often in a natural driving scenario.

One last aspect is the driving setup. As described half of the drivers were instructed by a co-pilot (passenger) while the rest were alone following a sat-nav. It was observed that drivers tend to be more expressive when they are accompanied. This might have increased the frequency of all the happiness label.

7 Conclusion

We presented the taxonomy of four categories of mental states that are believed to be related to the driving context. We have used a spontaneous dataset of drivers' videos, and evaluated two models of automatic mental states detection. Comparing our approach with a per video pre-trained classifier, our temporal per frame SVM classifier performs significantly better. We presented our real-time annotation tool and discussed the challenges of obtaining validated ground truth labels. We also discussed the challenges of having an unbalanced dataset, which is expected for this type of natural dataset. The dataset features and labels will be publicly available to the research community.

For future work, we would like to use different models such as Hidden Markov Models (HMM). It may also be worth using a multi-modal approach integrating speech and body with the facial features. Finally, more context-related data is needed. Not having enough data, especially for *Confused*, affected the classification results.

8 Acknowledgment

The work presented in this paper was funded and supported by Jaguar Land Rover, Coventry, UK.

References

1. Adams, A., Robinson, P.: Automated recognition of complex categorical emotions from facial expressions and head motions. In: Affective Computing and Intelligent Interaction (ACII), 2015 International Conference on. pp. 355–361. IEEE (2015)
2. Baltru, T., Robinson, P., Morency, L.P., et al.: Openface: an open source facial behavior analysis toolkit. In: 2016 IEEE Winter Conference on Applications of Computer Vision (WACV). pp. 1–10. IEEE (2016)
3. Baltrušaitis, T., Mahmoud, M., Robinson, P.: Cross-dataset learning and person-specific normalisation for automatic action unit detection. In: Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on. vol. 6, pp. 1–6. IEEE (2015)
4. Bartlett, M.S., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I., Movellan, J.: Recognizing facial expression: machine learning and application to spontaneous behavior. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). vol. 2, pp. 568–573. IEEE (2005)
5. Cohn, J.F., De la Torre, F.: Automated face analysis for affective. The Oxford handbook of affective computing p. 131 (2014)
6. Cowie, R., Douglas-Cowie, E., Savvidou*, S., McMahon, E., Sawey, M., Schröder, M.: 'feeltrace': An instrument for recording perceived emotion in real time. In: ISCA tutorial and research workshop (ITRW) on speech and emotion (2000)
7. Ekman, P.: An argument for basic emotions. Cognition & emotion 6(3-4), 169–200 (1992)
8. Ekman, P., Rosenberg, E.L.: What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS). Oxford University Press, USA (1997)
9. El Kaliouby, R., Robinson, P.: Real-time inference of complex mental states from facial expressions and head gestures. In: Real-time vision for human-computer interaction, pp. 181–200. Springer (2005)
10. Gudi, A., Tasli, H.E., den Uyl, T.M., Maroulis, A.: Deep learning based facs action unit occurrence and intensity estimation. In: Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on. vol. 6, pp. 1–5. IEEE (2015)
11. Gunes, H., Schuller, B.: Categorical and dimensional affect analysis in continuous input: Current trends and future directions. Image and Vision Computing 31(2), 120–136 (2013)
12. van den Haak, P., van Lon, R., van der Meer, J., Rothkrantz, L.: Stress assessment of car-drivers using eeg-analysis. In: Proceedings of the 11th International Conference on Computer Systems and Technologies and Workshop for PhD Students in Computing on International Conference on Computer Systems and Technologies. pp. 473–477. ACM (2010)
13. Hu, S., Zheng, G.: Driver drowsiness detection with eyelid related parameters by support vector machine. Expert Systems with Applications 36(4), 7651–7658 (2009)
14. Jones, C.M., Jonsson, I.M.: Automatic recognition of affective cues in the speech of car drivers to allow appropriate responses. In: Proceedings of the 17th Australia conference on Computer-Human Interaction: Citizens Online: Considerations for Today and the Future. pp. 1–10. Computer-Human Interaction Special Interest Group (CHISIG) of Australia (2005)
15. Katsis, C., Goletsis, Y., Rigas, G., Fotiadis, D.: A wearable system for the affective monitoring of car racing drivers during simulated conditions. Transportation research part C: emerging technologies 19(3), 541–551 (2011)

16. Katsis, C.D., Katertsidis, N., Ganiatsas, G., Fotiadis, D.I.: Toward emotion recognition in car-racing drivers: A biosignal processing approach. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans 38(3), 502–512 (2008)
17. Krippendorff, K.: Agreement and information in the reliability of coding. Communication Methods and Measures 5(2), 93–112 (2011)
18. Lee, H.C., Cameron, D., Lee, A.H.: Assessing the driving performance of older adult drivers: on-road versus simulated driving. Accident Analysis & Prevention 35(5), 797–803 (2003)
19. Lisetti, C.L., Nasoz, F.: Affective intelligent car interfaces with emotion recognition. In: Proceedings of 11th International Conference on Human Computer Interaction, Las Vegas, NV, USA. Citeseer (2005)
20. Mahmoud, M., Baltrušaitis, T., Robinson, P., Riek, L.D.: 3d corpus of spontaneous complex mental states. In: International Conference on Affective Computing and Intelligent Interaction. pp. 205–214. Springer (2011)
21. McKeown, G., Valstar, M.F., Cowie, R., Pantic, M.: The semaine corpus of emotionally coloured character interactions. In: Multimedia and Expo (ICME), 2010 IEEE International Conference on. pp. 1079–1084. IEEE (2010)
22. Nasoz, F., Ozyer, O., Lisetti, C.L., Finkelstein, N.: Multimodal affective driver interfaces for future cars. In: Proceedings of the tenth ACM international conference on Multimedia. pp. 319–322. ACM (2002)
23. Oehl, M., Siebert, F.W., Tews, T.K., Höger, R., Pfister, H.R.: Improving human-machine interaction–a non invasive approach to detect emotions in car drivers. In: International Conference on Human-Computer Interaction. pp. 577–585. Springer (2011)
24. O?Reilly, H., Pigat, D., Fridenson, S., Berggren, S., Tal, S., Golan, O., Bölte, S., Baron-Cohen, S., Lundqvist, D.: The eu-emotion stimulus set: A validation study. Behavior research methods pp. 1–10 (2015)
25. Ringeval, F., Sonderegger, A., Sauer, J., Lalanne, D.: Introducing the recola multimodal corpus of remote collaborative and affective interactions. In: Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on. pp. 1–8. IEEE (2013)
26. Roidl, E., Frehse, B., Höger, R.: Emotional states of drivers and the impact on speed, acceleration and traffic violations?a simulator study. Accident Analysis & Prevention 70, 282–292 (2014)
27. Rozin, P., Cohen, A.B.: High frequency of facial expressions corresponding to confusion, concentration, and worry in an analysis of naturally occurring facial expressions of americans. Emotion 3(1), 68 (2003)
28. Simon Baron-Cohen, Ofer Golan, S.W.: A new taxonomy of human emotions (2004)
29. Valstar, M.F., Almaev, T., Girard, J.M., McKeown, G., Mehu, M., Yin, L., Pantic, M., Cohn, J.F.: Fera 2015-second facial expression recognition and analysis challenge. In: Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on. vol. 6, pp. 1–8. IEEE (2015)
30. Whitehill, J., Bartlett, M., Movellan, J.: Automatic facial expression recognition for intelligent tutoring systems. In: Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on. pp. 1–6. IEEE (2008)
31. Zhang, X., Yin, L., Cohn, J.F., Canavan, S., Reale, M., Horowitz, A., Liu, P., Girard, J.M.: Bp4d-spontaneous: a high-resolution spontaneous 3d dynamic facial expression database. Image and Vision Computing 32(10), 692–706 (2014)