Predicting Folds in Poker Using Action Unit Detectors and Decision Trees

Doratha Vinkemeier
Computer Vision Lab
University of Nottingham
Nottingham, UK
doratha.vinkemeier@nottingham.ac.uk

Michel Valstar
Computer Vision Lab
University of Nottingham
Nottingham, UK
michel.valstar@nottingham.ac.uk

Jonathan Gratch
Institute for Creative Technologies
University of Southern California
Playa Vista, USA
gratch@ict.usc.edu

***Abstract*—Predicting how a person will respond can be very useful, for instance when designing a strategy for negotiations. We investigate whether it is possible for machine learning and computer vision techniques to recognize a person's intentions and predict their actions based on their visually expressive behaviour, where in this paper we focus on the face. We have chosen as our setting pairs of humans playing a simplified version of poker, where the players are behaving naturally and spontaneously, albeit mediated through a computer connection. In particular, we ask if we can automatically predict whether a player is going to fold or not. We also try to answer the question of at what time point the signal for predicting if a player will fold is strongest. We use state-of-the-art FACS Action Unit detectors to automatically annotate the players facial expressions, which have been recorded on video. In addition, we use timestamps of when the player received their card and when they placed their bets, as well as the amounts they bet. Thus, the system is fully automated. We are able to predict whether a person will fold or not significantly better than chance based solely on their expressive behaviour starting three seconds before they fold.**

***Keywords*-automatic facial analysis; human behaviour; machine learning;**

I. Introduction

It is widely believed that a person's face can reveal many things about them, including what they are thinking or feeling. For this reason, the human face has been an object of scientific study going back at least as far as Duchenne [1], who methodically studied the role of facial muscles in human expression, and Darwin [2] who studied human and animal expression in the context of the theory of evolution. Understanding human expression, and in particular human expression conveyed by the face, is considered central to our understanding of others and of ourselves and therefore there has been effort to decode what thoughts and states of mind the face reveals [3].

Understanding human expression is important in any field involving interaction with humans. Humans and computers interact extensively. Since computers are both partners in interactions and also powerful tools, it is becoming increasingly important for computers to correctly analyse and interpret human behaviour. Automatic Facial Analysis is a field of computer science that uses computer vision to analyse and interpret human facial expressions. In this paper, we are concerned with *sign-based* measurement of human facial expression, which describes the appearance of the face in terms of which facial muscles are active. This is in contrast to *message-based* measurement which describes the face in terms of the reasons for displaying the facial expressions [4] such as feeling one of the 'six basic emotions' - anger, disgust, fear, happiness, sadness, surprise.

Technology for automatically recognizing facial muscle action has advanced in the last years and many applications have been found in fields like advertising, HCI and medicine. There has been so much promising research into combining computer vision and medicine that Valstar proposed creating a new field, *Behaviomedics*, to facilitate its development [5]. Recent applications include depression [6] [7], automatically recognizing ADHD and ASD [8], and automatically measuring pain [9]. These applications take advantage of the potential for computer vision to provide objective, repeatable measurements, to pick up subtleties humans may miss, and to do so in a relatively unobtrusive way using equipment, like webcams, that is cheap, fast and easy to obtain.

In this paper we use computer vision and machine learning techniques to study pairs of people playing a simplified version of poker. In particular, we ask if, by using action unit detectors together with a decision tree, we can automatically predict a player's actions. Automatically predicting a person's decisions has been explored before in the setting of social dilemmas [10] and also in negotiations [11]. Here, in the context of poker, we are interested in whether the player is about to fold versus raise or call. We also seek to answer the question of when this signal, if it exists, is strongest. Thus, we are using computer vision and machine learning techniques to study a human behaviour whose expression is likely to be only fleeting as it is associated with a passing event, and whose expression is also not necessarily universal, as different players may respond differently, or not at all, to being placed before the decision to fold versus raise or call. While we focus on folding in this paper, we assume these techniques can be used for other behaviours/events both in poker as well as other settings.

To predict whether a person folds or not based on detected Action Units, we apply sets of decision trees, each operating

978-1-5386-2335-0/18/$31.00

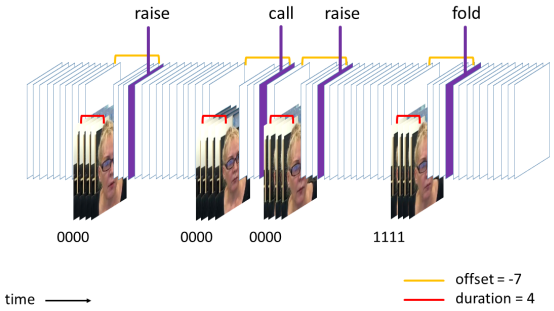

Figure 1. An illustration of which frames of each players game are extracted for a fixed offset/duration pair in order to learn the corresponding tree. In this case, the tree is that corresponding to offset = -7, duration = 4.

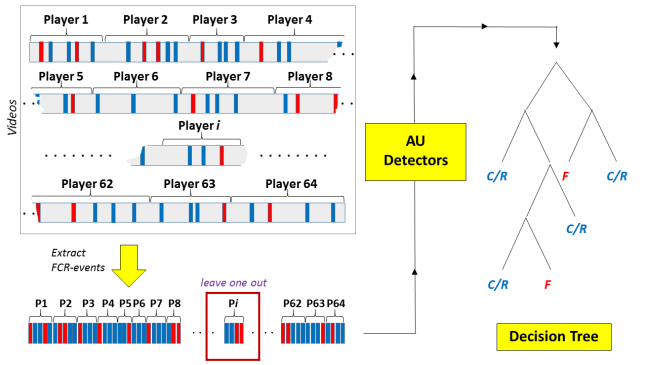

Figure 2. An overview of the method by which we create our decision trees showing the FCR-event frames being extracted from the poker videos, converted frame by frame into 12 action unit values each and then being used to learn a decision tree. To approximate the performance of the tree we use leave-one-subject-out.

on the AUs detected in a single video frame, and trained on data from different subjects only. To improve the results, we combine predictions of multiple adjacent video frames in a time window of a particular length to make a decision on whether a person will fold or not. Both the width and centre of the time window, relative to the fold/call/raise decision time, are explored in detail. All frames in the window of the given decision are labelled to be the positive class if the decision is a fold, and the negative class, that is raise/call, otherwise. Figure 1 gives an overview of this process.

To the best of our knowledge, this is the first study to use computer vision to analyse human behaviour in poker or any other card game. We will show that the detected facial expressions can be used to predict the intended action of a person with an accuracy that is well above chance levels, even with the relatively simple approach of fusing a small number of decision trees, each operating on instantaneous facial expressive behaviour only. In addition, we contribute to the body of knowledge on human behaviour by showing how facial expressions carry more information about intended actions as the time to decide draws near, quantifying for each time period the probabilities of success for predicting actions. Interestingly, the behaviour shortly after the action is made is most telling. Although this is clearly not useful for developing winning game strategies, it is of value to people studying human behaviour.

II. Related Work

In this paper, we use computer vision to study people playing poker. In particular, we ask if we are able to predict when they are about to fold. Poker is a multi-billion-dollar industry [12]. It is considered both a game of chance and a game of skill. As well as its entertainment and business aspects, there is also a lot of interest in poker from the perspectives of mathematics and game theory, as well as from the perspective of psychology. In 2015, the University of Alberta in Canada solved Texas Hold'em with a game-theoretic approach [13]. In 2017, for the first time, two computer programs, DeepStack [14] and Libratus [15], separately beat professional human poker players at Texas Hold'em. This represents a major advance in game theory and AI and has applications in other fields like security and finance.

Poker and other forms of gambling are also important topics in psychology where they are considered to be *"powerful tools for investigating risk-taking, decision making, and how the brain responds to personal gains and losses"* [16]. The poker face was studied in Schlicht et al.[17] and it was discovered that a player deliberates more and makes more betting mistakes when their opponent has a trustworthy face. Slepian et al. [18] looked for visual cues that could betray whether a poker player has a strong hand. The observations used were from humans not computers. They asked three groups of people who were not professional poker players to try and guess the strength of professional players' poker hands by watching very short (2 seconds) silent videos of the players pushing poker chips into to the pot. The first group watched videos of the players' faces only, the second group watched videos of the players' arms only, and the third group watched videos of the players' upper bodies. The group that watched faces only guessed the strength of the players' hands worse than random, implying they were deceived by the players' poker faces. The group that watched the full upper body did as well as, but not better than, random. However, the group that watched arms only did better than random, implying that motor actions can reveal how strong a poker player's hand is.

Perhaps automatic techniques can pick up motor cues in the face without falling into the trap of deception. Loetscher et al. [19] used precise measurement of eye movement to investigate if a person's eye position betrays what number they are thinking of. They had twelve right handed men sit in a dark room and call out 'random' numbers they thought up. After precisely measuring their eye movement, they

concluded that low numbers were associated with leftward and downward positions of the pupil and high numbers were associated with rightward and upward positions. The significance of this result for poker in guessing the value of an opponents hand from visual cues was quickly noted [20].

III. Methodology

We used a database of facial view videos of people playing poker to investigate if we could automatically predict when the players were about to fold just from their facial cues. First, we ran 12 separate action unit detectors on all the videos to produce twelve action unit values for each frame of the video. We chose the times for when the players were faced with the choice to fold, call or raise. From now on we refer to a player choosing between a fold, call or raise as a FCR-event. We aligned the timestamps for these FCR-events over all such events for all players at time $t = 0$. In order to find the period of highest predictability, we tested different offsets and durations. An offset denotes how many frames before t_0 to begin, and the duration signifies how many continuous frames to look at starting with the offset frame. We then extracted these frames and labelled those individual frames belonging to a fold as 1, and those belonging to a call or a raise as 0. These labelled frames were then used to train a decision tree. See Figure 2. We trained a separate decision tree for each offset/duration pair and then looked to see which trees performed the best. We also looked to see if there were periods that had higher predictability than other periods. Out of these results, we concluded when folds are most clearly revealed by the face.

A. The Database

The database used in this study was designed at the Institute for Creative Technologies at the University of Southern California in 2015 [21]. The participants were recruited through Craig's List, which is a website that posts classified ads. They were not professional poker players. The participants were offered $30 to play poker. They did not play for real money, but instead they could win tickets for a lottery worth $100 according to how well they played. The players played in pairs, A versus B, over a local area network (LAN) where they could see each other by means of a computer monitor. However, they could not speak to each other. The games were videoed at 30 frames per second by a webcam embedded in the monitors where the games were depicted for the players and across which the players communicated with each other. Each game produced two videos, a frontal view video of each player, as they were focused on the monitor for most of the game. See Figure 3.

The game was a simplified version of poker. It consisted of ten rounds of poker, where in each round the players were each dealt a single card whose value was between 2 and 10. As is typical in poker, the participants did not get to see their opponents' cards. The object of the game was to bet (add to the pot) a numerical value such that the player maximizes their wins and minimizes their losses. A player wins if a call occurs and their card is higher or their opponent folds, that is, chooses to give the round to their opponent in order to avoid a bigger loss. Player A went first on odd numbered rounds, player B went first on even numbered rounds. The first bet in a round could be zero or a positive bet, which meant the first play in the game could not be a call or fold. Otherwise, when it was the player's turn, they had the choice to either call, raise or fold. See Figure 4. Although from each player's perspective the card order seemed random, the sequence of ten cards was always fixed. All A players saw the same ten cards as did all B players. This was not a problem since no player played in more than one game. After rounds three and seven the players were given a brief, multiple choice questionnaire about what they thought of their opponent's betting strategy. This lasted usually around 20 seconds.

Since the games took place over a LAN and players made bets by means of mouse clicks, the major events of the game could be automatically timestamped. There were timestamps for the beginning of each round, when each of the bets were placed, and when the two short questionnaires occurred. Bet amounts and the answers to the questionnaires

Figure 3. The poker game as seen by a player, who is in the lower left corner of the image of their opponent.

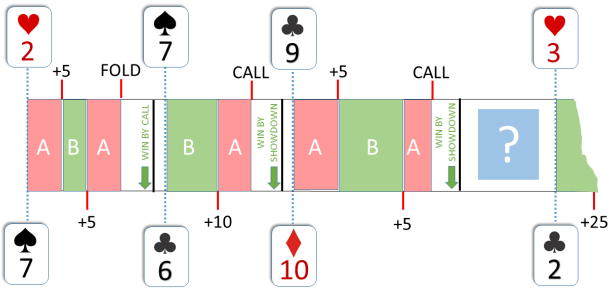

Figure 4. An excerpt from a game depicting rounds 1,2,3 and the beginning of round 4, as well as the questionnaire, represented by the blue square occurring between rounds 3 and 4. The upper sequence represents the sequence of cards and moves of Player A, the lower sequence represents those of Player B.

were also recorded. There were no other annotations to the database, nor were there later any added for this study, except for those of the action unit detectors, see below. Therefore, the database consisted only of the images of the videos and the timestamps and bet values, all of which were created automatically. Originally, there were altogether 104 videos. The 104 videos did not pair up into 57 game pairs. This was due to some participants not having giving their permission to share their videos. Forty of the individual participants' videos were recommended by the Institute for Creative Technologies for removal. This was due to various reasons: they involved confederates, the video quality was bad, there had been technical difficulties, or the player did not understand the game. We removed all forty of these before carrying out this study. Therefore, this study includes 64 videos.

B. The Action Unit Detectors

Altogether, over the 64 videos of the poker games, there are 675,432 frames of video, each frame consisting of 640×480 pixels. To reduce the complexity of the data while retaining relevant information, we ran twelve action unit detectors separately on the videos, producing twelve separate action unit intensity values for each frame of video. Thus, we replaced each frame of 640×480 pixels with a frame of 12 real-valued numbers between 0 and 1.

We created these automatic annotations for our database of poker videos using state of the art Action Unit (AU) Detectors by Jaiswal et al. [22] . These detectors use Convolutional and Bi-directional Long Short-Term Memory Neural Networks to learn a subset of action units in the Facial Action Coding System. The Facial Action Coding System was developed by Ekman and Friesen [23] in order to provide a method to systematically and objectively describe facial expressions in terms of the facial muscles that are activated when they occur. They are the basis of most sign-based methods of Automatic Facial Analysis. The idea of using facial muscle descriptions was introduced in 1872 by Charles Darwin [2] and his contemporary and collaborator Pierre Duchenne de Boulogne [1], who used this method to describe human facial expressions more precisely. In building the AU detectors, Jaiswal et al. used the SEMAINE database [24] as well as the DISFA [25] and the BP4D [26] databases to train their detectors. The SEMAINE database consists of videos of actors, sometimes represented by a virtual avatar, interacting with a non-actor. Where the actors try to drive an emotionally rich conversation, their non-actor counterparts respond spontaneously and naturally given the context. The BP4D and DISFA databases consist of videos each of which shows a human reacting spontaneously and naturally to emotion inducing stimuli. This differs from most databases where emotions are acted out by professional actors in an exaggerated way. That the AU detectors we chose were trained using more naturalistic data is an important characteristic for this work as we are analysing subtle and spontaneous natural behaviour.

Table I
SHORT DESCRIPTION OF THE 12 AUS USED.

Action Unit	AU Description
AU1	Inner Brow Raiser
AU2	Outer Brow Raiser
AU4	Brow Lowerer
AU5	Upper Lid Raiser
AU6	Cheek Raiser
AU9	Nose Wrinkler
AU12	Lip Corner Puller
AU15	Lip Corner Depressor
AU20	Lip Stretch
AU25	Lips Part
AU26	Jaw Drop
AU45	Blink

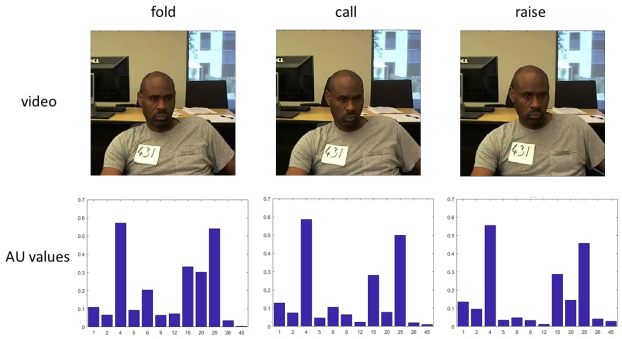

Figure 5. This figure shows the video frames along with their corresponding detected action unit intensities corresponding three different events: the player folds, calls and raises.

There are two types of action unit detectors presented in Jaiswal et al. [22], occurrence detectors and intensity estimation detectors. We chose to work with the intensity estimation detectors as we are interested in the different magnitudes to which the facial muscles are stimulated. We chose the 12 intensity estimation action unit detectors listed in Table I. We chose to work with these action units as they were the best performing of those available, not necessarily because we expected all in this group to be the most relevant. We intentionally tried not to make assumptions about this. Figure 5 shows the action unit intensities detected on a player at three instances: when he folds, calls and raises.

C. Decision Trees

We applied the twelve action unit detectors to the videos and replaced each frame of image pixel values with twelve real values between 0 and 1. After this, we did not use the original video images again. The appropriate frames (now consisting of only twelve real values) corresponding to FCR-events were extracted and labelled 1 if they were in a fold

event and 0 if they were in a call or raise event. We then used these labelled frames to learn a decision tree. We chose to use decision trees because of their clear conception, the speed with which they can be learned and their ability to pick out useful information while filtering out noise.

We used our labelled frames, which now consisted of twelve AU values as attributes, as input to learn the decision trees. The decision trees were the usual, recursively built, binary trees. Decision trees are well-established machine learning techniques. There are many variations of them. For a good introduction see Breiman [27] or Mitchel [28]. We use the CART version in [27].

Our tree is built using questions of the form, "Is the value of a particular action unit greater than some threshold?", where a search is made over all possible thresholds over all 12 action units to find the best threshold, or split of the data, at the current node. The CART algorithm reduces the impurity at child nodes by maximizing

$$\Delta i(s,n) = i(n) - p_L \cdot i(n_L) - p_R \cdot i(n_R)$$

over all possible candidates splits s. The current node where one is splitting is denoted n. The proportion of instances that will go to the left child according to split s is denoted p_L, the proportion of instances that will go to the right child according to s is denoted p_R. CART uses the Gini Index of Diversity to define the purity of a node n. The Gini Index of Diversity is defined as

$$i(n) = 2 \cdot p(1|n) \cdot p(0|n).$$

Here, $p(1|n)$ denotes the number of class 1 examples at node n and $p(0|n)$ denotes the number of class 0 examples at node n.

CART splits downward by querying the twelve AU values of each instance that is at the current node until it reaches its stopping criteria. Here, splitting stops when there is no further improvement in impurity possible, or when the node has too few instances and splitting might cause overfitting.

D. Action Unit Detectors and Decision Trees

We combined the action unit detectors with decision trees to discover if there is a facial expression, or a small enough set of facial expressions, common to enough to different players as to allow a classifier to predict whether a player was going to fold or not. We wanted to address the question whether such expressions exist and when and for how long their signals are strongest. Therefore, we searched a space of different offsets and durations centred around the players' FCR-events by building a separate decision tree for each *offset/duration* pair. Here *offset* refers to the time of the first frame of the window relative to the event and *duration* refers to the length of the window being considered. These two parameters determine the starting frame and how many continuous frames after this are used relative to each FCR-event to build the given tree.

More precisely, if e is a frame corresponding to a players choice to fold, call or raise, and if the current offset is o and the current duration is d, then the d frames $e+o$ to $e+o+d-1$ are labelled 1 if e is a fold and 0 if e is a call or a raise. This labelled set of frames is added to the set of those used to learn the current tree and the current tree differs from other trees only in its offset/duration parameters.

IV. Evaluation

We searched a space of decision trees $T_{\textit{offsets,durations}}$ over different offset and duration pairs in order to discover if and when folds could best be predicted. The distance between some FCR-events was just over nine seconds. In order to avoid overlapping events we focused on decision trees learned on frames that did not precede their corresponding FCR-event by more than nine seconds (-270 frames). Four seconds after a round ended with a fold or call, the players were dealt their cards for the next game, or the game was over. Therefore, we restricted our offsets to four seconds after the FCR-event. Since there were as few as nine seconds between rounds, only durations of nine seconds or less were considered. Therefore, the search space was restricted to offsets in the range of 9 seconds before the FCR-event to 5 seconds after and to durations of $\frac{1}{30}$ of a second (one frame) to 9 seconds.

In order to evaluate the performance of each of the decision trees, we used leave-one-player-out. So, for each offset/duration pair and for each of the 64 players p, we learned a decision tree without the frames for that player making that particular decision tree independent of that player. We then used the tree to classify the left out player's frames. In this way, we collected all the classifications for all the players and used these to estimate the performances of each $T_{\textit{offset,duration}}$.

For our performance measure, we chose to use the classification rate which we define as

$$R = \frac{C_1 + C_0}{N}.$$

Here C_1 is the number of correctly labelled fold instances, C_0 is the number of correctly labelled call/raise instances, and N is the total number of instances. There are 481 FCR-events in the database. Of these, 132 are folds, 184 are calls, and 165 are raises. The ratio of fold events to call and raise events is therefore 1:2.65. This is also the ratio of fold to call/raise instances used to learn and test any given tree since the number of instances used to learn a tree is just 481$\times$duration. In this case, simply assigning every instance the class 0 (call/raise) gives a classification rate of 0.73. However, we were interested how well a classifier can distinguish between a fold and a call/raise and we were expecting the baseline for the problem to be low.

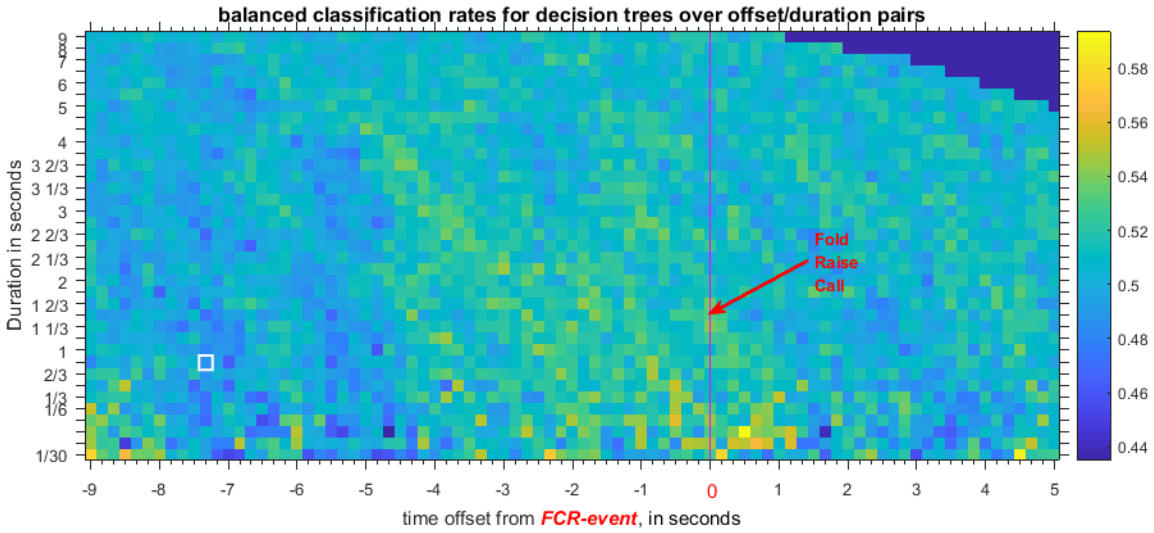

Figure 6. Each square in the heat map shows the balanced classification rate of a decision tree. Each entry represents a decision tree trained for a particular offset (x-axis) and window length (y-axis). The x-coordinate 0 represents the time of the FCR-event. Left and right of this are offsets in seconds (30 frames per second). The y-axis represents different durations in seconds increasing in ascending order. For example, the white rectangle at $x = -7\frac{1}{3}$ and $y = \frac{5}{6}$ gives the balanced performance ratio of $T_{-7\frac{1}{3},\frac{5}{6}}$, which is 0.4856.

Therefore, we have scaled the folds to have equal weight as the call/raises. This has led us to a balanced version of the classification rate, which we interpret as the classification rate in the case that in the test data the number of folds equals the number of calls together with raises. We define R^b as

$$R^b = \frac{P_1 + P_0}{2}.$$

Here, P_1 is the proportion of correctly labelled class 1 examples and P_0 is the proportion of correctly labelled class 0 examples. Similarly, we computed the balanced precision, recall and F1-measure for the trees. When the classes are balanced this gives

$$Precision^b = \frac{P_1}{P_1 + (1 - P_0)},$$

$$Recall^b = P_1$$

and

$$F1^b = 2 \cdot \frac{Precision^b \cdot Recall^b}{Precision^b + Recall^b}.$$

We show the results for the best decision tree, $T_{15,3}$, which occurs at *offset*=15 frames and *duration*=3 frames, in Table II. In this case, there are 1443 frames. Our method of leave one player out labelled 170 of the of the 396 class 1 frames correctly and 794 of the 1047 class 0 frames correctly.

We also looked into the statistical significance of our results. We compared the balanced classification rate of our classifier with a fair coin, that is, one that has a probability of $\frac{1}{2}$ of landing heads, and we used the binomial distribution

Table II
SHORT DESCRIPTION OF THE 12 AUS USED.

	regular	balanced
Classification rate	0.6681	0.5938
Precision	0.4019	0.6398
Recall	0.4293	0.4293
F1-measure	0.4151	0.5138

with the statistical significance level .99. Instead of using individual frames, which are clearly dependent, we considered different FCR-events to be independent of each other, as they never overlap and usually have a gap of several seconds between each other. There were altogether 481 such events. We considered the decision tree with the best performance, which occurs at 15 frames (half a second) after the player makes their choice to fold or not, and has a duration of 3 frames (a tenth of a second). Its balanced classification rate is .5938, which according to our significance test is significant. However, if one considers the worst classifier, which occurs just short of five seconds before the event and has a duration of 3 frames, as a negative classifier, it has a balanced classification rate of 0.5651, which, though much lower than the classification rate of .5938 of the best classifier, is also statistically significant. This points out the difficulty of analysing human data when the baseline is low, as we believe it is in this case of subtle human expression.

We also created a heat map of the decision trees, see Figure 6, with offset values increasing along the x-axis and duration values increasing along the y-axis. We did this in order to discover if there were consistent areas where

the detectors could better pick up folds versus calls and raises, which would indicate if and when there was a signal. The heat map can also indicate if the performance of the detectors makes sense in terms of human behaviour, or if it is random. The heat map shows that just more than four seconds before the FCR-event the classifiers begin to perform better, the performances peak around half a second after the FCR-event and rapidly decline at 1.5 seconds after the event. In order to view this from a different angle, we also created a bar graph, Figure 7, which serves as a cross section of the heat map shown in Figure 6. It also depicts the performance of the classifiers as time elapses using the identical offset schema as in Figure 6, but this time, for each offset i we plotted the best performance ratio from among the five classifiers made from durations of 1 to 5 frames. Adjacent classifiers made with longer durations otherwise begin to contain overlapping frames. Figure 7 thus shows nearly the same data as Figure 6, focused on the best classifiers at each offset. One can see from both that the ability to detect the fold versus the call/raise increases four seconds before the event, peaks at half a second after and decreases rapidly at 1.5 seconds after. At 9 seconds before FCR-events, the players are once again in FCR-events. At 4 seconds after an FCR-event, the players are receiving their next cards. Since the card order for the players is fixed, we believe the decision trees at these ends of the heat map are detecting other correlations between the games of the players.

V. Conclusion

We wanted to see if we could predict a player's objective actions from only their facial actions. In particular, we wanted to see if we could predict a fold. The strongest prediction we obtained was half a second after the event at $T_{15,3}$. However, in the four seconds leading up to the FCR-event, we also obtained classifiers with classification rates that were statistically significant at the 5% level; $T_{-100,4}$, just over 3 seconds before the event, had balanced classification rate 0.55, and $T_{-70,1}$, just over 2 seconds before the event, had balanced classification rate 0.57. We conclude that it is possible using action unit detectors together with decision trees to predict and detect subtle and spontaneous actions of humans.

Acknowledgment

This research was supported by the National Science Foundation under grant 14196221 and the US Army, the NIHR Nottingham Biomedical Research Centre and the National Institute for Health Research. The views represented are the views of the authors alone and do not necessarily represent the views of the Department of Health in England, NHS, the National Institute for Health Research or reflect the position or the policy of any Government, and no official endorsement should be inferred.

References

[1] Pierre Duchenne de Boulogne. *The Mechanism of Human Facial Expression*. Cambridge University Press; New Ed edition (12 Jan. 2008).

[2] Charles Darwin. *The Expression of Emotions in Man and Animals*. 200th Anniversary Edition, HarperCollinsit Publishers, 2009. 2016.

[3] P. Ekman, W. Friesen. "Unmasking the Face: A guide to recognizing emotions from facial expressions", Malor Books (1 April 2003).

[4] J. Cohn, F. De la Torre. *Automated Face Analysis for Affective Computing*. "Oxford Handbook of Affective Computing". OUP USA (15 Jan. 2015).

[5] M. Valstar. "Automatic Behaviour Understanding in Medicine", Proceedings of the International Conference on Multimodal Interaction (workshop section), 2014.

[6] Schere, S., G. Stratou,, G. Lucas, M. Mahmoud, J. Boberg, J. Gratch and L.-P. Morency (2014). "Automatic Audiovisual Behavior Descriptors for Psychological Disorder Analysis." Image and Vision Computing **32**(10): 648-658.

[7] J. Girard, J. Cohn, M. Mahoor, S. Mavadati, D. Rosenwald. "Social Risk and Depression: Evidence from Manual and Automated Facial Expression Analysis". Proc Int Conf Autom Face Gesture Recognit. 2013 : 18.

[8] S. Jaiswal, M. Valstar, A. Gillot, D. Daley. "Automatic Detection of ADHD and ASD from Expressive Behaviour in RGBD Data". Face & Gesture 2017.

[9] J. Egede, M. Valstar, B. Martinez. "Fusing Deep Learned and Hand-Crafted Features of Appearance, Shape, and Dynamics for Automatic Pain Estimation", 2017 IEEE 12th International Conference on Automatic Face & Gesture Recognition.

[10] Hoegen, R., G. Stratou and J. Gratch (2017). "Incorporating emotion perception into opponent modeling for social dilemmas. 16th International Conference on Autonomous Agents and Multiagent Systems." Sao Paulo, Brazil.

[11] Park, S., J. Gratch and L. P. Morency (2012). "I already know your answer: using nonverbal behaviors to predict immediate outcomes in a dyadic negotiation." Proceedings of the 14th ACM international conference on Multimodal interaction, ACM.

[12] "Poker: A big deal", The Economist, December 2007.

[13] Bowling, M., Burch, N., Johanson, M. and Tammelin, O. "Heads-up limit holdem poker is solved". Science, 347(6218), pp.145-149, 2015.

[14] M. Moravk, M. Schmid, N. Burch, V. Lis, D. Morrill, N. Bard, T. Davis, K. Waugh, M. Johanson, M. Bowling. "DeepStack: Expert-level artificial intelligence in heads-up no-limit poker", Science, March(2017).

[15] O. Solon."Oh the humanity! Poker computer trounces humans in big step for AI". The Guardian, 30 January, 2017.

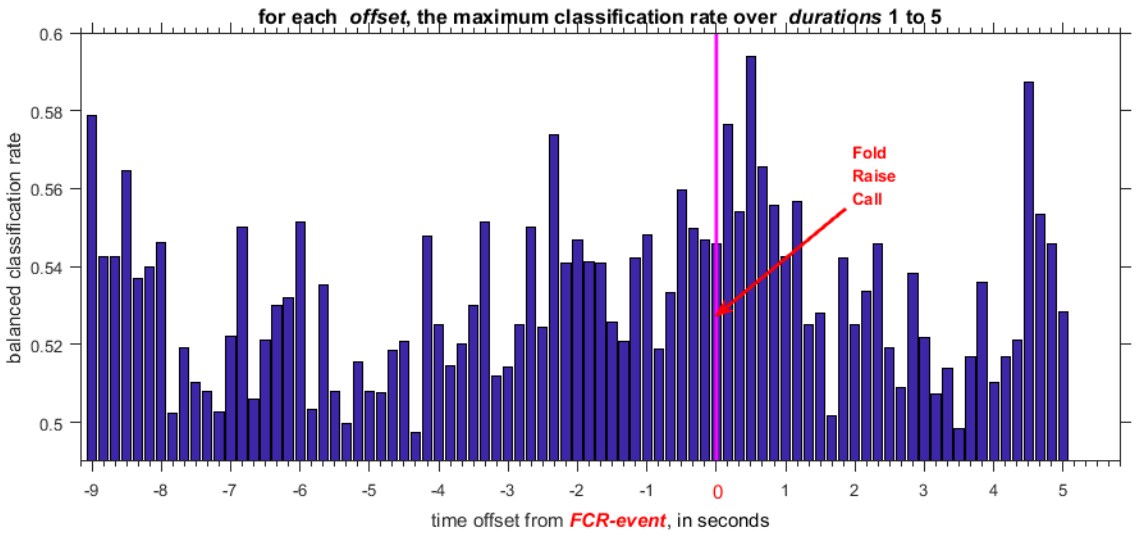

Figure 7. The x-axis of this figure is identical to the x-axis of Figure 6. Each bar in the bar graph corresponds to the best performance ratio of a decision tree at that offset from among the five durations of {1,2,3,4,5} frames.

[16] F. Jabr. "Two of a Kind: Studies Reveal New Insights into the Psychology of Gambling". Scientific American. August 12, 2010.

[17] E. Schlicht, S. Shimojo, C. Camerer, P. Battaglia, K. Nakayama. "Human Wagering Behaviour Depends on opponents' Faces". Plos One July 2010, Volume 5 Issue 7, e11663

[18] M. Slepian, S. Young, A. Rutchick, N. Ambady. "Quality of Professional Players Poker Hands Is Perceived Accurately From Arm Motions", Psychological Science, 24(11) 23352338 (2013).

[19] T. Loetscher, C. Bockisch, M. Nichols, P. Brugger. "Eye position predicts what number you have in mind". Current Biology , Volume 20 , Issue 6 , R264 - R265. 23 March 2010.

[20] University of Melbourne. "Poker face busted? Our eye position betrays the numbers we have in mind, new study", ScienceDaily, 24 March 2010.

[21] Ryan Lu, Sarvesh Pantage. "Emotion Modeling in Poker". University of Southern California, unpublished report, 2015.

[22] S. Jaiswal, M. Valstar. "Deep Learning the Dynamic Appearance and Shape of Facial Action Units". *Winter Conference on Applications of Computer Vision (WACV)*, accepted, 2016.

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

[24] G. McKeown, M. Valstar, R. Cowie, M. Pantic, and M. Schroder. "The semaine database: Annotated multimodal records of emotionally colored conversations between a person and a limited agent". IEEE Transactions on Affective Computing, 3:517, 2012.

[25] S. Mavadati, M. Mahoor, K. Bartlett, P. Trinh, J. Cohn. "DISFA: A Spontaneous Facial Action Intensity Database". IEEE Transactions on Affective Computing, Vol. 4, NO. 2, April-June 2013.

[26] S. Mavadati, M. Mahoor, K. Bartlett, P. Trinh, J. Cohn. "BP4D-Spontaneous: a high-resolution spontaneous 3D dynamic facial expression database". Image nd Vision Computing, 32, 2014, 692-706.

[27] L. Breiman, J. Friedman, R. Olshen, C. Stone. *Classification and Regression Trees*. Chapman and Hall/CRC; 1984.

[28] Tom Mitchell, *Machine Learning*, McGraw-Hill Science/Engineering/Math; (March 1, 1997).