Using Electrodermal Activity to Detect Deception and Suspicion during a Card Game

Jan Ondras
Department of Computer Science and Technology, University of Cambridge, Cambridge, UK

***Abstract*— In this project I focus on detection of deception and suspicion from electrodermal activity (EDA) measured on left and right wrists during the card game called *Cheat*.**

I aim to answer three research questions: (i) Is it possible to reliably distinguish deception from truth based on EDA measurements during a card game? (ii) Is it possible to reliably distinguish the state of suspicion from trust based on EDA measurements during a card game? (iii) What is the relative importance of EDA measured on left and right wrists? To answer my research questions I conducted a study in which 20 participants were playing the game *Cheat* in pairs with two EDA sensors placed on their wrists.

My experimental results show that EDA measures from left and right wrists provide more information for suspicion detection than for deception detection and that the person-dependent detection is more reliable than the person-independent detection. In particular, classifying the EDA signal with Support Vector Machine (SVM) yields accuracies 52% and 57% for person-independent prediction of deception and suspicion respectively, and 63% and 76% for person-dependent prediction of deception and suspicion respectively. Also, I found that: (i) the optimal interval of informative EDA signal for deception detection is about 1 s while it is around 3.5 s for suspicion detection; (ii) features extracted from EDA from both hands are important for classification of both deception and suspicion; and that (iii) EDA measured from left wrist is more dominant (as compared to the right) during the game for most participants and seems to be independent of player handedness.

To the best of my knowledge, this is the first work that uses EDA data to automatically detect both deception and suspicion during a card game.

I. INTRODUCTION

Electrodermal activity (EDA) is a widely used indicator of sympathetic nervous system (SNS) activity and is often used to describe the degree of a persons excitement, stress, anxiousness, as well as changes in arousal related to pain, anticipation, and other feelings that may be of interest [1]. EDA is also known as skin conductance or galvanic skin response and these terms will be further used interchangeably. Traditionally, EDA measurements involved attaching wired and gelled electrodes to the skin [2]. Recently, unobtrusive wearable devices such as the wireless Affectiva Q Sensor [3] used in my study have attained popularity.

EDA can be also used as an indicator of deceit since lying costs more mental effort than telling the truth [4] and the cognitive load activates SNS [5], [6]. In particular, the previous research [7], [8] found increased EDA for lying as compared to truth telling. Moreover, EDA is an autonomic-based physiological response which makes it hard to control and therefore less susceptible to strategic manipulations [9] and so it is a good indicator of deceit.

Suspicion (or trust) detection is a more recent and less researched field and to the best of my knowledge there was no attempt at detecting suspicion or trust from EDA. However, the state of suspicion is often associated with an increased cognitive load and stress as compared to the state of trust [10], [11] which along with the aforementioned findings about EDA suggests that it should be possible to detect suspicion based on skin conductance measurements.

Many studies have measured the presence of EDA asymmetry on the left and right palms [12], [13], [14] and attempted to relate bilateral EDA measures to verbal/spatial, positive/negative, emotional/nonemotional specialization, with conflicting findings. The classic understanding assumed that EDA represents one homogeneous change in arousal across the whole body, but recent works [15], [16] show that multiple brain structures contribute to elicitation of EDA, namely, two regions were identified: a limbichypothalamic source (EDA1) and a premotor-basal ganglia source (EDA2). The EDA1 system includes structures (such as amygdala, cingulate gyrus, anterior thalamus, fornix, hippocampus, and hypothalamus) that play a strong role in emotions and it is believed to be ipsilateral[1] [17] which means that activating key emotion regions on the right side of the brain (e.g. right amygdala) produces right palmar EDA activation and analogously for the left side. In contrast, the EDA2 system (including the basal ganglia and premotor cortex) is contralateral[2]. Together with findings that amygdala is the emotional center of the brain and that the left hemisphere primarily processes positive emotions while the right hemisphere processes primarily negative emotions [18], it might be tempting to conclude that higher skin conductance found on right hand reflects negative emotional state and analogously higher skin conductance on left hand reflects positive emotional state. Such conclusions may be flawed since the measured EDA might be also influenced by the other source of arousal (EDA2), for example, by hand movement. With the current knowledge the plausible suggestion that can be made is that underlying negative emotions (such as fear or anxiety) only *contributed* to greater right amygdala activation and thus to the EDA on right hand. Since EDA measures on only one side may lead to misjudgment of arousal [16], I investigate EDA signals from both hands.

In this work I hypothesise that EDA data obtained from

[1]Occurring on the same side of the body.
[2]Occurring on the opposite side of the body.

left and right wrists can be effectively used to detect deception and suspicion during the card game *Cheat*. Specifically, I aim to answer the following research questions: (i) Is it possible to reliably distinguish deception from truth based on EDA measurements during a card game? (ii) Is it possible to reliably distinguish the state of suspicion from trust based on EDA measurements during a card game? (iii) What is the relative importance of EDA measured on left and right wrists? I investigate these questions by conducting a controlled experiment where 20 participants are asked to play a two-player variant of the card game *Cheat* and answer a post-study questionnaire. In this context, I define the deception to be the action when a player discards a card different from what he/she claims and suspicion as a state when a player does not trust the opponent that the card discarded by the opponent was the same as he/she claimed. To summarise, my work has the following contributions:

- First of its kind dataset (named *DSDEDA*) for automatic deception and suspicion detection based on measurements from wearable EDA sensors, collected from 20 participants along with their personality traits (for future research).
- System for automatic detection of deception and suspicion from EDA measurements on-the-fly.
- Findings that:
  - EDA measures from left and right wrists provide more information for suspicion detection than for deception detection and that the person-dependent detection is more reliable than the person-independent detection;
  - optimal interval of relevant EDA signal for deception detection is about 1 s while it is around 3.5 s for suspicion detection;
  - features extracted from EDA from both hands are important for classification of both deception and suspicion;
  - EDA measured from left wrist is more dominant (as compared to the right wrist) during the game for most participants and seems to be independent of player handedness.

The rest of the report is structured as follows. Section II lists the related work in the field, Section III describes the details of the conducted study, and Section IV shows the feature extraction process. Next, the classification experiments and obtained results are presented in Section V while Section VI shows importance of individual features and asymmetry in EDA. Section VII discusses the results and Section VIII concludes the report and foreshadows future research directions.

II. RELATED WORK

A. EDA

EDA has been widely used for various tasks such as seizure detection [19], engagement recognition during social interactions [20], analysis of EDA during sleep [21], or depression prediction based on EDA asymmetry [22].

B. Deception detection

The detection of deception (or lie detection) is a long-standing problem addressed by many studies using various sources of information for detection. Neurophysiological signals such as Functional Magnetic Resonance Imaging (fMRI) [23], [24] and Event Related Potentials (ERP) [25] were investigated for this task. For instance, the work [26] used electroencephalography (EEG) features extracted through wavelet transformation and they achieved the correct detection rate of 86%. Another brain-imaging technique, functional near-infrared spectroscopy (fNIRS), which measures brain activity through hemodynamic responses associated with neuron behavior was also examined [27].

Other approaches used cues from videos, for example, [28] built an automated system that can infer deception or truthfulness from a set of features extracted from head and hands movements captured in a video, yielding 71% classification accuracy using both Support Vector Machine and a neural network.

However, in most lie detection settings several physiological signals including respiration, skin conductance, blood pressure, and pulse rate were employed [29]. The study [30] reports 74.5% accuracy when using physiological measures and accuracy of 86.5% when combined with fNIRS measurements.

Several works tried to detect deception during games. As a case in point, the study [31] based the deceit detection on EEG measures during a poker-like card game. Others [32], [33] investigated EDA in gaming but not for deception detection. For instance, Drachen et al. [33] researched correlations between heart rate, EDA and player experience in first-person shooter games. However, to the extent of my knowledge there was no previous work that would attempt at detecting deception from EDA and in a game environment.

C. Suspicion detection

Suspicion or trust detection is a more recent and less researched field. Previous work focused mostly on videos and analysed nonverbal behaviors achieving even above human accuracy [34]. However, to the best of my knowledge there was no attempt to detect suspicion or trust from EDA.

III. THE STUDY

A. Motivation

My motivation to choose a card game for detection of deception and suspicion was two-fold. Firstly, games are a very common scenario when people lie and are suspicious without negative consequences. Secondly, a game with real participants and their direct interactions allows more natural reactions of players as compared to computer-based games, and thus is expected to provide more realistic results. With this motivation I designed a study to answer the three research questions described in Sec. I.

B. Card game Cheat

I choose the card game *Cheat* (also called *Bluff* or *I Doubt It*) as it has very simple rules which allows players to better focus on their actions of deception and suspicion. To simplify the experiment I focused on a two-player variant of this game with the following rules.

All cards[3] are evenly dealt out to the two players and they can see their own cards. Goal of the game is to get rid of all cards at hand. First player calls out the suit (diamonds/clubs/spades/hearts) and discards one card face down on the discard pile. The suit called out by the first player is the *true suit* for the current discard pile. Players then take alternate turns discarding one card each time and calling out the same suit as the player in the first turn. Since the cards are discarded face down, players can cheat to win faster by discarding a card of different suit than required. If one player suspects the other player he/she can challenge the play by calling "Cheat!". Then the card played by the challenged player is exposed and one of two things happens: (i) if the exposed card is of the suit that was called, the challenger must pick up the whole discard pile; or (ii) if the card is different from the called suit, the person who played the card must pick up the whole discard pile. The player who did not pick up the pile begins the new round by discarding one card and calling out a suit that becomes a true suit for the new discard pile. The game ends when one of the players gets rid of all his/her cards at hand.

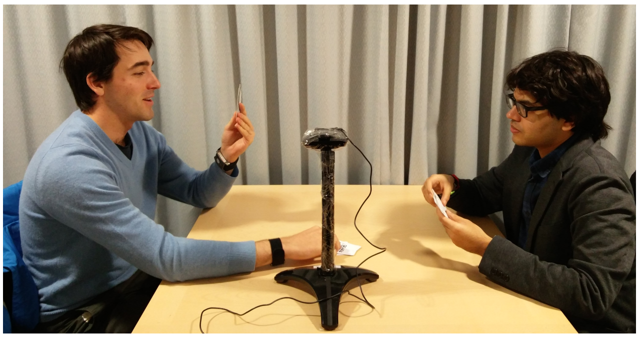

Fig. 1. Experimental setup: participants playing the card game *Cheat*. Each player is wearing two Affectiva Q EDA sensors (one on left and one on right wrist). The discard pile is recorded by web-camera.

C. Sensors

As can be seen in Fig. 1, each player was wearing two Affectiva Q EDA sensors [3] (one on left and one on right wrist) during the game. The Q sensor measures electrical conductance (in units of μS) across the skin by passing a minuscule amount of current between two electrodes that are in contact with the skin. Each Q sensor provides the following data: EDA, skin temperature, and 3-axis of acceleration of the wrist over time. These data were sampled at 32 Hz and I decided to use only EDA measurements for deception and suspicion detection.

The discard pile of cards was recorded by a web-camera (along with the audio) and the back side of every card was labeled with QR code corresponding to the suit on the other side of the card. This allowed later reconstruction of events (whether a player lied or told truth while discarding a card) and their localisation in time which was necessary for correct annotation of EDA data.

Prior to the experiment all four Q sensors were time-synchronised with the system time used by the web-camera.

D. Data collection

I built an in-house dataset named *DSDEDA* (Deception and Suspicion Detection from EDA) by recruiting 20 participants (5 female and 15 male) to play the game *Cheat* and to answer a post-study questionnaire.

The participants were aged 19–32 and came from various cultural and educational backgrounds. All but three participants said they were right-handed (participants with IDs 09, 11, 20 were left-handed). The participants were arranged into 10 pairs so that each participant played the game only once. Before the experiment they were informed about its procedure, game rules, and their rights by verbal introduction and through a signed consent form. However, the true goal of this study was not disclosed to them until after the experiment in order to avoid artificial behaviour (e.g. reluctance to bluff and blame the other player) and to allow more natural reactions during the game. Prior to the data collection phase the participants were given time to freely play the game to familiarise with its rules. In order to settle down the measured EDA values, before the game started, participants rested for a 2-minute baseline period while listening to relaxing music. Then, one or multiple games up to a maximum duration of 30 minutes were recorded (4 streams of EDA and video recording of the discard pile).

In the post-study questionnaire the participants were asked to provide information about their gender, age, handedness[4], and take a short personality test[5]. Also, they answered 3 self-report questions: *How many times did you play this game before? On average out of 10 opportunities to lie, how many times do you think you really lied? On average out of 10 opportunities to say "Cheat", how many times do you think you really said "Cheat"?*

E. Data segmentation and annotation

Firstly, I defined three event types: deception event (player discarded card lying), truth event (player discarded card telling truth), and suspicion event (player said "Cheat"). Using the audiovisual recordings of the discard pile I determined the times of these three events for all players and manually annotated EDA measurements as shown in Fig. 2.

Next, the time intervals (epochs) from which the EDA signal was used to extract features for associated events were determined. This process was motivated by the work [8] that demonstrates that the cues in EDA are present before the event of interest. The segmentation procedure for various types of epochs follows.

[3] A stripped deck (US) or shortened pack (UK) of 32 cards.

[4] Tendency to use either right or left hand more naturally than the other.

[5] 20-item measure of BIG5 personality (in terms of Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism), available at: https://discovermyprofile.com/miniIPIP/introduction.html

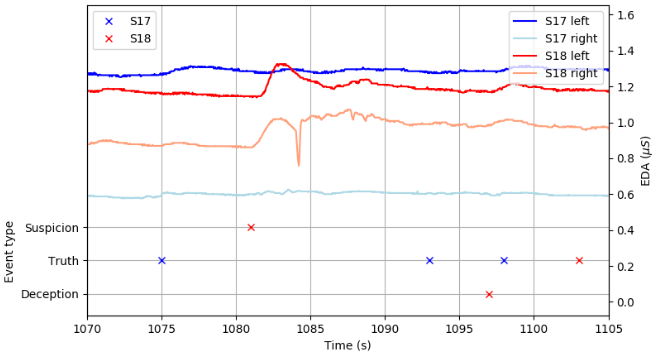

Fig. 2. Annotated EDA signals from left and right hands of participants with IDs 17 and 18. EDA is annotated with three event types (Suspicion, Truth, and Deception event). The suspicion event is followed by a strong response in EDA from both hands of the corresponding player while the deception event is followed by much weaker response.

1) Deception and truth epochs: The *deception epoch* associated to the deception event k (or the *truth epoch* in case of the truth event) was defined as the time interval

$$[\max(t_{k-1}, t_k - \tau_{MDEL}); t_k]$$

where t_k is the time of the associated deception event and t_{k-1} is the time of the previous event (if the previous event was not deception/truth event, then this epoch along with the event was ignored since it was the first one in the game or the first one after the pile was picked up and it contained lots of noise as the players were not well focused yet). τ_{MDEL} is the *maximum deception epoch length* and it prevents epochs from being too long. This thresholding was necessary because in most cases long time between two consecutive events $k-1$ and k meant a player's distraction in the earlier stage of the time interval between these two events. Since it was not clear how to set the parameter τ_{MDEL}, I used cross-validation to find its optimal value in the discrete range $\tau_{MDEL} \in \{0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5\}s$. This range was estimated by looking at the distribution of lengths of time intervals between two consecutive deception/truth events over all players.

2) Suspicion and trust epochs: The *suspicion epoch* associated to the suspicion event k was defined as the time interval

$$[\max(t_{k-1}, t_k - \tau_{MSEL}); t_k]$$

where t_k is the time of the associated suspicion event and t_{k-1} is the time of the previous deception event. τ_{MSEL} is the *maximum suspicion epoch length* and it serves the same purpose as the threshold τ_{MDEL} and its value was also chosen by cross-validation over the discrete range $\tau_{MSEL} \in \{0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5\}s$. Analogously, this range was estimated by looking at the distribution of lengths of time intervals between deception event and following suspicion event over all players.

There is no observable event that would mark the end of a trust epoch (when the player did not say "Cheat" and trusted the other player). To tackle this issue I made an assumption that the length of the trust epoch is approximately τ_{MSEL} if the next deception event is far enough (namely, further than $2 \times \tau_{MSEL}$ from the start of the trust epoch) and it is the half of the distance between the start of the trust epoch and the next event otherwise. In other words, the *trust epoch* following the deception event $k-1$ was defined as the time interval

$$[t_{k-1}, \min(t_{k-1} + \tau_{MSEL}, \frac{t_k + t_{k-1}}{2})]$$

where t_k is the time of the next event. The end of the trust epoch can be thought of as an imaginary trust event.

Lastly, both endpoints of each epoch (of all four types) were delayed by δ because there is a delay between a stimulus and skin conductance response [35]. This is also confirmed by Fig. 2 where we can observe that the response in EDA is delayed after event occurrence. The value of the delay δ was observed to be 1–4 s, and so I determined its value by cross-validation over the discrete range $\delta \in \{1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0\}s$.

Each epoch was then labeled according to its type resulting in the distribution shown in Tab. I. The labeled dataset excludes the deception and truth epochs from one participant (with ID 05) who confused card suits during the game which made me unable to correctly label his/her epochs (this was not a problem for suspicion detection task and so the suspicion and trust epochs of this participant were kept). As we can see, the collected data is unbalanced with a strong bias towards trust labels for suspicion detection task. To mitigate this issue, random over-sampling technique was applied to the data (as further described in Sec. V).

TABLE I

DISTRIBUTION OF FOUR TYPES OF LABELED EPOCHS BEFORE BALANCING.

	Deception	Truth	Suspicion	Trust
#labeled epochs	496	635	300	1180

IV. FEATURE EXTRACTION

Similarly to [36], as a preprocessing step, I performed epoch normalisation to allow for comparison between different epochs and between different study participants. EDA measurements from each epoch were scaled into the range $[0, 1]$ independently. In order to remove high-frequency noise, I applied a 5$^{\text{th}}$ order low-pass Butterworth filter with a cut-off frequency 3 Hz using the Python library *SciPy* [37].

As suggested by [38], I chose a set of six features including: 1) Mean; 2) Standard deviation; 3) Mean of the absolute values of the first differences of the raw signal; 4) Mean of the absolute values of the first differences of the normalised signal; 5) Mean of the absolute values of the second differences of the raw signal; 6) Mean of the absolute values of the second differences of the normalised signal. For each epoch, these six features were extracted from EDA signals from both hands resulting in a 12-dimensional feature vector per epoch.

V. CLASSIFICATION

I approached the detection of deception and suspicion as two separate binary classification tasks. For each task I report person-independent and person-dependent testing accuracies as well as the optimal hyperparameters for detection on-the-fly. For training, hyperparameter tuning, and testing I used

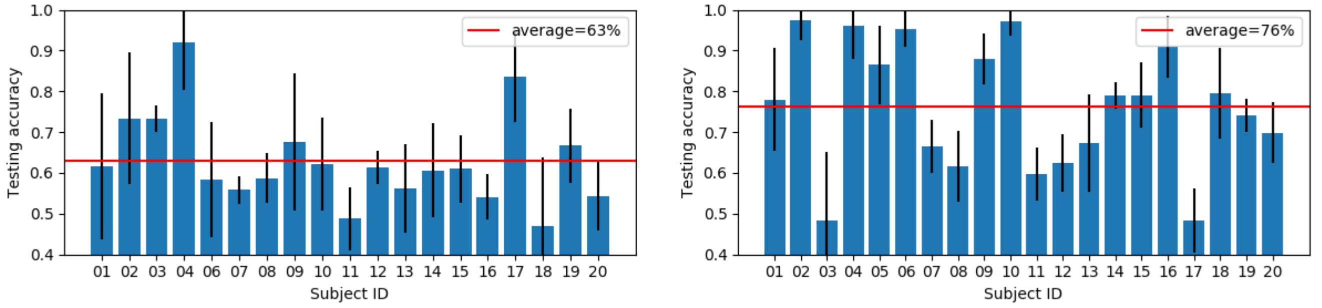

Fig. 3. Testing accuracies with standard deviations for person-dependent *deception* (left) and *suspicion* (right) detection for all study participants (excluding ID 05 for deception detection, see Sec. III-E).

Support Vector Machine (SVM) classifier with linear kernel implemented in the Python framework *scikit-learn* [39].

A. Deception detection

Firstly, the previously discussed biased nature of the data was addressed by random over-sampling of the minority class using the toolbox *imbalanced-learn* [40]. Next, the nested cross-validation was used to evaluate the performance of the developed method in person-independent and person-dependent manner.

1) Person-independent detection: The data from all subjects was used to develop and test a general model capable of deception detection independent of a person being investigated. In this case the outer loop of the nested cross-validation was leave-one-subject-out (LOSO) cross-validation and was used for testing while the inner loop was 5-fold cross-validation and served to tune 3 hyperparameters: δ_D, τ_{MDEL}, and C_D. The parameters δ_D (deception/truth epoch delay) and τ_{MDEL} (maximum deception epoch lenght) were optimised in ranges defined in Sec. III-E. The SVM's regularisation parameter C_D was tuned in the discrete range $C_D \in \{2^{-30}, 2^{-29}, \ldots, 2^{15}\}$. The mean testing accuracy (and the standard deviation) over all LOSO folds for person-independent deception detection was 52 ± 7 %.

2) Person-dependent detection: In this case a person-specific model was trained, tuned, and tested on each subject separately and so the outer loop of the nested cross-validation was changed to 5-fold cross-validation. Otherwise, the procedure was the same as in the person-independent detection (same set of hyperparameters was optimised but this time for each subject separately). The testing accuracies for each subject are shown in Fig. 3 (left). The mean testing accuracy (and the mean standard deviation) over all subjects for person-dependent deception detection was 63 ± 10 %.

3) Optimal parameters for detection on-the-fly: Lastly, optimal parameters for detection of deception on-the-fly were determined using LOSO cross-validation on the whole dataset, and consequently, a model ready for classification on-the-fly was trained on the whole dataset. The best hyperparameters that maximise the validation accuracy were found to be $(\delta_D, \tau_{MDEL}, C_D) = (3.0s, 1.0s, 2^3)$. Figure 4 (left) shows the mean (over all folds) validation accuracy for various parameter settings with hyperparameter C_D already optimised for each pair (δ_D, τ_{MDEL}).

B. Suspicion detection

For classification of suspicion/trust epochs I followed the same procedure as for deception detection with the only difference that the hyperparameter τ_{MDEL} was replaced by τ_{MSEL} and parameters δ_D, C_D were relabeled to δ_S, C_S. The testing accuracy was 57 ± 12 % and 76 ± 8 % for person-independent and person-dependent classification respectively. Figure 3 (right) shows testing accuracies for all subjects in the person-dependent case. The optimal parameters that maximise the validation accuracy for suspicion detection on-the-fly were found to be $(\delta_S, \tau_{MSEL}, C_S) = (3.0s, 3.5s, 2^{-5})$ and the mean validation accuracy for various parameter settings (with C_S already optimized for each pair (δ_S, τ_{MSEL})) is shown in Fig. 4 (right).

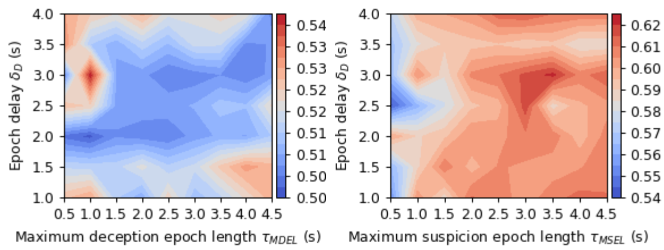

Fig. 4. Mean validation accuracy from person-independent (LOSO) cross-validation on the whole dataset for *deception* (left) and *suspicion* (right) detection task with SVM's hyperparameters C_D and C_S already optimised.

C. Summary of results

All results from this section are summarised in Tab. II.

TABLE II
SUMMARY OF RESULTS FROM DECEPTION/TRUTH AND SUSPICION/TRUST CLASSIFICATION TASKS FOR PERSON-INDEPENDENT (PI) AND PERSON-DEPENDENT (PD) METHODS.

		Deception	Suspicion
Testing accuracy [%]	PI	52 ± 7	57 ± 12
	PD	63 ± 10	76 ± 8
Best parameters	$\delta_{\{D,S\}}$ [s]	3.0	3.0
	$\tau_{M\{D,S\}EL}$ [s]	1.0	3.5
	$C_{\{D,S\}}$	2^3	2^{-5}

VI. FEATURE IMPORTANCE AND ASYMMETRY IN EDA

A. Feature importance

To answer my third research question (Sec. I), I assessed the relative importance of the chosen features and most importantly, I compared the informativeness of extracted features between left and right hand. In particular, I examined SVM weights when the person-independent model was trained on the whole dataset using the optimal parameters determined according to Sec. V-A.3.

Since I used linear-kernel SVM, there was no kernel transformation to a higher dimensional feature space and so the trained weights could be used for feature ranking, as suggested in [41] and studied in detail by [42]. The reasoning is that the larger the magnitude $|w_i|$ of weight w_i is, the larger influence the i^{th} feature has on the predictions of the classifier. The work [43] further suggests to use squares of weights as a ranking criterion to magnify relative differences between weights. Figure 5 shows squares of trained SVM weights corresponding to features extracted from EDA signals from both hands, for deception and suspicion detection tasks.

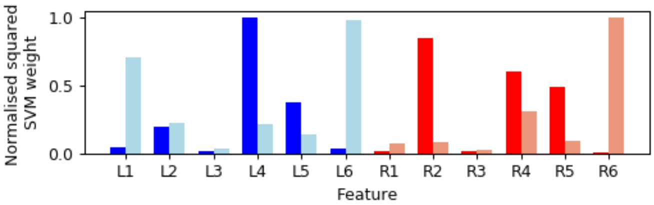

Fig. 5. Feature ranking in terms of normalised squares of SVM weights trained on the whole dataset, for both deception (dark colours) and suspicion (bright colours) detection tasks. L1–L6 (blue) and R1–R6 (red) denote features extracted from EDA signals from left and right hand respectively. For definition of types of features 1–6 see Sec. IV.

B. Asymmetry in EDA

For 10 study participants the difference between individual EDA measurements from left and right hand never changed sign during the whole game and for the other 10 participants the difference was consistent in sign for almost whole game (regardless of periods of deception, truth, suspicion, or trust). Therefore, there was no point in evaluating the left-right difference in an epoch-wise way and so, similarly to [44], I calculated average EDA level from each wrist for every participant (omitting the baseline period) and subtracted the left hand from the right hand mean value to obtain mean difference Δ_{L-R}. Figure 6 shows the distribution of Δ_{L-R} comprising all participants of the study.

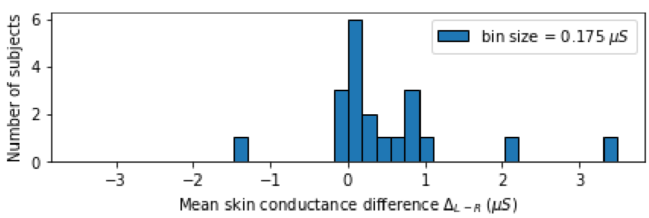

Fig. 6. Distribution of mean skin conductance difference Δ_{L-R} between EDA measured on left and right hand during the whole game, omitting the baseline period. All 20 participants of the study are included.

VII. ANALYSIS AND DISCUSSION

A. Detection of deception and suspicion

The results in Tab. II clearly show that the detection of suspicion from EDA is more reliable than detection of deception. Detecting lie from skin conductance as a single source of information is challenging and as mentioned in Sec. II other methods that achieved higher detection accuracies often combined multiple sources of input. As illustrated by Fig. 2, the skin conductance response to the suspicion event has about 3-times the magnitude of a response to the deception event which probably also contributed to the worse deception detection performance.

Comparing the person-independent (PI) and person-dependent (PD) detections, we can conclude that the PI classification is a more challenging task than PD which is reflected in lower testing accuracies for both deception and suspicion detection tasks. As shown by Fig. 3, in the PD case the testing accuracies vary a lot between subjects (47%–92% and 48%–98% for deception and suspicion respectively) which suggests that it is much more difficult to develop a reliable model for some people than for others, in other words, the nature of EDA signals is highly person-specific. We can also see that for some participants the testing accuracy was even below the baseline – in this case the chance level[6] of 50% for both detection tasks. This might be caused by the fact that internal factors such as hydration and medications can affect EDA measurements. Moreover, there are people who have essentially no measurable EDA [45].

As it can be seen from Tab. II, all 4 mean testing accuracies are above the baseline. However, the accuracy in the PI case, and especially, for deception detection is very close to the baseline. Also, the standard deviations are relatively large. One possible reason might be the fact that the game environment is very challenging for automatic detection of deception and suspicion as it is very dynamic with wide range of response times. Moreover, players were not constrained not to talk which often caused significant distractions. Thus, for future experiments it may be appropriate to design a more controlled experiment.

The obtained optimal values of parameters τ_{MDEL} and τ_{MSEL} for detection on-the-fly suggest that the most informative EDA signal for deception detection is captured within 1 second before the deception/truth event and the most informative EDA signal for suspicion detection is captured within 3.5 seconds before the suspicion event or imaginary trust event. Also, Fig. 2 illustrates that the response to the suspicion event is longer than the response to the deception event. The epoch delays δ_D and δ_S were optimised to the same value which was expected.

B. Feature importance

As it can be seen from Fig. 5, features extracted from EDA signals from both hands are important and this is the case for both deception and suspicion detection tasks. Also, it can

[6]Expected accuracy if classes are assigned by random guessing.

be observed that for some feature types[7] there is a symmetry between the importance of left-hand and right-hand features. For example, features of types 4 and 5 from both hands seem to be most informative for deception detection and the feature of type 6 for suspicion detection. However, it is not the case for all feature types as illustrated by feature type 2 for deception detection and feature type 1 for suspicion detection whose importances significantly differ between left and right hand. All these results confirm the conclusions of [16] that measurement from multiple points of EDA arousal is more informative than measuring only from the traditional single non-dominant hand.

C. Asymmetry in EDA

My results confirm the asymmetry in EDA measured on left and right wrist. As we can see from Fig. 6 the magnitude of the left-hand EDA dominates over the right-hand EDA for majority of subjects. Specifically, 16 out of 20 study participants had positive mean skin conductance difference Δ_{L-R} between left and right hand. According to [16], this could be interpreted as follows: participants with dominant right-hand EDA perceived the whole game more anxiously and stressfully than participants with dominance in left-hand EDA. However, as described in Sec. I, making such general conclusions is challenging and may be flawed.

Looking at the obtained results it seems that in this case the dominance in skin conductance did not depend on subject's handedness as there was 1 left-handed participant with negative Δ_{L-R} and 2 left-handed participants with positive Δ_{L-R} and also there were 3 right-handed participants with negative Δ_{L-R}. However, data from more subjects would have to be collected to make a firm conclusion.

It is important to note that the above-described findings were observed irrespective of participant's position (left/right side of the table) which means that it is unlikely that they were caused by some systematic differences between Q sensors.

VIII. CONCLUSION AND FUTURE WORK

A. Conclusion

This work presents a novel dataset for automatic detection of deception and suspicion from EDA measurements. Using SVM classifiers I developed models to automatically detect deception and suspicion on-the-fly. My experimental results show that detection of suspicion is more reliable than of deception, and that person-dependent models perform better than person-independent. Next, I found that the optimal interval of informative EDA signal is about 3.5-times shorter for deception detection than for suspicion detection task. Results from feature ranking show that features extracted from EDA from both hands are important for deception and suspicion classification tasks and that the importance of some feature types is symmetric between left-hand and right-hand features while it is asymmetric for other types of features. I also verified that there is an asymmetry in EDA measured on left and right wrist.

[7] As defined in Sec. IV.

B. Future work

This work opens up several directions for further research. Firstly, using the method described in [46], the measured EDA could be decomposed into phasic (rapidly changing) and tonic (slowly changing) components in order to investigate which of them is more informative for detection of deception and suspicion.

Next, it would be interesting to compare the results obtained when RBF kernel SVMs are used and when other over-sampling techniques such as Synthetic Minority Over-sampling Technique (SMOTE) [47] or Adaptive Synthetic (ADASYN) sampling method [48] are employed.

Then, the developed models could be tested on-the-fly by conducting another controlled experiment.

Another research avenue could use the data from post-study questionnaires and explore relationships between participants' personality traits and their deceitful/truthful and suspicious/trustful behaviours. As a starting point, one can look at how certain personality traits affect frequency of deception and suspicion events, as depicted in Fig. 7.

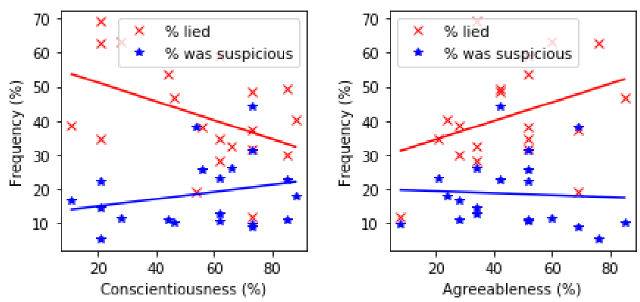

Fig. 7. Future research directions: frequency of deception (red) and suspicion (blue) events against personality traits: Conscientiousness (left) and Agreeableness (right). Each datapoint corresponds to one study participant.

IX. ACKNOWLEDGMENTS

I am grateful to my mother Iveta Ondrasova who helped me with manual annotation of EDA measurements.

REFERENCES

[1] Hugo D Critchley. Electrodermal responses: what happens in the brain. *The Neuroscientist*, 8(2):132–142, 2002.

[2] Don C Fowles, Margaret J Christie, Robert Edelberg, William W Grings, David T Lykken, and Peter H Venables. Publication recommendations for electrodermal measurements. *Psychophysiology*, 18(3):232–239, 1981.

[3] Q Affectiva. Sensor 2.0 datasheet, 2012.

[4] Aldert Vrij. *Detecting lies and deceit: Pitfalls and opportunities*. John Wiley & Sons, 2008.

[5] Johan Engström, Emma Johansson, and Joakim Östlund. Effects of visual and cognitive load in real and simulated motorway driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, 8(2):97–120, 2005.

[6] Nargess Nourbakhsh, Yang Wang, Fang Chen, and Rafael A Calvo. Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks. In *Proceedings of the 24th Australian Computer-Human Interaction Conference*, pages 420–423. ACM, 2012.

[7] Makoto Nakayama. Practical use of the concealed information test for criminal investigation in japan. *Handbook of polygraph testing*, pages 49–86, 2002.

[8] Sabine Ströfer, Matthijs L Noordzij, Elze G Ufkes, and Ellen Giebels. Deceptive intentions: can cues to deception be measured before a lie is even stated? *PloS one*, 10(5):e0125237, 2015.

[9] Nurit Gronau, Gershon Ben-Shakhar, and Asher Cohen. Behavioral and physiological measures in the detection of concealed information. *Journal of Applied Psychology*, 90(1):147, 2005.

[10] Thomas Schill, Carmen Toves, and Nerella V Ramanaiah. Interpersonal trust and coping with stress. *Psychological Reports*, 1980.

[11] Ana Cristina Costa, Robert A Roe, and Tharsi Taillieu. Trust within teams: The relation with performance effectiveness. *European journal of work and organizational psychology*, 10(3):225–244, 2001.

[12] Y Miossec, MC Catteau, E Freixa i Baqué, and JC Roy. Methodological problems in bilateral electrodermal research. *International journal of psychophysiology*, 2(4):247–256, 1985.

[13] Kenneth Hugdahl. Hemispheric asymmetry and bilateral electrodermal recordings: A review of the evidence. *Psychophysiology*, 21(4):371–393, 1984.

[14] J Kayser and G Erdmann. Bilateral electrodermal activity: Effects of lateralized visual input of emotional stimuli. *Journal of Psychophysiology*, 5:110–111, 1991.

[15] Wolfram Boucsein. *Electrodermal activity*. Springer Science & Business Media, 2012.

[16] Rosalind W Picard, Szymon Fedor, and Yadid Ayzenberg. Multiple arousal theory and daily-life electrodermal activity asymmetry. *Emotion Review*, 8(1):62–75, 2016.

[17] H Schliack and R Schiffter. Neurophysiologie und-pathophysiologie der schweißsekretion. In *Normale und Pathologische Physiologie der Haut II*, pages 349–458. Springer, 1979.

[18] HA Sackheim, MS Greenberg, AL Weiman, RC Gur, JP Hungerbuhler, and N Geschwind. Hemispheric asymmetry in the expression of positive and negative emotions. *Archives of Neurology*, 39:210–218, 1982.

[19] Ming-Zher Poh, Tobias Loddenkemper, Claus Reinsberger, Nicholas C Swenson, Shubhi Goyal, Mangwe C Sabtala, Joseph R Madsen, and Rosalind W Picard. Convulsive seizure detection using a wrist-worn electrodermal activity and accelerometry biosensor. *Epilepsia*, 53(5), 2012.

[20] Javier Hernandez, Ivan Riobo, Agata Rozga, Gregory D Abowd, and Rosalind W Picard. Using electrodermal activity to recognize ease of engagement in children during social interactions. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 307–317. ACM, 2014.

[21] Sara Taylor, Akane Sano, and Rosalind Picard. Structure of electrodermal responses during sleep. *REM*, 1(3):4.

[22] Szymon Fedor, Peggy Chau, Nicolina Bruno, Rosalind W Picard, Joan Camprodon, and T Hale. Can we predict depression from the asymmetry of electrodermal activity? *Journal of Medical Internet Research*, 18(12), 2016.

[23] Tatia Lee, Ho-Ling Liu, Li-Hai Tan, Chetwyn CH Chan, Srikanth Mahankali, Ching-Mei Feng, Jinwen Hou, Peter T Fox, and Jia-Hong Gao. Lie detection by functional magnetic resonance imaging. *Human brain mapping*, 15(3):157–164, 2002.

[24] Daniel D Langleben, James W Loughead, Warren B Bilker, Kosha Ruparel, Anna Rose Childress, Samantha I Busch, and Ruben C Gur. Telling truth from lie in individual subjects with fast event-related fmri. *Human brain mapping*, 26(4):262–272, 2005.

[25] J Peter Rosenfeld. Event-related potentials in the detection of deception, malingering, and false memories. 2002.

[26] Vahid Abootalebi, Mohammad Hassan Moradi, and Mohammad Ali Khalilzadeh. A new approach for eeg feature extraction in p300-based lie detection. *Computer methods and programs in biomedicine*, 94(1):48–57, 2009.

[27] Xiao-Su Hu, Keum-Shik Hong, and Shuzhi Sam Ge. fnirs-based online deception decoding. *Journal of Neural Engineering*, 9(2):026012, 2012.

[28] Thomas O Meservy, Matthew L Jensen, John Kruse, Judee K Burgoon, Jay F Nunamaker, Douglas P Twitchell, Gabriel Tsechpenakis, and Dimitris N Metaxas. Deception detection through automatic, unobtrusive analysis of nonverbal behavior. *IEEE Intelligent Systems*, 20(5):36–43, 2005.

[29] Jeffrey J Walczyk, Frank P Igou, Alexa P Dixon, and Talar Tcholakian. Advancing lie detection by inducing cognitive load on liars: a review of relevant theories and techniques guided by lessons from polygraph-based approaches. *Frontiers in psychology*, 4, 2013.

[30] M Raheel Bhutta, Melissa J Hong, Yun-Hee Kim, and Keum-Shik Hong. Single-trial lie detection using a combined fnirs-polygraph system. *Frontiers in psychology*, 6, 2015.

[31] Anna Caterina Merzagora, Scott Bunce, Meltem Izzetoglu, and Banu Onaral. Wavelet analysis for eeg feature extraction in deception detection. In *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*, pages 2434–2437. IEEE, 2006.

[32] J Christopher Westland. Electrodermal response in gaming. *Journal of Computer Networks and Communications*, 2011, 2011.

[33] Anders Drachen, Lennart E Nacke, Georgios Yannakakis, and Anja Lee Pedersen. Correlation between heart rate, electrodermal activity and player experience in first-person shooter games. In *Proceedings of the 5th ACM SIGGRAPH Symposium on Video Games*, pages 49–54. ACM, 2010.

[34] Jin Joo Lee, W Bradley Knox, Jolie B Wormwood, Cynthia Breazeal, and David DeSteno. Computationally modeling interpersonal trust. *Frontiers in psychology*, 4, 2013.

[35] Psychlab. Skin conductance explained. http://www.psychlab.com/SC_explained.html [Online; accessed 10/01/2018].

[36] Oya Celiktutan, Efstratios Skordos, and Hatice Gunes. Multimodal human-human-robot interactions (mhhri) dataset for studying personality and engagement. *IEEE Transactions on Affective Computing*, 2017.

[37] Eric Jones, Travis Oliphant, and Pearu Peterson. SciPy: open source scientific tools for Python. 2014. http://www.scipy.org/ [Online; accessed 10/01/2018].

[38] Rosalind W. Picard, Elias Vyzas, and Jennifer Healey. Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE transactions on pattern analysis and machine intelligence*, 23(10):1175–1191, 2001.

[39] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[40] Guillaume Lemaître, Fernando Nogueira, and Christos K. Aridas. Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *Journal of Machine Learning Research*, 18(17):1–5, 2017.

[41] Vikas Sindhwani, Pushpak Bhattacharya, and Subrata Rakshit. Information theoretic feature crediting in multiclass support vector machines. In *Proceedings of the 2001 SIAM International Conference on Data Mining*, pages 1–18. SIAM, 2001.

[42] Janez Brank, Marko Grobelnik, Natasa Milic-Frayling, and Dunja Mladenic. Feature selection using support vector machines. *WIT Transactions on Information and Communication Technologies*, 28, 2002.

[43] Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. Gene selection for cancer classification using support vector machines. *Machine learning*, 46(1):389–422, 2002.

[44] Ming-Zher Poh, Tobias Loddenkemper, Nicholas C Swenson, Shubhi Goyal, Joseph R Madsen, and Rosalind W Picard. Continuous monitoring of electrodermal activity during epileptic seizures using a wearable sensor. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pages 4415–4418. IEEE, 2010.

[45] Rosalind W Picard, Szymon Fedor, and Yadid Ayzenberg. Response to commentaries on multiple arousal theory and daily-life electrodermal activity asymmetry. *Emotion Review*, 8(1):84–86, 2016.

[46] Alberto Greco, Gaetano Valenza, Antonio Lanata, Enzo Pasquale Scilingo, and Luca Citi. cvxeda: A convex optimization approach to electrodermal activity processing. *IEEE Transactions on Biomedical Engineering*, 63(4):797–804, 2016.

[47] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.

[48] Haibo He, Yang Bai, Edwardo A Garcia, and Shutao Li. Adasyn: Adaptive synthetic sampling approach for imbalanced learning. In *Neural Networks, 2008. IJCNN 2008.(IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on*, pages 1322–1328. IEEE, 2008.