

Perinasal Indicators of Deceptive Behavior

Malcolm Dcosta¹, Dvijesh Shastri², Ricardo Vilalta³, Judee K. Burgoon⁴, Ioannis Pavlidis¹

¹ Computational Physiology Lab, Department of Computer Science, University of Houston, Texas

² Department of Computer Science and Engineering Technology, University of Houston-Downtown, Texas

³ Pattern Analysis Lab, Department of Computer Science, University of Houston, Texas

⁴ Center for the Management of Information, University of Arizona, Arizona

mdcosta2@uh.edu, shastrid@uhd.edu, vilalta@cs.uh.edu, jburgoon@cmi.arizona.edu, ipavlidis@uh.edu

Abstract—High-stakes lying causes detectable changes in human behavior and physiology. Lie detection techniques based on behavior analysis are unobtrusive, but often require labor-intensive efforts. Lie detection techniques based on physiological measurements are more amenable to automated analysis and perhaps more objective, but their often obtrusive nature makes them less suitable for realistic studies. In this paper we present a novel lie detection framework. At the core of this framework is a physiological measurement method that quantifies stress-induced facial perspiration via thermal imagery. The method uses a wavelet-based signal processing algorithm to construct a feature vector of dominant perinasal perspiration frequencies. Then, pattern recognition algorithms classify the subjects into deceptive or truthful by comparing the extracted features between the hard and easy questioning segments of an interview procedure. We tested the framework on thermal clips of 40 subjects who underwent interview for a mock crime. We used 25 subjects to train the classifiers and 15 subjects for testing. The method achieved 80% success rate in blind predictions. This framework can be generalized across experimental designs, as the classifiers do not depend on the number or order of interview questions.

I. INTRODUCTION

The surge in terrorism over the past decade has motivated governments around the world to invest in various defensive and offensive technologies. One area that has drawn the interest of government agencies is a quick and reliable method to aid interviewers in detecting lies. This paper presents such a method based on thermal imaging. The method associates changes in facial perspiration patterns with deceptive behavior.

Behavioral psychologists have shown that high-stakes lying causes detectable changes in the human face, body, and voice [1]. The literature abounds with approaches that link behavioral cues to deception. Some of these approaches target non-verbal cues, such as macro and micro facial expressions [2][3][4]. Gestures, including hand, finger, and leg movements are also indicators of deceptive behavior [5]. Other non-verbal techniques capitalize on pupil dilation [6], blinking rate [7], and gaze duration [8]. Approaches for verbal behavior analysis target acoustic features related to pitch, energy, and frequency of the audio signals [9]. Overall, behavior analysis approaches are unobtrusive, naturally lending themselves to realistic experimentation. However, several

of these approaches require labor intensive data analysis and in some cases have a measure of subjectivity.

Another distinct line of research in lie detection targets physiological indicators. Lykken demonstrated that instantaneous changes in physiological responses can be associated with deceptive behavior [10]. The polygraph technology capitalizes on this fact. It monitors changes in peripheral physiological responses, including respiration rate, heart rate, and finger perspiration, linking these measurements to deceptive behavior [11][12][13]. In comparison to behavior-based approaches, physiology-based approaches are more amenable to automated data analysis and perhaps more objective. Until recently, however, these approaches used tethered probes to collect bodily data - an obtrusive and motion-restrictive framework that lessens the validity of polygraph examinations [14].

Studying brain activity patterns is another approach to lie detection. Mohamed *et al.* investigated the neural correlates during deception and truth telling by using a functional magnetic resonance imaging (fMRI) technique [15]. Izzetoglu *et al.* examined the feasibility of using functional near infrared (fNIR) imaging to quantify cognitive load [16]. They showed that oxygenation changes during a 'lie' task are larger than oxygenation changes during a 'truth' task. Other researchers explored the use of electroencephalography (EEG) in lie detection [17][18]. All these measurements offer valuable information on how the brain operates in lying. Nevertheless, many of the brain activity methods are not suitable for realistic studies due to the obtrusive nature of the associated sensing technology.

In this paper, we propose a thermal imaging approach that is not only highly automated but also unobtrusive. It is not the first time thermal imaging is used in lie detection applications. Pavlidis *et al.* and Tsiamyrtzis *et al.* used the heat signature of the periorbital region during an interview to determine subject deceptiveness [19][20][21]. The thermal imaging approach we adopted in this paper uses a different physiological indicator, that of perinasal perspiration. Perinasal perspiration responses are concomitant to finger perspiration responses during sympathetic arousal [22]. In [23], Shastri *et al.* proposed a thermal image processing method to quantify perinasal perspiration; they validated their method against the clinical standard, which is electro-dermal activity (EDA) on the fingers. Hence, the perinasal perspiratory indicator extracted via thermal imaging can be used in place

This work was supported by the National Center for Credibility Assessment (NCCA).

of the finger perspiratory indicator extracted via EDA probes - the beginning of unobtrusive polygraphy. Recently, Pavlidis *et al.* have demonstrated the field potential of perinasal imaging in a surgical training study, where they associated sympathetic responses with laparoscopic performance [24]. However, to the best of our knowledge, perinasal perspiration has never been explored for lie detection analysis.

In the remainder of this paper, we first discuss the design of the mock crime experiment that provided us the data (section II). Next, we present the details of our method (section III). Then, we discuss the experimental results (section IV). Finally, we conclude the paper in section V.

II. EXPERIMENTAL DESIGN

We received thermal facial imaging data from a mock stealing experiment staged within university premises. A group of behavioral psychologists from the Center for the Management of Information at the University of Arizona designed and executed the experiment. The experimental details are given in [25] and [26]. The study subjects were recruited from within and without the University of Arizona campus via flyers, newspaper advertisements, and online postings.

When a subject arrived for the experiment, a pre-recorded set of instructions was waiting for her/him. The instructions programmed the subject as guilty/deceptive or innocent/truthful; the instructions also included initial directions for the next steps in the experimental procedure. After listening to the instructions and signing the consent form, the subject walked to a room where a ring was stored. In that room a confederate asked the subject to wait until s/he locates the individual who could give her/him further information, as promised in the pre-recorded instructions. The subject was left alone in the room for several minutes, to provide the opportunity to steal the ring if s/he was programmed deceptive. When the confederate returned, s/he escorted the subject to the interview room. There, the subject was interviewed for her/his involvement in the mock crime. A professional polygraph examiner provided by the National Center for Credibility Assessment (NCCA) conducted the interview. All the subjects were instructed to prove their innocence regardless of whether they took the ring or not. Those who successfully convinced the interviewer of their innocence received \$50 in addition to the standard monetary compensation of \$15 per hour. Those who failed to convince the interviewer of their innocence received only the standard monetary compensation. The experimental design aimed to engage the subjects in the process by creating stakes.

The questioning was structured according to the Behavior Analysis Interview (BAI) design [27]. The complete list of the questions is given below:

- 1) What color are the walls in this room?
- 2) Are there any lights on in this room?
- 3) Where were you born?
- 4) What is the name of the building we are in?
- 5) Did you ever take anything valuable from a place where you worked?
- 6) Have you ever lied to a person in position of authority?
- 7) How do you feel about taking the credibility assessment examination?
- 8) You know you are going to be tested about a crime committed in this building today. If you were involved in the crime in any way, you should tell me now.
- 9) Would you please describe everything you did for the first two hours after you awoke today? Think about what you were doing, how you felt, and what happened.
- 10) Now I'd like you to describe everything you did and saw from the moment you left the Esquire Building until you arrived here.
- 11) At any time were you on the fourth floor of this building?
- 12) We sometimes verify the information that people in this study give us. If I call the receptionist in Room 429, is there any reason that he will say you might have been near his desk? I'm not saying that you are being dishonest, but we put you on the spot here and maybe you misremembered.
- 13) Is there any reason why your fingerprints should be on a desk on the 4th floor of this building? Maybe you just opened a receptionist's desk to look for a pen?
- 14) Is there any reason why we may see you entering room 429 on any surveillance camera tapes?
- 15) I'd like you now to describe in reverse order everything you did and saw from the moment you arrived here back to when you left the Esquire Building.
- 16) Is there anyone who can vouch for you coming directly to this room from the Esquire Building?
- 17) What is the item that was taken?
- 18) A ring was taken. Do you know where it is now?
- 19) What do you think should happen to the person who stole the ring?
- 20) Under any circumstances would you give that person a break?
- 21) What kind of person would steal jewelry from someone's desk?
- 22) Is there anything that you said in this interview so far that you would like to change?
- 23) At any time during this study, were you instructed to lie?
- 24) How do you think this credibility assessment of you will turn out today?

The interviewer asked all these questions in order. In addition, the interviewer had freedom to ask any number of follow-up questions outside this list. Therefore, the length of the interview varied between 10 and 12 minutes, depending on the number of the follow-up questions and the length of the subject's answer to each question.

We can partition the interview questions into six groups using a relevancy criterion. Specifically, we can group the interview questions into two irrelevant question sets (IR1, and IR2), and four relevant question sets (R1, R2, R3, and R4). The first four interview questions (Q1-Q4) and question Q9 are irrelevant questions that make up set IR1 and set IR2,

respectively (green-colored rectangles in Fig. 1(b)). These irrelevant questions are not directly linked to the mock crime, but they are purposefully included to establish the subject’s baseline behavior [27].

The relevant questions (red-colored rectangles in Fig. 1 (b)) are provoking and related to the mock crime but from different perspectives. Specifically, questions Q5-Q6 investigate the subject’s deceptive trait. They make up the first relevant set (R1). Questions Q7-Q16 (except Q9) focus on the subject’s explanation for incriminating evidence about the ‘theft’. These questions make up the second relevant set (R2). Questions Q17-Q21 are consequential questions, aiming to investigate the subject’s views on the consequences of criminal actions associated to the ‘theft’. They make up the third relevant set (R3). The last three questions (Q22-Q24) are concluding questions, where the subject is given the opportunity to change her/his story. These questions make up the fourth relevant set (R4).

We received data for 164 subjects. Out of these 164 subjects we were able to complete data processing for 67 subjects. NCCA incrementally released ground-truth information for only 40 out of these 67 subjects. Hence, in this paper we report results for this set of 40 subjects (17 males and 23 females). We were not able to complete data processing in the following cases: for 22 subjects that had mustache; for 17 subjects where the thermal camera was out of focus; for 14 subjects that exhibited excessive motion, causing the facial tracking algorithm to fail; for 31 subjects with corrupt thermal files; and, for 13 subjects exhibiting abnormally low and featureless perinasal signals.

No contact probes were attached to the subjects. The experimental setup did not pose any restrictions on the subjects’ postures while sitting. Throughout the interview, the subjects’ faces were recorded via a thermal imaging system. The system consisted of a ThermoVision SC6000 Mid-Wave Infrared (MWIR) camera from FLIR Systems [28], a MWIR 100 mm lens, and an HP Pavilion desktop. The distance between the camera and the subject was 13 *ft*. The thermal data was recorded at 25 frames per second.

The conversation between the subject and the interviewer was recorded via two boom microphones, one microphone per individual. The audio was synchronized with the thermal image recording to facilitate audio-video mapping. We used this mapping to segment the perinasal perspiration signal into irrelevant and relevant portions.

III. METHODOLOGY

Figure 1 illustrates the methodological framework. It is divided into four steps. In the first step, the perspiration signal is extracted from the perinasal region of interest (ROI) through image processing algorithms (Fig. 1(a)). In particular, the ROI is tracked in every frame of the thermal video, and within the ROI, the perspiration intensity is computed. In the second step, the extracted signal is split into multiple segments, where each segment represents an interview portion of irrelevant or relevant questions (Fig. 1(b)). The segments’ start and end points are derived from the interview

audio. Next, each perspiration segment is cleaned from high-frequency noise via an FFT (Fast Fourier Transformation)-based approach. In the third step, the dominant frequency of each signal segment is computed via a wavelet-based approach and a feature vector is constructed (Fig. 1(c)). In the final step, the feature vector is supplied to pattern classifiers to categorize the subjects into deceptive and truthful groups (Fig. 1(d)).

A. Perspiration Signal Extraction

The perspiration signal extraction process begins by selecting a tracking region of interest (TROI) on the perinasal area (white-colored rectangle in Fig. 1(a)). A tracking algorithm estimates TROI’s position in every incoming frame. We use the particle filtering-based tracker reported in [29]. This tracker is driven by a probabilistic template mechanism with spatial and temporal smoothing components that cope well with thermophysiological changes due to transient perspiration and quasi-periodic breathing. Therefore, this tracker is a suitable choice for the perinasal region that is affected by both of these thermophysiological factors.

Within the TROI, a measurement region of interest (MROI) is selected on the maxillary part of the perinasal region (black-colored rectangle in Fig. 1(a)). To compute the perspiration intensity within the MROI, we use the morphology-based algorithm reported in [23]. Specifically, this algorithm uses a variant of black top-hat transformation that is suitable for localization of objects having small target size and background fuzziness (e.g., perspiration spots). The isotropic nature of its structuring element makes the algorithm shift and rotation invariant. The algorithm localizes the perspiration spots and computes the perspiration signal energy for every frame. The perspiration signal energy is indicative of perspiration activity in the MROI. The energy remains low in the absence of perspiration activity, but elevates with the onset of perspiration and gradually returns to the baseline level during the recovery phase.

The tracking step, along with the perspiration computation, is iteratively executed until the end of the thermal sequence. Thus, we obtain a 1D perspiration signal (Fig. 1 (b)) from a sequence of 2D thermal frames.

B. Signal Partitioning and Noise Cleaning

We split the perspiration signal into multiple segments, where each segment represents an irrelevant or relevant portion of the interview, per the criteria established in section II. The start and end points of the segments are derived from the interview audio, which is in sync with the thermal data.

As observed in Fig. 1 (b), the perspiration signal typically contains high frequency noise due to systemic thermal noise and imperfections in the tracking algorithm. To suppress such noise, we use the FFT-based noise cleaning algorithm reported in [21]. The noise-cleaned signal is depicted in black color in Fig. 1 (b).

C. Feature Extraction

The sympathetic nervous system (SNS) releases norepinephrine neurotransmitters that prepare the body to cope

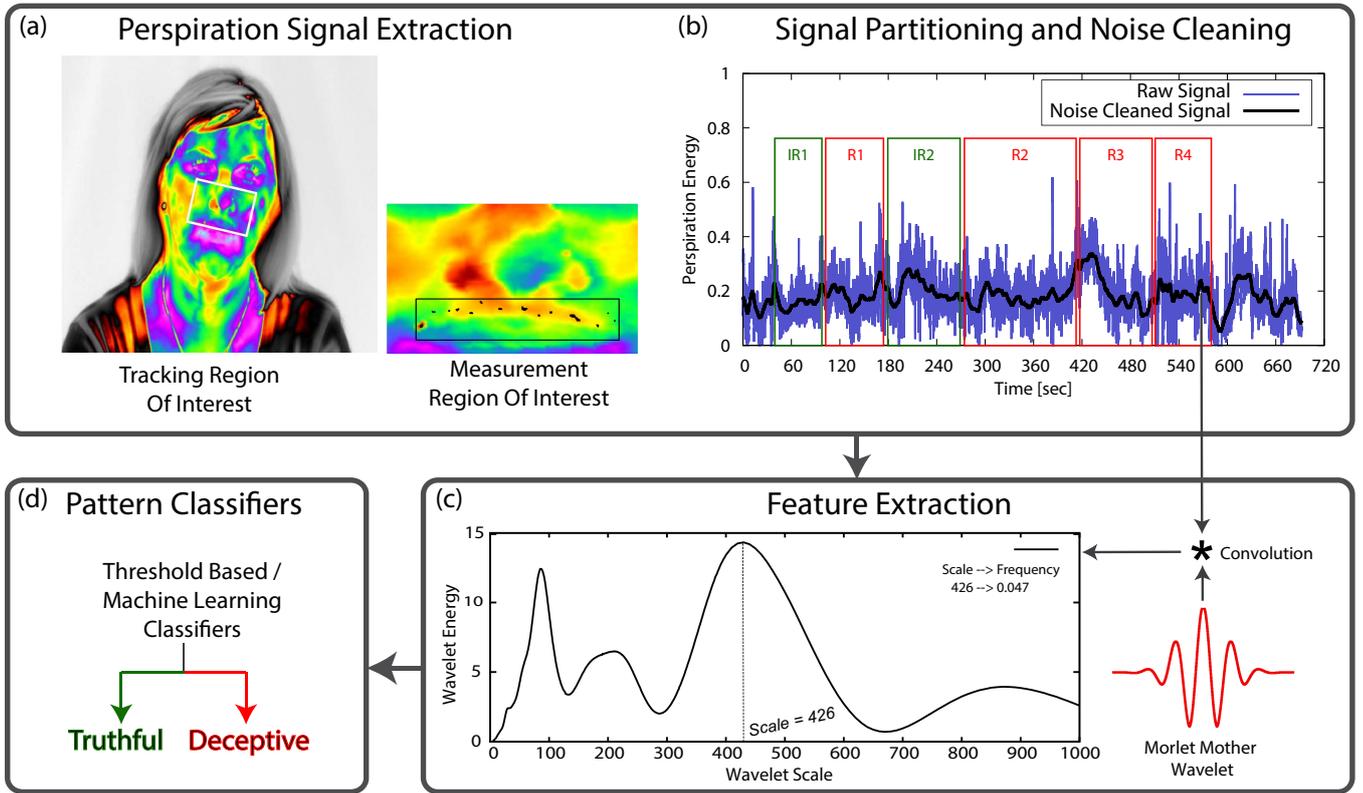


Fig. 1: Methodological framework.

with a threatening event. As a result, breathing rate, heart rate, and sweat gland activity accelerate. Perspiration response is an outcome of the sweat gland activity. Nicolaidis and Sivadjian found that perspiration is secreted from the glands in a pulsate manner [30]. In [31], Storm *et al.* demonstrated a positive correlation between norepinephrine levels and perspiration pulsations. Furthermore, Kamei *et al.* gauged emotional stress as a function of perspiration pulsations in [32]. Our approach capitalizes on this finding. Specifically, we hypothesize that deceptive subjects demonstrate considerable elevation in perspiration pulsation when faced with the relevant interview questions compared to the irrelevant ones. In contrast, for truthful subjects there is no significant difference in perspiration pulsation between the relevant and irrelevant segments.

To validate our hypothesis we devised a wavelet-based frequency computation approach. We opted for wavelet transformation rather than FFT because the perspiration signals are non-stationary in nature. The wavelet transformation characterizes each perspiration segment by its dominant signal frequency. The algorithmic steps are as follows:

1) *Normalization:* Each perspiration signal segment is defined as a discrete function $S(i), i = 1, 2, 3, \dots, n$, where n is the number of data samples. To alleviate inter-subject variations, we normalize the signal amplitude as follows:

$$S_N(i) = \frac{S(i) - \text{Min}(S)}{\text{Max}(S) - \text{Min}(S)},$$

where Min and Max are the minimum and maximum signal amplitude values, respectively. This normalization transforms the original perspiration signal S to S_N with amplitude values in $[0,1]$.

2) *Signal Extension:* The normalized signals are extended beyond the boundary limits before computing the wavelet coefficients. Convolution of a wavelet with a finite length signal looks for data points beyond the signal end points. As there are no data points beyond the signal end points, this introduces an error in the wavelet energy computation, which is known as the border discontinuity error. This error leads to wrong global maxima in wavelet energy curves, which in turn result in incorrect feature values. Figure 2 illustrates the impact of the border discontinuity error. The red-colored signal in the figure shows the wavelet energies without the signal extension, while the blue-colored signal shows the wavelet energies after having symmetrical extension. The red-colored signal gives an incorrect global maximum.

The purpose of the signal extension is to define the data points beyond the signal boundary. There are many ways to extend a finite length signal, including zero-padding, wraparound, and symmetric extension. As our signals are non-stationary in nature, we selected the symmetric extension technique. Signal extensions of length n , $n/2$, and $n/4$ produced nearly identical results. Therefore, we decided to use the $n/4$ extension, which is computationally efficient.

3) *Wavelet Energy Computation:* We apply the Continuous Wavelet Transform (CWT) on the normalized and extended signals. The Morlet mother wavelet is convoluted with

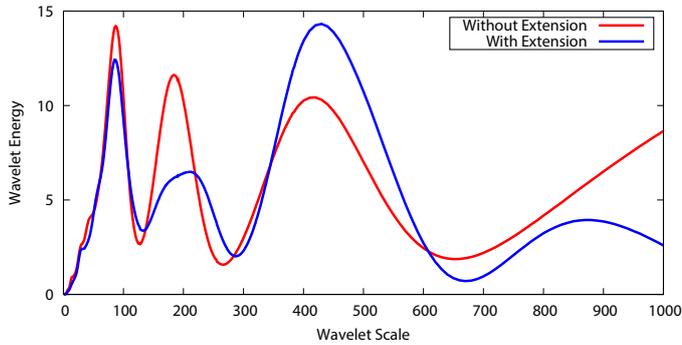


Fig. 2: The wavelet energy curves of a signal obtained with and without signal extension. Wavelet scale uniquely maps to wavelet frequency.

the signal at all scales exploring all frequency components. This generates a 2D energy spectrum with the time scale in the horizontal axis and the frequency scale in the vertical axis. We compute the 1D wavelet energy signal from the 2D spectrum by aggregating the wavelet energy at each frequency f as follows:

$$E_f = \sum_{t=1}^n |WT_f(t)|^2,$$

where $WT_f(t)$ is the wavelet coefficient at frequency f and time t . The output from this step is illustrated in Fig. 2 as a blue-colored 1D energy signal. The energy signal features global and local maxima. The frequency scale corresponding to the global maximum determines the most dominant frequency in the signal. We assign the dominant frequency f_i to be the feature value of perspiration segment i . Thus, a total of six feature values (f_{IR1} , f_{IR2} , f_{R1} , f_{R2} , f_{R3} , f_{R4}) are computed for each perspiration signal.

D. Pattern Classifiers

The amount of perspiration secretion is affected by various factors, including the subject’s age, gender, metabolic rate and body mass index. Furthermore, different subjects react differently to the same stimuli [20]. In our study, the stressful investigation of the mock crime elicited different amounts of perspiration in the study subjects. Therefore, an effective pattern classifier has to ameliorate inter-subject variability through an intra-subject analysis. We propose five classifiers that facilitate intra-subject analysis.

All subjects experienced some amount of stress due to the fact that they faced an interview process. We name this stress *interview stress*. The *interview stress* elevates physiological responses regardless of whether the questions asked were irrelevant or relevant to the mock crime. In addition to the *interview stress*, the deceptive subjects likely experienced stress caused by their deceptive behavior. They had to make sure their lies were non-detectable by avoiding contradictory statements. They also experienced fear of being caught. This mixture of cognitive and emotional loading elevated their stress beyond the *interview stress*. We call this type of stress *guilty knowledge stress*. We theorize that the

deceptive subjects experienced heightened *guilty knowledge stress* when faced with the relevant questions. Therefore, by measuring the difference in the subject’s perspiratory responses between the two question sets (relevant vs. irrelevant) one may determine whether the subject was ‘deceptive’ or ‘truthful’. This is the basis of our pattern classifiers, which are described in detail next:

1) *Threshold-based Classifier*: We compute the average perspiration frequency of the irrelevant sets (IR1, and IR2), and the average perspiration frequency of the relevant sets (R1, R2, and R3). The relevant set R4 (concluding set) is excluded for this classifier because it over-smooths the average frequency value. Based on these average values, each subject i is classified as ‘Deceptive’ (D) or ‘Truthful’ (T) according to the following formula:

$$f_R - f_{IR} \rightarrow \begin{cases} > 0 \text{ then, subject}(i) \text{ is D} \\ \leq 0 \text{ then, subject}(i) \text{ is T,} \end{cases}$$

$$f_R = \text{avg}(f_{R1}(i), f_{R2}(i), f_{R3}(i)),$$

$$f_{IR} = \text{avg}(f_{IR1}(i), f_{IR2}(i)),$$

where f_R and f_{IR} are the average perspiration frequencies of the relevant sets and the irrelevant sets, respectively. Hence, the classifier categorizes the subject as deceptive if the average perspiration frequency of the relevant sets is higher than that of the irrelevant sets. Otherwise, the subject is categorized as truthful. A training set is used to design the classifier.

2) *Machine Learning Classifiers*: We explore four representative machine learning classifiers. Specifically, from the Bayesian approaches we use a Naïve Bayes classifier. From the tree-based approaches, we use a decision tree classifier and a decision stump classifier. The decision tree classifier uses the C4.5 algorithm to generate decision trees. The decision stump classifier is used with the AdaBoost method in an iterative process. From the neural network approaches, we use a multi-layered perceptron classifier. The classifiers are modeled on the training set and validated via the *leave-one-out* cross validation method. A separate testing set is used to evaluate the classifiers’ performance. We use the Weka v3.7 tool for modeling and predictions [33]. The default parameters that the tool provides for each classifier are used in the analysis.

Along with the six feature values, a set of meta-features supplied to most machine learning classifiers. The meta-feature set includes the average frequency of the relevant set (f_R), the average frequency of the irrelevant set (f_{IR}), and the difference in the average frequencies ($f_R - f_{IR}$). The information gain (0.682) of the difference meta-feature is considerably higher than the raw feature values and drives the process. In particular, the decision stump generates a decision tree with the difference meta-feature as the root node. Essentially, this is the threshold-based method with a difference as the threshold value. The decision stump features

a threshold value of -0.004895. The decision tree approach generates a two-level binary tree, where the root node is the difference meta-feature ($f_R - f_{IR}$), while the second node is the relevant set meta-feature (f_R). The Naïve Bayes is a probabilistic classifier. It generates a probability value for each prediction it makes. The Naïve Bayes classifier does not use the meta-features, as it assumes the features are conditionally independent. The neural network generates a single-layer network, where the input nodes are the six raw features. The output nodes are the Deceptive and Truthful groups.

IV. EXPERIMENTAL RESULTS

From the pool of 40 subjects we used 25 for training and 15 for testing. Initially, we obtained from the National Center for Credibility Assessment (NCCA), the ground-truth values for the training set only. For the testing set, we were asked to make blind predications. We submitted to NCCA the prediction results of the testing set. NCCA released the ground-truth results after registering our predictions. The predictions we report in this paper for the testing set are the ones that we submitted to the government agency (Table I).

1) *Success Rate on the Training Set:* Our classifiers scored above 80% successful prediction rate on the training set. In particular, all but the neural network classifier performed well for both truthful and deceptive instances. Taking the majority vote on the classifiers gave 96% overall success rate. The classifiers produced low false positive rates (1-true negative rate), meaning only a small number of truthful subjects were misclassified as deceptive. They also produced low false negative rates (1-true positive rate), meaning a low number of deceptive subjects were misclassified as truthful.

2) *Success Rate on the Testing Set:* We performed blind predictions on the testing set, which rarely happens in deception detection studies. Most studies are limited to modeling classifiers on training sets only, which is prone to over-fitting. The classifiers achieved 80% success rate on the testing set with the exception of the Naïve Bayes classifier that achieved 78.6% success rate. In particular, the classifiers performed well for the truthful instances. They produced low false positive rates (11%). However, the false negative rates were higher (33.34%). The likely cause of the higher false negative rate is the small population of deceptive subjects (8/25 = 32%) in the training set.

V. CONCLUSION

To the best of our knowledge, this is the first research effort that investigated the value of facial perspiration in high-stakes deceptive behavior. We validated the proposed framework for a mock crime experiment where the subjects faced intense investigation. A pool of 40 subjects were used in our analysis (25 for training and 15 for testing). We devised a threshold-based classifier and modeled four machine learning classifiers. The classifiers scored above 80% successful prediction rate on the training set and close to

80% successful prediction rate on the testing set, indicating that the proposed method scales up.

This research makes two significant contributions in the field of deception analysis - one at the feature level and one at the system level. At the feature level, we demonstrated that high-stakes lying causes detectable changes in facial perspiration patterns. Specifically, most deceptive subjects in our experiment exhibited significant increase in perspiration frequency when faced with questions related to the mock crime. Most truthful subjects, on the other hand, did not show any substantial changes in perspiration frequency between the relevant and irrelevant question sets. These findings indicate that the perspiration frequency can be used as a discriminating feature for classifying deceptive from truthful behavior.

Our contribution at the system level is the proposed thermal imaging-based deception detection framework. The framework features unobtrusive measurement, rapid analysis, and generalizable classifiers. Specifically, the framework quantifies facial perspiration responses in a contact-free manner. Thus, it eliminates the need for contact probing that may compromise the validity of a sympathetic measurement. Given a thermal video with a synced interview audio, the framework offers a semi-automated process for deception detection. In particular, its image processing-based perspiration extraction module is near real-time. Only its signal processing-based feature extraction module needs some human intervention for audio demarcations. This process, however, requires only a few mouse clicks and thus, it is fast. Most importantly, the proposed framework features a generalizable classification method. Unlike the previous deception detection approaches that are limited to a specific interview question or a set of interview questions [10][19][21], the proposed classifiers operate on the difference in perspiratory responses between the relevant and irrelevant question sets. Thus, this approach makes the framework a natural fit to behavior analysis interviews (BAI) with any number of irrelevant and relevant questions!

VI. ACKNOWLEDGMENTS

This work was supported by a research contract from the National Center for Credibility Assessment (NCCA). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agency. We are grateful to Prof. Burgoon's team at the University of Arizona for their help with the thermal data collection.

REFERENCES

- [1] P. Ekman, *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage (Revised Edition)*. WW Norton & Company, 2009.
- [2] P. Ekman and W. V. Friesen, *Facial Action Coding System*. Consulting Psychologists Press, Stanford University, Palo Alto, 1977.
- [3] M. G. Frank, M. A. Menasco, and M. O'Sullivan, "Human behavior and deception detection," *Wiley Handbook of Science and Technology for Homeland Security*, 2008.
- [4] J. J. Gross, *Handbook of emotion regulation*. Guilford Press, 2007.
- [5] A. Vrij and G. R. Semin, "Lie experts' beliefs about nonverbal indicators of deception," *Journal of Nonverbal Behavior*, vol. 20, no. 1, pp. 65-80, 1996.

TABLE I: Experimental Results

Set	Subject	Ground Truth	Threshold Based	NN	DS	DT	NB (Probability)	
Training	D001	Deceptive	1	1	1	1	1 (1)	
	D002	Truthful	1	1	1	1	1 (1)	
	D003	Truthful	1	1	1	1	1 (1)	
	D004	Truthful	1	1	1	1	1 (1)	
	D005	Truthful	1	0	1	0	1 (0.80)	
	D006	Truthful	1	0	1	1	1 (0.95)	
	D007	Truthful	1	1	0	1	1 (0.79)	
	D008	Truthful	1	1	1	1	1 (1)	
	D009	Deceptive	1	1	1	1	1 (1)	
	D010	Truthful	1	1	1	1	1 (1)	
	D011	Deceptive	1	0	1	1	0 (1)	
	D012	Truthful	1	1	1	1	1 (1)	
	D013	Truthful	1	1	0	1	1 (0.69)	
	D014	Deceptive	1	1	1	1	1 (1)	
	D015	Truthful	1	1	1	1	1 (1)	
	D016	Truthful	1	1	1	1	1 (0.94)	
	D017	Deceptive	1	1	1	1	1 (1)	
	D018	Truthful	1	1	1	1	1 (1)	
	D019	Truthful	1	1	1	1	1 (1)	
	D020	Truthful	1	1	1	1	1 (1)	
	D021	Truthful	1	1	1	1	1 (1)	
	D022	Deceptive	1	1	1	1	1 (1)	
	D023	Deceptive	0	0	1	1	1 (0.76)	
	D024	Truthful	1	1	1	1	1 (1)	
	D025	Deceptive	0	0	0	0	1 (0.87)	
Overall Success Rate			92%	80%	88%	92%	96%	
Success Rate for Truthful Instances			100%	88.23%	88.23%	94.11%	100%	
Success Rate for Deceptive Instances			75%	62.5%	87.5%	87.5%	87.5%	
Testing	D026	Deceptive	1	1	1	1	1 (0.98)	
	D027	Deceptive	0	0	0	0	0 (1)	
	D028	Truthful	1	1	1	1	1 (0.91)	
	D029	Deceptive	0	0	0	0	0 (1)	
	D030	Deceptive	1	1	1	1	1 (0.96)	
	D031	Truthful	1	1	1	1	1 (1)	
	D032	Deceptive	1	1	1	1	0* (0.50)	
	D033	Truthful	1	1	1	1	1 (1)	
	D034	Truthful	1	1	1	1	1 (1)	
	D035	Truthful	1	1	1	1	1 (0.95)	
	D036	Truthful	1	1	1	1	1 (1)	
	D037	Truthful	1	1	1	1	1 (1)	
	D038	Truthful	1	1	1	1	1 (1)	
	D039	Truthful	0	0	0	0	0 (0.98)	
	D040	Deceptive	1	1	1	1	1 (0.89)	
	Overall Success Rate			80%	80%	80%	80%	78.6%
	Success Rate for Truthful Instances			88.88%	88.88%	88.88%	88.88%	88.88%
Success Rate for Deceptive Instances			66.66%	66.66%	66.66%	66.66%	60%	

1: Correctly Classified, **0:** Incorrectly Classified, **NN:** Neural Network, **DS:** Decision Stump, **DT:** Decision Tree, **NB:** Naïve Bayes, *Refuse to classify

- [6] J. C. Nunnally, P. D. Knott, A. Duchnowski, and R. Parker, "Pupillary response as a general measure of activation," *Perception & Psychophysics*, vol. 2, no. 4, pp. 149–155, 1967.
- [7] S. Porter and L. Brinke, "The truth about lies: What works in detecting high-stakes deception?" *Legal and Criminological Psychology*, vol. 15, no. 1, pp. 57–75, 2010.
- [8] I. Nwogu, M. Frank, and V. Govindaraju, "An automated process for deceit detection," in *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2010, pp. 76 670R–76 670R.
- [9] M. C. Sezgin, B. Gunsel, and G. K. Kurt, "A novel perceptual feature set for audio emotion recognition," in *IEEE International Conference on Automatic Face & Gesture Recognition and Workshops (FG 2011)*. IEEE, 2011, pp. 780–785.
- [10] D. T. Lykken, "The GSR in the detection of guilt," *Journal of Applied Psychology*, vol. 43, no. 6, pp. 385–388, 1959.
- [11] D. J. Krapohl, J. B. McCloughan, and S. M. Senter, "How to use the Concealed Information Test," *Polygraph*, vol. 35, no. 3, pp. 123–138, 2006.
- [12] H. W. Timm, "Analyzing deception from respiration patterns," *Journal of Police Science & Administration*, 1982.
- [13] B. Verschuere, G. Crombez, A. De Clercq, and E. H. Koster, "Autonomic and behavioral responding to concealed information: Differentiating orienting and defensive responses," *Psychophysiology*, vol. 41, no. 3, pp. 461–466, 2004.
- [14] W. Yankee, "An investigation of sphygmomanometer discomfort thresholds in polygraph examinations," *Police*, vol. 9, no. 6, pp. 12–18, 1965.
- [15] F. B. Mohamed, S. H. Faro, N. J. Gordon, S. M. Platek, H. Ahmad, and J. M. Williams, "Brain mapping of deception and truth telling about an ecologically valid situation: Functional MR imaging and polygraph investigation - Initial Experience 1," *Radiology*, vol. 238, no. 2, pp. 679–688, 2006.
- [16] K. Izzetoglu, G. Yurtsever, A. Bozkurt, B. Yazici, S. Bunce, K. Pourrezaei, and B. Onaral, "NIR spectroscopy measurements of cognitive load elicited by GKT and target categorization," in *Proceedings of the 36th Hawaii International Conference on System Sciences*. IEEE, 2003.
- [17] J. P. Rosenfeld, J. W. Ellwanger, K. Nolan, S. Wu, R. G. Bermann, and

- J. Sweet, "P300 scalp amplitude distribution as an index of deception in a simulated cognitive deficit model," *International Journal of Psychophysiology*, vol. 33, no. 1, pp. 3–19, 1999.
- [18] M. Soskins, J. P. Rosenfeld, and T. Niendam, "Peak-to-peak measurement of P300 recorded at 0.3 Hz high pass filter settings in intraindividual diagnosis: complex vs. simple paradigms," *International Journal of Psychophysiology*, vol. 40, no. 2, pp. 173–180, 2001.
- [19] I. Pavlidis and J. Levine, "Thermal image analysis for polygraph testing," *IEEE Engineering in Medicine and Biology Magazine*, vol. 21, no. 6, pp. 56–64, 2002.
- [20] I. Pavlidis, N. L. Eberhardt, and J. A. Levine, "Seeing through the face of deception," *Nature*, vol. 415, no. 6867, pp. 35–35, 2002.
- [21] P. Tsiamyrtzis, J. Dowdall, D. Shastri, I. Pavlidis, M. Frank, and P. Ekman, "Imaging facial physiology for the detection of deceit," *International Journal of Computer Vision*, vol. 71, no. 2, pp. 197–214, 2007.
- [22] D. Shastri, A. Merla, P. Tsiamyrtzis, and I. Pavlidis, "Imaging facial signs of neurophysiological responses," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 2, pp. 477–484, 2009.
- [23] D. Shastri, M. Papadakis, P. Tsiamyrtzis, B. Bass, and I. Pavlidis, "Perinasal imaging of physiological stress and its affective potential," *IEEE Transactions on Affective Computing*, vol. 3, no. 3, pp. 366–378, 2012.
- [24] I. Pavlidis, P. Tsiamyrtzis, D. Shastri, A. Wesley, Y. Zhou, P. Lindner, P. Buddharaju, R. Joseph, A. Mandapati, B. Dunkin *et al.*, "Fast by nature-how stress patterns define human experience and performance in dexterous tasks," *Scientific Reports*, vol. 2, 2012.
- [25] D. C. Derrick, A. C. Elkins, J. K. Burgoon, J. F. Nunamaker, and D. D. Zeng, "Border security credibility assessments via heterogeneous sensor fusion," *IEEE Intelligent Systems*, vol. 25, no. 3, pp. 41–49, 2010.
- [26] N. W. Twyman, A. C. Elkins, J. K. Burgoon, and J. F. Nunamaker, "A rigidity detection system for automated credibility assessment," *Journal of Management Information Systems*, vol. 31, no. 1, pp. 173–202, 2014.
- [27] J. Reid and J. Buckley, *Criminal interrogation and confessions*. Jones & Bartlett Publishers, 2011.
- [28] F. Systems, <http://www.flir.com>.
- [29] Y. Zhou, P. Tsiamyrtzis, P. Lindner, I. Timofeyev, and I. Pavlidis, "Spatiotemporal smoothing as a basis for facial tissue tracking in thermal imaging," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 5, pp. 1280–1289, 2013.
- [30] S. Nicolaidis and J. Sivadjan, "High-frequency pulsatile discharge of human sweat glands: Myoepithelial mechanism," *Journal of Applied Physiology*, vol. 32, pp. 86–90, 1972.
- [31] H. Storm, K. Myre, M. Rostrup, O. Stokland, M. Lien, and J. Raeder, "Skin conductance correlates with perioperative stress," *Acta Anaesthesiologica Scandinavica*, vol. 46, no. 7, pp. 887–895, 2002.
- [32] T. Kamei, T. Tsuda, S. Kitagawa, K. Naitoh, K. Nakashima, and T. Ohhashi, "Physical stimuli and emotional stress-induced sweat secretions in the human palm and forehead," *Analytica chimica acta*, vol. 365, no. 1, pp. 319–326, 1998.
- [33] W. Tool, <http://www.cs.waikato.ac.nz/ml/weka/>.