Deception in the Eyes of Deceiver
A Computer Vision and Machine Learning Based Automated Deception Detection

Wasiq KHAN [a], *Keeley Crockett* [b], *James O'Shea* [b], *Abir Hussain* [a], *Bilal M Khan* [c]

[a] School of Computer Science and Mathematics, Liverpool John Moores University, Liverpool, Byrom Street, L3 3AF, United Kingdom,

Email: W.Khan@ljmu.ac.uk (Corresponding Author), A.Hussain@ljmu.ac.uk

[b] Department of Computing and Mathematics, Manchester Metropolitan University, Manchester, M15 6BH, United Kingdom,
Email: K.Crockett@mmu.ac.uk (Corresponding Author), J.D.Oshea@mmu.ac.uk

[c] University of California Los Angeles, Institute of the Environment and Sustainability (IoES), California CA, USA
Email: M.bilal@ucla.edu

***Abstract*—There is growing interest in the use of automated psychological profiling systems, specifically applying machine learning to the field of deception detection. Several psychological studies and machine-based models have been reporting the use of eye interaction, gaze and facial movements as important clues to deception detection. However, the identification of very specific and distinctive features is still required. For the first time, we investigate the fine-grained level eyes and facial micro-movements to identify the distinctive features that provide significant clues for the automated deception detection. A real-time deception detection approach was developed utilizing advanced computer vision and machine learning approaches to model the non-verbal deceptive behavior. Artificial neural networks, random forests and support vector machines were selected as base models for the data on the total of 262,000 discrete measurements with 1,26,291 and 128,735 of deceptive and truthful instances, respectively. The data set used in this study is part of an ongoing programme to collect a larger dataset on the effects of gender and ethnicity on deception detection. Some observations are made based on this data which should not be interpreted as scientific conclusions, but pointers for future work. Analysis of the above models revealed that eye movements carry relatively important clues to distinguish truthful and deceptive behaviours. The research outcomes align with the findings from forensic psychologists who also reported the eye movements as distinctive for the truthful and deceptive behavior. The research outcomes and proposed approach are beneficial for human experts and has many applications within interdisciplinary domains.**

***Key Words:* - Deception detection, Credibility assessment, Facial micro-gestures, Nonverbal behavior analysis, Eye movements, Psychological profiling**

1. INTRODUCTION

There has been increasing interest in automatic detection of deceptive behavior, particularly from law enforcement, national security, border controls, internet fraud detection and government agencies (Crockett et al., 2017). Several explanations of the term 'deception' have been produced however; the most commonly accepted definition is provided by Vrij (2008) as 'a successful or unsuccessful deliberate attempt, without forewarning, to create in another a belief which the communicator considers to be untrue'. While most of the earlier research on deception detection is based on physiological sensors, such as the polygraph (Larson et al., 1932) or the subjective perception of trained experts undertaking a facial or frame-by-frame analysis (Ekman et al., 1991), each approach can potentially lead to biased human judgments, poor classification of deception and excessive analysis time throughput limits (Bond & DePaulo, 2006).

While many of the technological approaches to detecting deception include non-linguistic vocal sounds (Howard & Kirchhubel, 2011), Electroencephalography (EEG) signals (Roshni & Bhavana, 2014), posture and body movement (Ekman et al., 1991; Sebanz, & Shiffrar, 2009), brain imaging (Kozel, 2005), physiological sensors (Abouelenien et al., 2014), psycholinguistics and gesture (Rosas et al., 2015; Meservy et al., 2005) and thermal imaging (Bashar, & Reyer, 2014). Each method has overlapping as well as distinct indicators of deception. Information content from these indicators has been modelled through various approaches to identify the deceptive behaviour in different scenarios and application domains (O'Shea et al.,2018). Several computational models have indicated facial micro-gestures (Feldman et al., 1979; O'Shea et al.,; Chen et al., 2018; Rothwell et al., 2006; Happy & Routray, 2015), gaze aversion (O'Shea et al.,2018; Freire et al., 2004), eye interactions (Proudfoot et al., 2015; Nunamaker et al., 2016) and eye blink rate (Borza et al., 2018) as important clues within automated deception detection. However, the reliability and efficiency in terms of distinguishing deception and truthful behaviours for these methods (human and machine) are significantly lower for real-time practical applications (Abouelenien et al.; Mendels et al., 2017; Zimmerman, 2016).

The psychological studies in general, indicate that NVB does not contain useful clues to catch the deceptive behaviour. For instance, meta-analysis by Bond & DePaulo (2006) clearly indicate that audible information leads to better human judgements about deception detection as compared to visual information. However, several relevant works have been addressing the significance of eye interactions and gaze attributes to discriminate deceptive behaviours (Marchak, 2013; Fukuda; Dionisio et al., 2001). Maintaining eye contact between interviewer and interviewee increases cognitive load and debilitate deceivers resulting more cues to deceit and therefore, supporting distinction between deceivers and truth tellers (Vrij, Mann, Leal & Fisher, 2010). The main reason in establishing and maintaining eye contact is to enrich information possessed within the eyes in social interactions, emotional state, beliefs and desires (Frischen et al., 2007). Some of the existing studies also reported

that criminal psychologists and investigative professionals use eye contact and gaze aversion to detect and filter deceiving behavior (Vrij, 2008; Taylor & Hick 2007). For instance, Mann., Vrij., and Bull (2004) revealed the use of averted gaze and eye contact as one of the clues reported by 73% (72 out of 99) of British police personnel, to catch the lying behavior of a person. However, these studies and the corresponding outcomes do not directly validate the argument of significance of such NVB in deception detection or improved human judgements.

Despite the above computational studies support the argument that eye interactions and eye related movements might contain some distinguishing clues for the deception detection, the focus on eye movements and gaze for automated deception detection is limited specifically in terms of multi-dimensional analysis and yet to emerge in real-time practical deception detection systems. Secondly, these methods either focus on specialized attributes (e.g. blinks, gaze) or a combination of limited expressions (e.g. eyes, face, body movements) and hence the trained classifiers lacks the efficiency to distinguish the truthful and deceptive behaviours (Abouelenien et al.; Mendels et al., 2017; Zimmerman, 2016). Furthermore, there are very limited, publicly available video datasets featuring participants engaged in roleplaying truthful and deceptive behaviours (e.g. Lloyd et al., 2019). Strict ethical approvals are required to undertake experiments which capture personal, sensitive data and due to the GDPR, participants are now more informed and empowered on whether their data can be used only for a current study or be made publicly available. Specifically, in case of video dataset where the personal identification and privacy are the major ethical concerns. Guozhen (2015) reported that common interpersonal verbal and non-verbal attributes between deceivers and truthful subjects such as vocal dynamics and smiling behaviours are difficult to classify. In addition, the embedded deception by deceivers within the truthful behaviour makes it further challenging to classify the deceptive behaviour (Bond & DePaulo, 2006). Hence, experiments must be defined to clearly establish the ground truth answers for deceivers and truthful scenarios.

This proposed work describes progressive research of an intelligent system that is designed to measure the degree of deception of an individual whilst they are being asked a number of questions in a role-playing interview. The system presented in this paper was inspired by earlier work undertaken as part of iBorderCtrl (Intelligent Portable Control System) which was a research and innovation project that was aiming to evaluate if it was possible that state of the art technologies could be integrated to enable more efficient thorough border control for third country nationals crossing the land borders of EU member states (Crockett et al., 2017). A risk assessment module in iBorderCtrl was used to combine scores obtained from novel and existing systems and biometric tools to classify the traveler in terms of risk, thus supporting the decision-making of the border guard at the land border crossing point. In the prototype system, the risk assessment of a potential traveler was calculated using input from all tools and this depended on the technological readiness level (TRL) of the tool and the data quality. Subsequently, tools with a low TRL, had little impact on the overall risk assessment and if data quality was not good, the result from the tool was excluded from any risk-based calculation.

This paper describes, for the first time, a methodology and experimental design to identify the micro-movements within the eyes and face which can be combined together to distinguish the truthful and deceptive behaviours. The work attempts to investigate *which specific fine-grained eye and facial micro-movements contain distinguishing clues for the classification of deception derived through Non-verbal Behaviour (NVB)?*

The motivation behind this research stems from the need for accurate, non-biased deception detection in automated deception detection systems. Although such systems should only be used as part of a human in the loop system, it is critical to classify the right suspects as deceptive without miss-classifying truthful people in order to build trust in the system. In this work, video NVB data is collected from consenting individuals from a holiday-inspired role-playing exercise while they are being asked a series of questions about items they packed in their suitcase. Each individual is asked to behave truthfully or deceptively. The data is pre-processed into a set of one-second image vectors which are labelled with the ground truth of each question to formulate a dataset where it is not possible to re-identify any individuals. This experiment is used to build different classification algorithms that include multi-layer perceptron neural networks, support vector machines, and random forests to classify each time slot as

being deceptive or truthful. Clustering-based attribute analysis indicates the significance of eye movements compared to facial micro-movements. Further experiments were conducted to utilize the identified eyes features to train multiple classification algorithms for the solid validation of the research question.

This manuscript is organised as follows: Section 2 provides a description of prior work in the field of deception detection systems with emphasis on automation and the role of eye interactions in deception detection. Section 3 presents the psychological studies which have been focusing on the distinctive clues within the eyes of deceivers. Section 4 describes the detailed proposed methodology for feature engineering and the creation of deception models from NVB. The experimental methodology, including a description of the dataset and attribute analysis are discussed in section 5. Performance evaluation and results are presented in section 6 followed by a detailed discussion on the importance of eyes and facial micro-movements in deception detection from NVB for both truthful and deceptive subjects.

2. RELATED WORK

There has been a long history of human interest in identifying deceptive behaviour. Trovillo (1939) addressed the historic evidence date back to the Hindu Dharmasastra of Gautama (900 – 600 BC) and the Greek philosopher Diogenes (412 – 323 BC). In 1921, Larson invented the Polygraph (Larson et al., 1932), which has been considered as one of the popular methods for lie detection and works by measuring physiological changes in a person in accordance with stress factors. Typically, the polygraph instrument captures physiological changes such as pulse rate, blood pressure and respiration that can be interpreted by psychological experts to identify truthful or deceptive behaviour. With respect to different scenarios, a polygraph test takes up to four hours which leads to limitations on it use in real time conditions. Research studies have been supporting the validity of the polygraph as well as criticizing its use in specific cases. A meta-study by Axe et al., (1985) found 10 studies from a pool of 250 (that were sufficiently rigorous to be included), indicated that the controlled question test could perform significantly better than chance under specified narrow conditions. However, the deception classification contained a high number of false positives, false negatives and inconclusive instances. In addition, substantial information about the interviewee's background (e.g. occupation, work record and criminal record) was required to be captured before the examination in order to construct a good set of control questions.

Vocal cues, voice stress and acoustic features have also been employed as indicators to distinguish the act of deceit (Hirschberg et al., 2005). Distinctive additional micro tremors appear due to cognitive overload during the deceptive behaviour (Walczyk et al., 2013). However, the performance of deception detection using voice stress analysis has been described as "charlatanry" (Eriksson & Lacerda, 2007). Likewise, linguistics has also investigated the changes in language and its structure to classify signs of deception. Linguistic inquiry and word count analysis for deception detection revealed that truth tellers' statements contain more first-person pronouns and self-references (e.g. mine, our) while liars statements contain more words referring to certainty (e.g. totally, truly) and to other- references (they, themselves) (Eriksson & Lacerda, 2009; Abouelenien et al., 2017). A variety of statistical features including mean length of sentence, mean length of clause and clauses per sentence have been extracted from transcribed interviews to evaluate the linguistic hypothesis that liars use less complex and less detailed sentences.

Vrij et al., (2009) reported on the use of thermal imaging of the facial periorbital area to analyse the variations in blood flow specifically when answering unexpected questions. A thermal facial pattern-based approach introduced by (Pavlidis et al., 2002) claims the deception detection accuracy is comparable to that of polygraph tests. Likewise, a thermodynamic model of blood flow variations using the thermal images of facial periorbital area to detect the deceptive behaviour is presented in (Pavlidis & Levine, 2001, 2002). Relationships between different facial emotions (such as stress, fear, and excitement) and deceptive behaviour using thermal imaging is addressed in (Merla & Romani, 2007). Bashar & Reyer, 2014) used thermal variation monitoring of the periorbital region and a nearest neighbor classifier that was trained on a high-dimensional feature vector extracted using an average value from each sub-region to detect deception. Experimental results indicated that the classification accuracy did not differ significantly from a random chance distribution based on leave-one-person-out

methodology and five-fold cross validation.

In addition to the aforementioned methods, analysis of eye interactions and facial micro-expressions also have been studied as a non-verbal deception detection method (Ekman, 2001). During the act of deceit, relatively short involuntary facial expressions may appear that can be helpful to detect deceptive behaviour. Furthermore, the analysis of facial expressions in terms of asymmetry and smoothness features (Ekman, 2003) indicate their relationship with the deceptive behaviour. Face orientation and intensity of facial expressions is also used to classify the act of deceit (Tian et al., 2005). Likewise, geometric features (Owayjan et al., 2012) and micro-expressions (Pfister & Pietikäinen, 2012) extracted from the facial data have also been used to classify the deceptive behaviour. Related research in (Pons & Masip, 2018) indicated the usefulness of facial micro-gestures towards the identification of comprehension levels. Buckingham et al., (2014) used artificial neural networks sequentially to identify the micro-gestures and perform the classification respectively. Pérez-Rosas et al., (2015) proposed the multi-model deception detection methodology that used a novel dataset acquired from real public court trials. A variety of linguistic and gesture modalities including facial features were combined together to classify the deceptive behaviour. Results reported a classification accuracy between 65-75% with varying combinations of modalities. Furthermore, the results indicated that the system outperformed human experts in terms of correct identification of deceptive behaviour. One of the recent machine-based research studies that uses the direction of gaze, eye movements and blink rate to distinguish the truthful and deceptive behaviours is presented in (Borza, 2018). The research outcomes indicated the normalised eye blink rate was an important clue of deception detection. Research carried out in (Marchak, 2013; Nunamaker et al., 2016; Levine, 2014), (Schuetzler, 2012; Kumar, 2016; Pak & Zhou, 2011; Lim, 2013) also indicate the significance of eye interaction and associated corresponding features towards effective deception detection. Eyes blink rate, pupil dilation and gaze are the most common examples of such a feature set. Research studies indicate the relationship between these attributes and cognitive effort variations in deceptive and truthful subjects (Fukuda, 2001). Like other psychological clues for deception detection, additional cognitive efforts performed by deceivers undergo additional cognitive processes compared to truthful individuals that leads to an increased pupil diameter for deceivers (Proudfoot et al., 2015; Dionisio et al., 2001). In a similar study by Marchak (2013), compared to truthful participants, a suppressed eye blinking rate is noticed for participants involved in a mock crime to transport an explosive device to be used for a disturbance.

3. PHYSIOLOGICAL EXPERTS AND DECEPTION DETECTION

The psychological research on behavioral cues to deception is longstanding. The literature on cues to deception mainly focuses the lies within the social opinions, facts, emotions, transgressions, and serious matters such as criminal investigations while analyzing diverse behavioral aspects (DePaulo et al., 2003). Deceptive behavior is cognitively more demanding than telling the truth since liars are required to invent a story and must monitor their construction in order to maintain consistency with what the observer knows or might know in further investigation (Vrij et al., 2011). Therefore, application of cognitive techniques such as seeking eye contact from the interviewee or telling the story in a reverse chronological order that leads to an increase in cognitive load and may produce some clues for detecting a lie. Similarly, avoidance and denial strategies that liars employ to distance themselves more from events, can lead to a deception detection strategy. Deceivers experience more negative emotions like anxiety and arousal during lying as a natural human response from their nervous system (Siering et al., 2016). This is also associated with attempts to overcompensate by controlling their behaviours. Likewise, truthful statements have more contextual embedding (Kleinberg et al., 2018) as compared to deceptive ones. A meta-analysis (DePaulo et al., 2003) conclude that liars are less forthcoming and tell less compelling tales as compared to truth tellers. While the literature presents in-depth analysis of variety of clues to deception detection in diverse scenarios, these methods have limitations due to biased human judgments (Bond & DePaulo, 2006), poor identification of deceptive clues and large amounts of time needed to analyse various feature combinations specifically in real-time scenarios.

Previous research also indicates the occurrence of variations within NVB during the cognitive interviews of suspects, regardless of any specific reason and interestingly, differently in truthful and

deceptive subjects (Frosina et al., 2018). A similar work (Vrij et al., 2008) reported that use of cognitive load during interviews affects the NVB and helps in making the deceptive judgements. For instance, blinking rate increases in argument of cognitive load while direct eye gaze decreases (Vrij et al., 2008; Frosina et al., 2018). However, some studies reported increasing direct eye gaze in perspective of cognitive load (Mann et al., 2012). This contradiction might be due to several factors that include subjective affect, experimental and/or data capturing design and other confounding factors (e.g. interview environment). A study on deliberate eye contact within the truthful and deceptive subjects is presented in (Mann et al., 2012) where the passengers were asked to produce truthful and deceptive statements about their future planned travel. The amount of time was recorded that the interviewees were looking away from interviewer. Research outcomes observed deliberate eye contact in liars, compared to truthful subjects. Another psychological study by Mann et al., (2013) on deceptive and truthful NVB is conducted using an additional interviewer who remained silent but exhibits neutral, suspicious or positive attitudes during the interview process. The outcomes of the study indicated that the truthful participants provided significantly more detailed answers as compared to the deceptive subjects but only while the second interviewer behavior was supportive. However, one of the key findings of this research was the duration of eye contact for both groups which indicated that the liars produced more deliberated eye contact than truth tellers. These psychological studies give an indication of comparatively deliberated eye contact from deceptive subjects, however, does not investigate the reason why liars produce more deliberated eye contact to the interviewer. The investigation is carried out in (Mann et al., 2013) that hypothesizes that a deliberated eye contact from deceptive subjects is due to their convincing behavior. The research indicates two key findings: a) deceptive subjects produced comparatively more deliberated eye contact, b) deceptive subjects reported that the reason behind the solid eye contact is trying to convince the interviewer and to conclude whether they (deceivers) were believed or not?

While aforementioned studies specifically the technological models (presented in Section 2) support the argument that eye related attributes possess some useful clues for the deception detection, psychological studies generally conclude the NVB as weak identifiers for the deception detection. For instance, a meta-analysis study (Bond & DePaulo, 2006) revealed that human can make better deception detection judgments by using audible information as compared to visual clues. Bond & DePaulo undertook a comparative experiment based on previous studies that concluded that distinguishing deceptive from truthful behavior is superior when measured through audiovisual or audio contents only rather than visual clues. DePaulo et al., (2003) presented a comprehensive analysis based on 158 clues to deception to investigate the behaviour differences between the deceptive and truthful subjects. While investigating whether the deceptive accounts less compelling than truthful ones, the study indicated the significance of verbal and vocal clues in distinguishing truthful and deceptive behaviours as compared to non-verbal clues. The study also indicate that the liars have more tensed vocal, high pitch voices and pupil dilation than the truthful subjects. Likewise, Ekman and Friesen (1969) found that compared to visual information that can easily be controlled by deceiver, more clues exist within the body movements. Subsequently, this was contradicted by the Bond & DePaulo (2006) meta-analysis study. Sporer and Schwandt (2007) also concluded that there is no evidence of NVB (specifically gaze aversion, eye contact) as distinctive deception indicator. However, factors like content, motivation, preparation and experimental design are comparatively more important in the deception detection context.

The criticisms of the psychological community that there are no meaningful single non-verbal indicators of deception (such as averted gaze) was addressed by the Silent Talker (ST) that was designed to model multiple (typically 36) NVB channels to classify the level of deception through trained ANNs (Rothwell et al., 2006, 2007). Unlike other deception detectors that deploy the underlying explanatory model, ST uses the conceptual modelling of NVB speculating that interviewee's NVB will be affected by certain mental states (e.g. stress, cognitive load behaviour control, duping delight) associated with deceptive behaviour. Modelling of eye related micro-gestures in ST is divergent to most of the existing research that focus on eye tracking and associated impacts on psychological attributes to detect the deceptive behaviour. A micro-movement defined in ST represents a very fine-grained non-verbal

gesture such as the face upward movement, left-eye half closed, right eye fully closed etc. Micro-movements are significantly different from micro-expressions (proposed in other systems), because they are much more fine-grained and require no functional psychological model of why the behaviour has taken place (Rothwell et al., 2006, 2007).

4. METHODOLOGY

4.1. Data Collection

Following the ethical approval, data were collected from 100 participants in total (50 truthful, 50 deceptive) through asking consenting adults to role-play either a truthful or deceptive scenario about packing a suitcase and taking it to an airport for a holiday. A truthful scenario involved being themselves and answering questions truthfully about a planned trip they were taking in the future. Those undertaking a deceptive scenario, were given one of four randomly selected short, fake, descriptive profiles of individuals who they were asked to roleplay. When asked questions by the Avatar they were therefore asked to answer deceptively in accordance with the fake profile.

Table 1. List of questions used during the Avatar interview for video data capturing

Q. No	Question Contents	Q. No	Question Contents
1	*What is your family name?*	7	*Which country does this person live?*
2	*What is your first name?*	8	*What is in your case?*
3	*When were you born?*	9	*Have you seen any posters of prohibited items?*
4	*Where were you born?*	10	*Are there any items from the lists of prohibited items in your case?*
5	*What is your current citizenship?*	11	*How many items are in the case?*
6	*You will be asked at the border to name a person who will confirm your identity. For now, just tell me your relationship with this person.*	12	*If you open the case and show me what is inside, will it confirm that your answers were true?*

All participants were then interviewed by an Avatar border guard with 12 travel related interview questions (Table 1) and the answers were based on whether they were playing a truthful or deceptive scenario. The interviews were conducted using a well-established 'Wizard of Oz' methodology where a human (known as the Wizard) manually controls the simulated Avatar (i.e. border guard in this case) to conduct the interview. The interview question flow is controlled in such a way that the interviewee experience is similar as if they were interacting with a real Avatar. Detailed information about truthful and deceptive scenarios, and 'Wizard of Oz' methodology we use for conducting the interviews is provided in our previous work (O'Shea et al., 2018). In total, 1200 short video clips were recorded from 100 interviewees (i.e. for 12 questions, one video per each question/answer from 100 participants). Each interview (i.e. set of 12 questions and answers) lasts between 3 to 6 minutes depending upon the length of answers produced by the participant to Avatar. Table 1 shows a list of questions used during the interview process. In total, there are 100 participants containing a balanced distribution for truthful (T: 50) and deceptive (D: 50) with mixed gender: male (M), female (F) and ethnicity: Asian/Arabic (A), European (E). It can be noted that the dataset is mixed in terms of gender and ethnicity and that the coverage across these two factors is quite balanced in the dataset as shown below:

	M	F
T	32	18
D	28	22

	A/A	EU
T	23	27
D	25	25

In this paper, the focus was on the inclusivity of a wide range of participants to investigate which specific fine-grained eye and facial micro-movements contained distinguishing clues for the classification of deception. Therefore, the inclusion criteria for volunteer participants was to be aged 18 or over and did not include vulnerable, participants with mental illness or learning difficulties. There was no payment to the participants and therefore we did not exclude volunteers based on any other

criteria in accordance with the University Ethical procedures. The balance of participants gender and authenticity was that naturally occurring within the diverse population from which the volunteers came. This was intended to avoid the potential confounding factor of a narrow pool of ethnicities.

4.2. Video Data Processing: Silent Talker Overview

The Silent Talker (ST) system (Rothwell et al., 2006, 2007), which uses features extracted from the NVB of interviewees to determine whether they are deceiving or telling the truth, is used as a basis to create a dataset of deceptive and truthful behaviour for this study. ST uses multifaceted interactions between multiple channels of micro-gestures over time to determine whether the behaviour is truthful or deceptive. Over a time interval, typically one second, complex combinations of micro-gestures can be extracted from the interviewee's behaviour. The core, original architecture is shown in Fig.1.

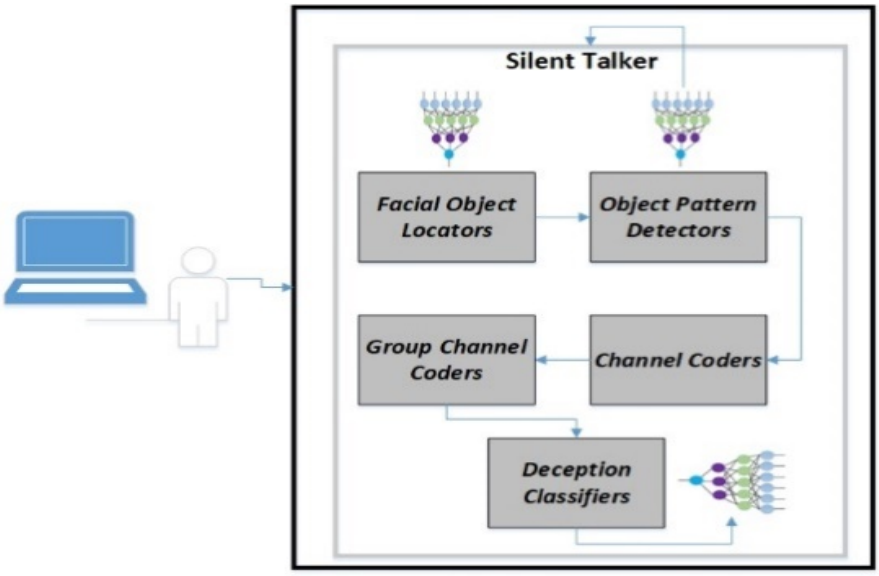


Fig. 1. Silent Talker core architecture for video data processing, object locators, state information extraction and deception classification

When processing a video interview, for a frame to contain useful data, the location of the face and the quality of the image must be determined. Once the face has been located within the frame, facial objects are found (e.g. eyes, nose) and an assessment takes place to determine if a sufficient number of facial objects found in frame. The frame, classified as 'good' if a required number of objects are detected, is then forwarded to a series of pattern detectors. The presence, location and states of each facial feature are transferred from the pattern detectors to the frame vector. This frame vector passes to the Channel coder which extracts the channel data and normalizes the data and passes to Channel accumulator which performs an aggregation of the channel data from all frames contained within the duration of the current timeslot, which may be fixed or variable in length. The Channel accumulator produces a single, normalized vector known as grouped channel data. The grouped channel data then is classified by the deception classifiers and classification of each frame vector is obtained. ST is patented and more information can be found in previous works (Rothwell et al., 2006, 2007).

Optimized Object Locators: In this work, the original face detection within facial object locators was replaced with the well-known OpenCV pre-trained library (Bradski, 2000) publicly available for the facial landmark recognition. It uses the Haar Cascade algorithm proposed by (Viola & Jones, 2001) that uses Haar-like features to encode local appearance of objects using two dimensional Haar functions. These functions consist of two or more rectangular regions enclosed in a template. The feature value of a Haar-like feature with - rectangles is obtained using:

$$f = \sum_{i=1}^{k} \omega^{(i)} . \mu^{(i)} \tag{1}$$

Where $\mu^{(i)}$ represents the mean intensity of the pixel in the image enclosed by the i^{th} rectangle with

mean value of μ. The $\omega^{(i)}$ presents the weight associated to the i^{th} rectangle. The weights assigned to rectangles usually set to default integer numbers such that:

$$\sum_{i=1}^{k} \omega^{(i)} = 0 \quad (2)$$

Detailed mathematical formulation of Haar Cascade with varying Haar-like features can be found in (Viola & Jones, 2001; Li et al., 2002; Jones & Viola, 2003). The pseudo code used in this work for extracting the channel data from video stream using ST and Haar Cascade is presented in Algorithm 1.

Algorithm 1. Channel data extraction from video files using Silent Talker and Haar Cascade

Inputs: Video data stream ***(v)***
Output: Vector **M**: extracted Cannel data
Procedure:
Set ***s-index*** to 1st video frame in ***v***
Step 1: Take one Slot (***ʂ***: 1 sec) of ***v***
For-each video frame (*f)* in ***ʂ***
Search for '*face'* using **Haar** Cascade:
If a 'face' is identified
$\mathbf{F_c}$ ← Rectangular coordinates around the 'face'
Search for 'eyes' within $\mathbf{F_c}$ using **Haar** Cascade:
If two 'eyes' are identified
L_c ← Left eye coordinates
R_c ← Right eye coordinates
Obj[] ← ST: **Object Locator*(**F_c, L_c, R_c*)***
goodFrame ← ST: **is_Good_Frame*(Obj[],f)***
if (***goodFrame***)
M ← ST: **Channel Coder(*obj[], f*)**
End loop
Increase ***ʂ-index*** by 1 to get next overlapped slot ***ʂ***
Go to Step 1 until last *ʂ* in *v*

Table 2. NON-VERBAL CHANNEL LIST EXTRACTED FROM THE VIDEO DATASET USING ALGORITHM 1

Channel NO	Channel Name	Channel Category	Channel No	Channel Name	Channel Category
1	face vertical movement (fvm)	face	19	left eye shift (lshift)	eyes
2	face horizontal movement (fhm)	face	20	left eye closed (lclosed)	eyes
3	face scale (fs) change (forward/backward movement)	face	21	left eye half left (lhleft)	eyes
4	face blush (fblu)	face	22	left eye half right (lhright)	eyes
5	face blanch (fbla)	face	23	left eye half closed (lhclosed)	eyes
6	face upward movement (fum)	face	24	right eye blink (rblink)	eyes
7	face downward movement (fdm)	face	25	right eye left (rleft)	eyes
8	face left movement (flm)	face	26	right eye right (rright)	eyes
9	face right movement (frm)	face	27	right eye shift (rshift)	eyes
10	face forward movement (ffm)	face	28	right eye closed (rclosed)	eyes
11	face backward movement (ffm)	face	28	right eye half left (rhleft)	eyes
12	face vertical shift (fvs)	face	30	right eye half right (rhright)	eyes
13	face horizontal shift (fhs)	face	31	right eye half closed (rhclosed)	eyes
14	face vertical shift with noise (fvsn)	face	32	face movement clockwise (fmc)	face angle
15	face horizontal shift with noise (fhsn)	face	33	face movement anti-clockwise (fmac)	face angle
16	left eye blink (lblink)	eyes	34	face movement angle-change (fma)	face angle
17	left eye left (lleft)	eyes	35	face movement right (fmuor)	face angle
18	left eye right (lright)	eyes	36	face movement left (fmuol)	face angle

4.3. Feature Vectors (Dataset)

The extracted dataset (referred to as DT-Deception in rest of the manuscript) contains 36-dimentional numerical features (facial and eye micro-movements) that were generated from participants' video frames using Algorithm 1, to produce numerical anonymized image vectors. These numerical vectors contain no personal identifiable data and cannot be used to re-identify a participant thus creating a truly anonymized dataset. Each vector represented a one second time slot of the video and represents the facial and eyes micro-movements (channel state information). A feature vector comprised of 36 facial channels of NVB (Table 2) with 2 additional attributes (i.e. gender and ethnicity) and was labelled with the ground truth (deceptive, truthful) based upon the participant scenario. We removed all the duplicate vectors from the extracted dataset. In total, 1,26,291 deceptive vectors and 128,735 truthful vectors were generated from 1200 videos (i.e. 12 videos per participants). As stated in Algorithm 1, only 'good frames' are accumulated, thus the cleaned dataset contained only valid feature vectors. Furthermore, the dataset is normalised [-1 to 1], representing the state of the NVB channels produced by ST. Detailed information about 'good frames', data scaling and time slots is available in our previous works (O'Shea et al., 2018; Rothwell et al., 2006, 2007; Buckingham et al., 2014).

4.4. Analysis of Facial and Eye Micro-Movements

Identification of the most significant features from granulated NVB micro-movements, extracted through the channel extractor was one of the fundamental undertakings of the proposed work. Various psychological and computational studies have previously addressed the critical elements for behavioral distinction such as eye gaze and facial features as clues to deception. Therefore, the assessment of feature importance through multiple well-known clustering techniques was performed using Principal Component Analysis (PCA) and Self-Organizing Maps (SOM).

The component loadings in PCA, represent correlation coefficients between the extracted features of DT-Deception and the principal components (obtained through PCA). The component rotations provide the maximized sum of variances of the squared loadings. The absolute sum of component rotations gives the degree of importance (as in Fig. 2) for the corresponding features in dataset. The first 20 PCs cover the 95% variance in this case and hence, the rest of the 18 PCs were eliminated. Then the absolute sum of rotations across the first 20 PCs is calculated that represent feature ranking within DT-Deception. A detailed description of the workflow of PCA and attribute loadings can be found elsewhere (Hervé & Williams, 2010).

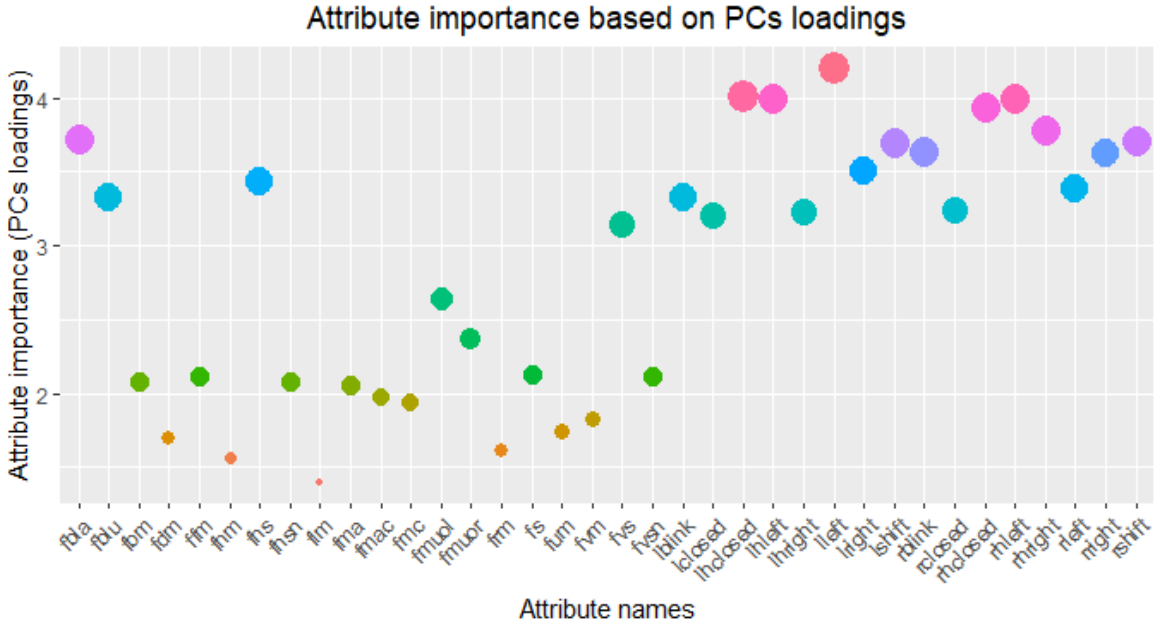


Fig. 2. Importance measure for facial and eyes micromovement using components' loadings from PCA. Small size of circle indicates less importance for the corresponding feature (x-axis) and vice versa.

Fig. 2 summarizes the importance of features representing the eyes and facial micro-movements (listed in Table 1). Firstly, there is a clear difference in the importance measure between most of the eyes related features (e.g. rshift, rright, rleft, rhclosed, lshift, rshift, rhright) and facial features (e.g. fbm, fdm, fhs, flm, fvm, etc.). For instance, there is a significant difference between the face and eyes related feature importance (p-value = 9.3×10^{-6}). The mean importance measure (i.e., feature loading as identified by the PCA) for eyes related channels was 3.42, higher than that of facial channels (2.39). These observations are analogous to the previous psychological as well as computational studies (Marchak, 2013; Nunamaker et al., 2016; Levine, 2014), (Fukuda, 2001; Dionisio et al., 2001), (Schuetzler, 2012; Kumar, 2016; Pak & Zhou, 2011; Lim, 2013) which mainly focus on eye interactions in deception detection and adaptive profiling activities.

One of the difficulties in dealing with high-dimensional feature space is the effective visualization and interpretation of relationships between the variables. To overcome this, we use the SOM with the ability to summaries the high-dimensional data into a typically two-dimensional space. The outcomes from SOM are used to further investigate the patterns and inter-relationships within the feature-space specifically at the individual levels of micro-movements.

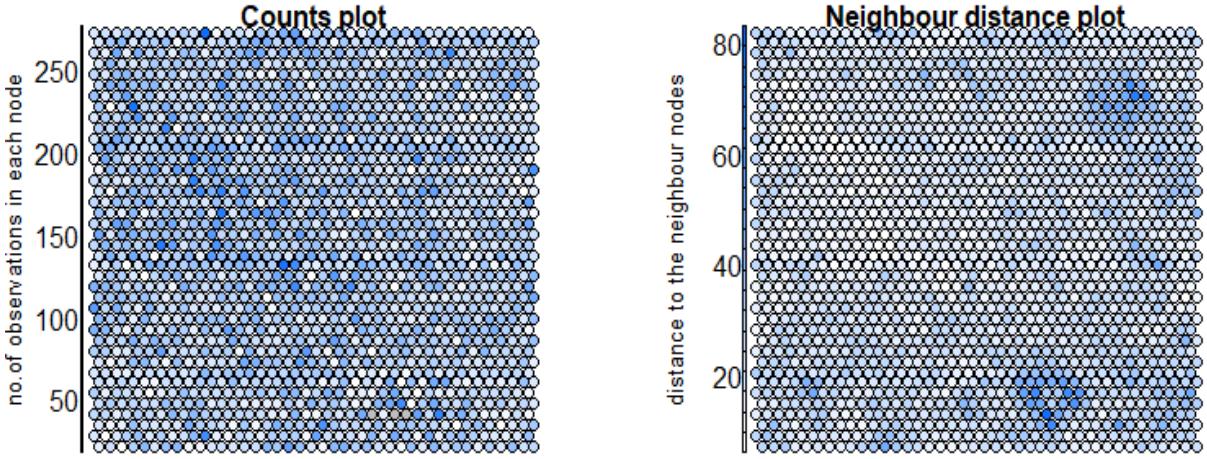


Fig. 3. (a: left side) SOM count plot for SOM map quality showing how may feature vectors assigned to each node. Color intensity increased towards blue with increasing number of observation (b: right side) SOM neighbor distance plot where more bluish color represents higher distance between neighboring nodes and vice versa

The count plot over the entire set of features is shown in Fig. 3(a) which represents the distribution of samples (i.e. DT-Deception vectors in our case) per node within the SOM model. It gives an indication of map quality with the ideal one presenting a homogenous colour distribution and minimum empty (i.e. white colour) nodes. Fig. 3 (b) demonstrates the neighbor distance plot representing the unified distance matrix and uses the Euclidean distance between the codebook vectors of neighboring neurons in SOM. There are some regions demonstrating the comparatively high intensity (darker blue) colour and hence giving the indication of two groups (i.e. truthful and deceptive in our case). However, the overall variations in colour distribution are low indicating the non-linearity in the deception detection problem.

A heat map representation produced by the SOM is the most efficient tool to provide the two-dimensional visualization for multiple variables' distributions within the trained model. More specifically, in this study, heat map is used to investigate the inter-relationships between the facial micro-movements that can be cross compared with the PCA based feature importance. Fig. 4(a) shows the SOM heat maps representing most (12) of the facial features in the DT-Deception. Interestingly, patterns for facial features in Fig. 4a indicate likely correlations. For instance, fum and ffm are correlated to fdm and fbm respectively. Likewise, fvsn and fmc are likely to be correlated with fvm and fmac respectively. Such correlations via SOMs also demonstrate consistency with PCA based feature importance (Fig. 2), indicating low importance for most of the facial features and discernible overlaps with the SOM based correlated variables (e.g. flm. fbm, fmac, fdm, fum, and fvm).

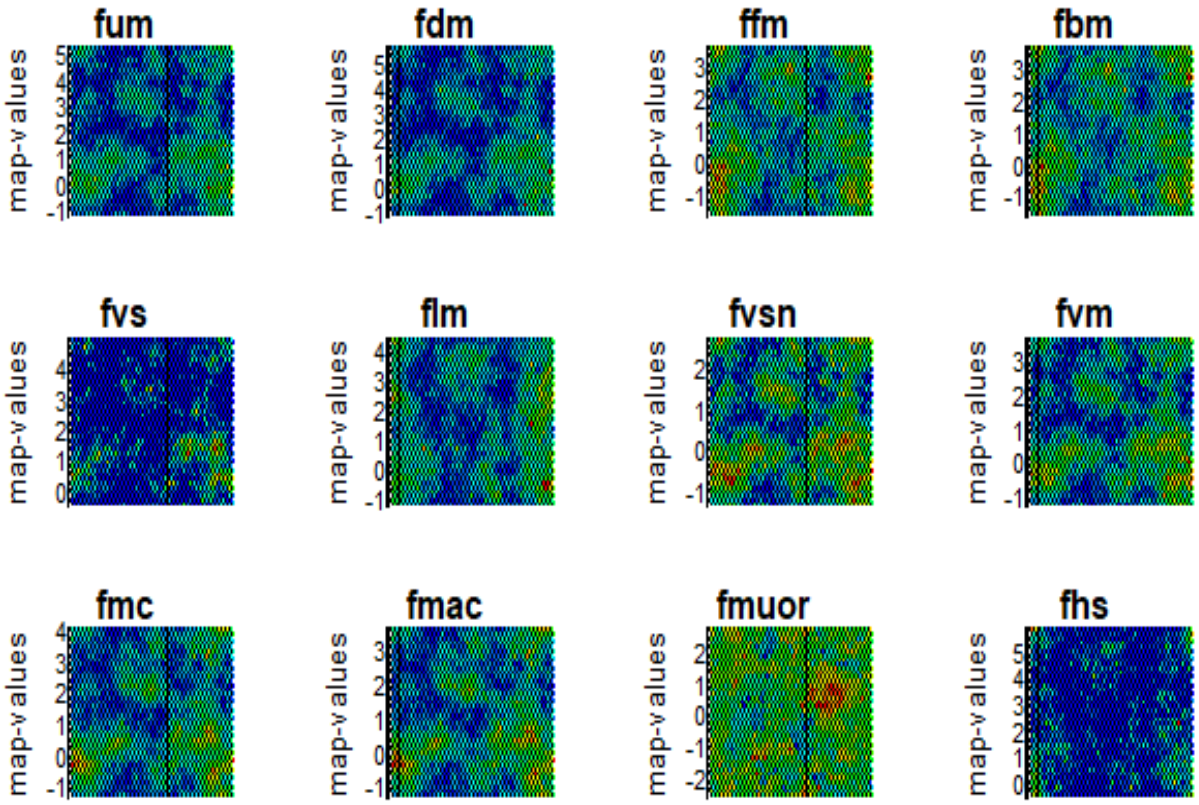


Fig. 4 (a). Heatmap representation of individual facial micro-movements using SOM. Colour intensity (blue to red) indicates the map values from low to high respectively

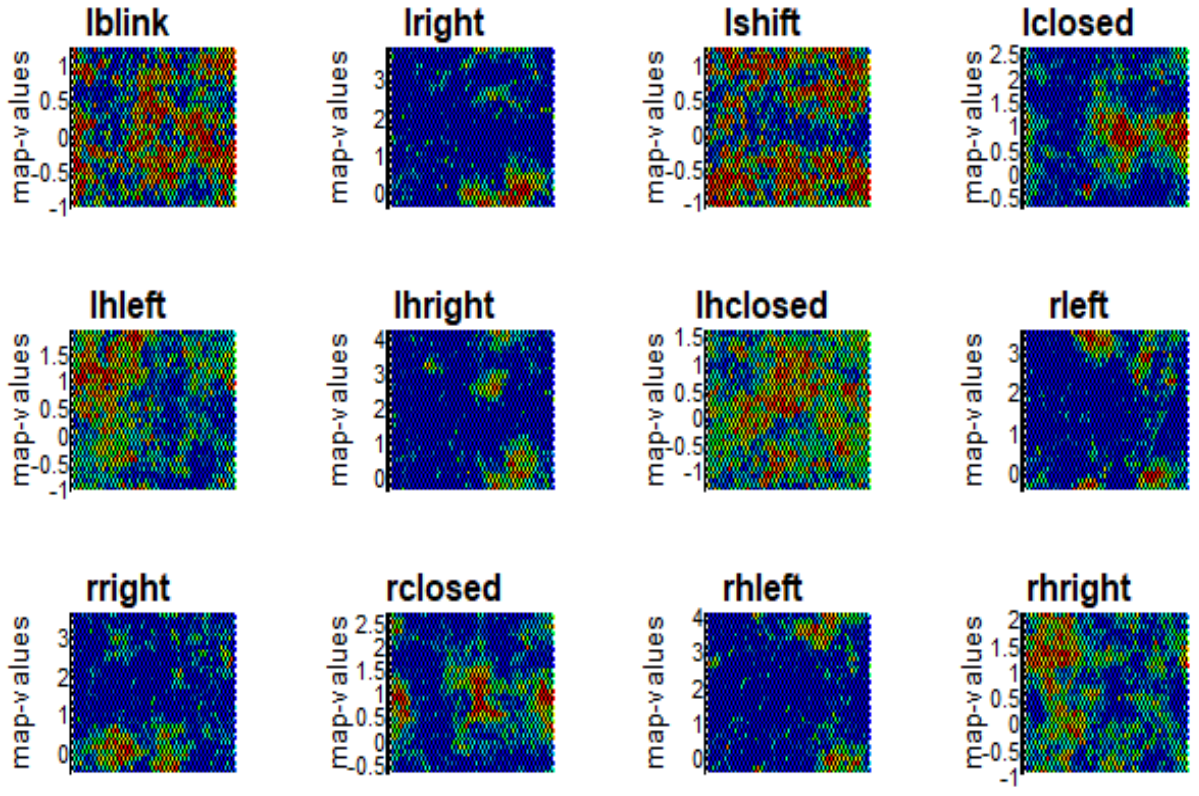


Fig. 4 (b). Heat map representation of individual eye related micro-movements using SOM. Colour intensity (blue to red) indicates the map values from low to high respectively

In contrast to heat maps representing the facial movements in Fig. 4(a), Fig. 4(b) demonstrates discrimination within the SOM heat maps pattern distribution for eyes related features. This indicates little to no correlation between most of the eye related features that might be useful for the deception classifiers to distinguish the truthful and deceptive behaviours more effectively. The maps also credence the PCA based feature importance that clearly indicates the significance of eye related features (Fig. 2).

4.5. Deception Classification

In this study, multiple machine learning (ML) algorithms are used for the deception classification and the performance comparison. The following ML algorithms are used for the comparative analysis of final deception classifiers in this study.

Artificial Neural Networks (ANN): Neural Network is a problem-solving methodology based on the connectionist paradigm. They are comprised of networks of interconnected neurons, whose weights are adapted until a solution emerges. In the current study, a feed-forward multi-layered ANN with fully connected input layer, two hidden layers and an output layer is used. The resilient backpropagation algorithm (Aristoklis et al., 2005) is used for the network training that updates the weights based on sign of corresponding derivate to find out the local minima of the error function. A separate learning rate is used for each weight that changes during the training process. This resolves the problem associated with traditional backpropagation algorithm that use an overall learning rate for the entire networks and training process. Further details about mathematical formulation and different variations are explained in (Aristoklis et al., 2005).

Random Forest (RF): Random Forest (RF) has wide application areas and is suitable for both regression and classification tasks. Random forests comprise of multiple decision trees, each of which acts as a weak classifier, typically characterized by poor prediction performance, however in aggregate form, it offers robust prediction. Therefore, this classifier can be thought of as a meta-learning model. Further technical details of RF and explanations of feature bagging and decision trees structures can be found in (Breiman, 2001; Ho, 2002). The RF algorithms efficiently and effectively produces partitions of high-dimensional features based on the divide-and-conquer strategy, over which a probability distribution is located. Moreover, it permits density estimation for arbitrary functions, which can be used in clustering, regression, and classification tasks. Classification results are obtained by averaging the decisions formed through the layers of the forest, permitting the collective knowledge of the decision-tree learners to be incorporated. Equations 3 and 4 summarize the RF.

$$f(x) = \frac{1}{m} \sum_{i=1}^{m} f(x, x_i p) \quad (3)$$

where x is the partial dependence variable and $x_i p$ refers to the data variable.

$$f(x) = log t_j - \frac{1}{J} \sum_{k=1}^{J} (log t_k(y) \quad (4)$$

where '*J*' refers to the number of classes (2 in our case), and '*j*' refers to the individual class (i.e. truthful/deceptive in this study). In addition, t_k belongs to the proportion of total votes for class '*j*'.

Support Vector Machines: Support Vector Machines (SVMs) are a type of supervised learning and can be used for classification and regression problems. An SVM is based on soft margin classification as stated by (Cortes & Vapnik, 1995), which lends itself on concepts of statistical method theory. Given a training dataset containing instance-label pairs $\{(x_1, y_1), \dots, (x_N, y_N)\}$ where $x_i \in R^d$ and $y_i \in \{-1, +1\}$, SVM solves the optimization problem:

$$\min_{x,b,\xi_i} \frac{1}{2} ||w||^2 + C \sum_{i=1}^{N} \xi_i \quad (5)$$

subject to: $y_i(\langle \emptyset(x_i), w \rangle + b) - 1 + \xi_i \geq 0, \xi_i \geq 0, i = 1 \dots, N$; where $\emptyset(x_i)$ is a non-linear kernel that maps the training data onto a high-dimensional space. To separate the two classes, SVM works by finding the separating hyperplane that maximizes the margin between observations. The slack variables ξ_i allow misclassification of difficult or noisy patterns. C > 0 is the regularization parameter, which controls the degree of overfitting. Finding the support vectors is made possible using the Lagrange multipliers and the separating hyperplane is found by solving the optimization problem, which

allows the selection of the support vectors that maximize the margin between the two classes (e.g. *truthful* and *deceptive* in this case). In addition, several kernel functions are available to support the transformation of the input data into a higher-dimensional space, where linear separability is possible. Detail about SVM and kernel functions can be found in (Cortes & Vapnik, 1995).

5. EXPERIMENTAL METHODOLOGY

This section describes the methodology of a quantitative empirical study of non-verbal behaviour consisting of between groups experiments. The aim of this work is to establish whether the NVB in terms of micro-movements within the eyes can be used as distinctive clues for deception classification. Furthermore, we investigate comparative analysis of multiple well-known ML algorithms to be trained and tested over the different feature set identified earlier by the PCA and SOM based clustering algorithms (Section 3). Various psychological and computational research studies including (O'Shea et al., 2018), (Marchak, 2013; Nunamaker et al., 2016; Levine, 2014), (Schuetzler, 2012; Kumar, 2016; Pak & Zhou, 2011; Lim, 2013), and (Schuetzler, 2012; Kumar, 2016; Pak & Zhou, 2011; Lim, 2013) have been addressing the importance of eye interactions as sources of indications in non-verbal deception detection. However, there is not a single study to investigate the machine-intelligence based level of significance for the micro level facial and eye movements and to validate the premise using a well-defined controlled and balanced dataset. In order to answer the research question outlined in the introduction, experiments were designed based on a holiday role-playing scenario (described in Section 4.1) using the DT-deception dataset and following experimental setup.

Algorithm 2. Recursive Experimental Steps for Deception Classifiers Training/Testing & Performance Measure

- Let $DT - Deception$ is a feature vector dataset containing all truthful and deceptive feature vectors for 100 participants.
- Let $C = \{SVM, RF, ANN\}$ is set of classifiers used and I_C is a set of input channels shown in Table 1 such that $\forall\, I_C \in R| -1 \leq R \leq 1$.
- Let $O = \{T, D\}$ outputs each classifier in *C* where T and D represents the truthful and deceptive output class O respectively.

Training/Testing of the *C* is performed recursively using following steps.

1: *Set Training Data* = $\{I_{TR} \in I_C \Rightarrow I_{TR} \ltimes I_C\}$ from *DT*

2: *Set Test Data* = $\{I_{TS} \in I_C \Rightarrow I_{TS} \ltimes I_C \,\&\, I_{TS} \notin Training\}$ from *DT*

3: Initialize a classifier from *C* following the corresponding configurations

4: Train classifier until it converges

5: Store the output O from each classifier in C as *[Confusion matrix, Classification]* = $\{O : O \Rightarrow C_i(I_{TR}, I_{TS})\}$

Repeat Steps 1-5 s.t in each iteration (1: 10), I_{TS} belongs to non-repeated unseen pairs of participants (10 truthful, 10 deceptive) from *DT-Deception.*

5.1. Experimental Setup

To investigate the outlined research questions, multiple experiments were conducted using the DT-Deception dataset and Algorithm 2 for the deception classifiers training. For a fair and reliable evaluation of the deception classification performances, a leave-Pair-Out (LPO) strategy is used for the training and testing, which is commonly used strategy in ML (Max et al., 2017). As the extracted dataset contains multiple features vectors/slots (extracted from video data) captured from same subject, a standard approach like cross-validation might result a biased classification performance. For example, a random partition of the train/test samples might contain features extracted from same subject (i.e. videos) that will cause high classification accuracy. We, therefore used the LPO strategy by leaving-out 20 percent of the entire DT-Deception (i.e. feature vectors) extracted from 20% of population (i.e. 10 truthful, 10 deceptive participants) for testing (totally unseen data) and train the deception classifiers with rest of (i.e. 80%) dataset extracted from 80 participants' videos. We performed 10 recursive runs where in each run, the deception classifier is tested on 10 randomly selected non-repeated combination

of pairs (10 truthful, 10 deceptive) of unseen participants' data while trained over the rest of dataset (40 truthful, 40 deceptive). Hence, overall in 10 iterations, the deception classifiers were tested over 100 unseen pairs (100 truthful, 100 deceptive) selected randomly while trained over rest of the dataset in each iteration. The following experiments (labelled A to C) are designed with a consistent deception network configuration and left-out non-repeated truthful and deceptive pairs of participants that were randomly selected.

A) The deception classifiers are trained on the entire feature set(i.e. all 36 features) from the DT-Deception dataset (with 80% of training data split) while recursively (i.e. 10 runs) being tested over the unseen random pairs (i.e. 20% of test data extracted from 10 truthful, 10 deceptive participants).

B) The deception classifiers are trained on the important feature set only (identified through clustering methods) using the DT-Deception dataset (with 80% of training data split) while recursively (i.e. 10 times) tested on randomly chosen unseen pairs (i.e. 20% of test data extracted from 10 truthful, 10 deceptive participants) of similar population used in the experiment A.

C) Using the classifiers' predictions in experiment B, identification of the best compromise between sensitivity and specificity while varying the classifier's decision stump.

Experiments A) and B) were used to test the following hypothesis:

$H_{AB}0$: *There is no significant difference between the performances of different classifiers trained over the entire feature space from DT-Deception vs trained over the important features only.*

$H_{AB}1$: *There is a significant difference between the performance of different classifiers trained over the entire feature space vs trained over the important features only.*

For each experiment (A-C), various statistical metrics (i.e. sensitivity, specificity, accuracy, positive prediction rate, negative prediction rate) are used to evaluate the classification performance based on confusion matrices retrieved from the deception classifiers containing:

True Positive (TP): *Actual deceptive cases are correctly classified as deceptive;* **False Positive (FP):** *Actual truthful cases are incorrectly classified as deceptive;* **True Negative (TN):** *Actual truthful cases are correctly classified as truthful;* **False Negative (FN):** *Actual deceptive are incorrectly classified as truthful.* The *positive* and *negative* conditions represent the *deceptive* and *truthful* classes respectively.

To set the baseline for aforementioned experiments, a number of classification trials were conducted to compare the deception classification performances of ANN, RF and SVM algorithms to choose the parametric configurations and models' tuning. Firstly, multiple random train/test trials were run by partitioning the entire dataset into training and testing proportions of 80% and 20% respectively. It was ensured that the test data contains fair distribution of both truthful and deceptive classes. The final parametric configurations were set empirically based on several recursive trials.

6. RESULTS AND DISCUSSIONS

Table 3 summarizes the statistical results achieved for experiments (A, B) where deception classifiers were trained over the entire feature space (Table 2) as well as important features only, using the same training and testing data proportions. Overall accuracies are shown across both truthful and deceptive scenarios. The models were recursively (10 runs) trained over the DT-Deception dataset which represents the extracted feature vectors from 80 participants (40 Truthful, 40 Deceptive) while leaving 20 randomly selected subjects (10 truthful, 10 deceptive) out for testing in each run.

TABLE 3. DECEPTION CLASSIFIERS' OVERALL PERFORMANCES (TRUTHFUL AND DECEPTIVE) WITH RECURSIVE TRAIN/TEST RUNS WITH LEAVING RANDOM UNSEEN PAIRS OUT FOR TESTING

Train/Test runs (*r*) with 10T,10D random pairs left ou	**Full Feature-set Vs Important Features**	**Accuracy**		
		SVM	**ANN**	**RF**
r_1	Important features	0.75	0.70	0.74
	Full feature-set	0.76	0.68	0.74
r_2	Important features	0.79	0.73	0.81
	Full feature-set	0.79	0.76	0.80
r_3	Important features	0.75	0.72	0.81
	Full feature-set	0.75	0.71	0.77
r_4	Important features	0.73	0.71	0.74
	Full feature-set	0.72	0.68	0.72
r_5	Important features	0.78	0.68	0.80
	Full feature-set	0.78	0.68	0.80
r_6	Important features	0.74	0.73	0.75
	Full feature-set	0.74	0.71	0.75
r_7	Important features	0.76	0.72	0.77
	Full feature-set	0.76	0.74	0.76
r_8	Important features	0.81	0.73	0.83
	Full feature-set	0.81	0.71	0.83
r_9	Important features	0.79	0.76	0.84
	Full feature-set	0.81	0.79	0.85
r_{10}	Important features	0.76	0.69	0.73
	Full feature-set	0.77	0.74	0.75
Avg. Accuracy	**Important Features**	**0.77**	**0.72**	**0.78**
	Full Feature-set	**0.77**	**0.72**	**0.77**

The average accuracy on all runs (r_1 to r_{10}) when using the entire feature space (36 features) was 77%, 72% and 77% for SVM, ANN and RF respectively. This is identical to the corresponding classifiers accuracies when using the important features only (24 features), except in the case of the RF classifier which indicated a slightly higher accuracy (i.e.78% in this case). The maximum classification accuracy is recorded 85% (Full feature-set) and 84% (important features only) in r_9 using RF. The average combined (i.e. truthful and deceptive class) accuracy results in Table 3 also show that RF has outperformed both ANN and SVM. As mentioned earlier, RF can be thought of as a meta-learning model utilizing a bagging concept where a combination of decision trees is used. Individual decisions trees may be weak in classification accuracy and typically characterized by poor classification performance, however in aggregate form, these trees offer robust classification and prediction.

We conducted the Welch two-sample t-test for the accuracy distributions from all runs for all features vs important feature to investigate the hypothesis $\mathbf{H_{AB}}$. The test resulted in the p-value of 0.74 with t score of 0.33 at the 95% confidence interval. This clearly accepts the $\mathbf{H_{AB}(0)}$ that there is no significant difference in the classifiers' performances despite the elimination several features identified irrelevant by the PCA and clustering algorithms (Section 4.4). These outcomes align with the existing computational studies (Borza, 2018; Proudfoot et al., 2015; Pak & Zhou, 2011) which have been indicating the eye gaze and eye interactions can provide significant clues for deception detection. On the other hand, except few of the psychological studies such as (Marchak, 2013; Fukuda, 2001; Dionisio et al., 2001), existing research in general indicate the visual clues as weak identifiers. For instance, the meta-analysis (Bond & DePaulo, 2006; DePaulo et al., 2003) clearly indicating that video medium is less significant than audio and vocal medium in distinguishing deceptive and truthful behaviours. A study conducted over crime professional investigators (Vrij, 2008; Taylor & Hick, 2007) reported that a large proportion of candidates use the eye related NVB to catch the deceivers however, it does not validate the effectiveness of such clues. Furthermore, it is important to note that the natures of the studies in the meta-analyses tend to focus on the differences between summary statistics across experimental groups of human judges, rather than the classifications of individuals by machine in the present study that models the composite of multiple facial micro-movements (not individual nonverbal indicator) using machine intelligence.

TABLE 4. OVERALL DECEPTION CLASSIFIERS' PERFORMANCES WITH RECURSIVE TRAIN/TEST RUNS WITH LEAVING RANDOM UNSEEN PAIRS OUT FOR TESTING

Classifiers	Features for Train/Test	Sen %	Spec %	F1-Score %	Acc %
SVM	Full Features	0.70	0.84	0.78	0.77
RF		0.72	0.83	0.79	0.77
ANN		0.69	0.75	0.75	0.72
SVM	Important Features	0.69	0.84	0.78	0.77
RF		**0.72**	**0.84**	**0.80**	**0.78**
ANN		0.70	0.74	0.74	0.72

Table 4 summarizes various statistical metrics retrieved across all the runs (r_1 to r_{10}) while using different deception classifiers. It can be observed that sensitivity is relatively lower than specificity (in both scenarios; full features vs important features), given the almost balanced dataset for both truthful and deceptive groups. The best compromise between the sensitivity and specificity is produced by the ANN classifier however, it sacrifices the overall accuracy to some extent (i.e. 72% compared to 78% from RF and 77% from SVM). On average, 72% sensitivity and 84% specificity produced by the RF classifier indicates that 28% of the cases were identified as *false negatives* whereas 16% of the cases were identified as *false positives* across the entire experiments (presented in Table 3). This indicate that the deception classifier predicts relatively better the truthful class (i.e. negative class) as compared to the deceptive one (i.e. positive class, PPV). However, the biasness towards truthful class is expected due to the nature of the deception detection problem. On the subject of bias, it was observed in a psychological study by Mann (Mann et al., 2012, 2013), that the embedded truth in NVB emitted from deceivers, could make the classification accuracy more biased towards truthful behavior. Similarly, the experimental outcome from Bond & DePaulo (2006) also indicated biasness towards identification of truthful behavior as compared to deception detection.

TABLE 5. IDENTIFICATION OF OPTIMAL COMPROMISE BETWEEN SEN, SPEC, NPV AND PPV PRODUCED BY RF ALGORITHM DURING A RANDOMLY CHOSEN RUN (r) FOR BOTH GROUPS (TRUTHFUL AND DECEPTIVE)

Cutoff Threshold	Sen %	Spec %	Acc %
0.95	0.52	0.97	0.74
0.9	0.53	0.92	0.73
0.85	0.54	0.89	0.71
0.8	0.55	0.86	0.70
0.75	0.56	0.84	0.70
0.7	0.58	0.83	0.71
0.65	0.61	0.83	0.72
0.6	0.63	0.82	0.72
0.55	0.66	0.82	0.74
0.5	**0.70**	**0.81**	**0.75**
0.45	0.73	0.81	0.77
0.4	0.77	0.80	0.79
0.35	**0.80**	**0.80**	**0.80**
0.3	0.84	0.79	0.81
0.25	0.88	0.79	0.83
0.2	0.91	0.79	0.85
0.15	0.95	0.79	0.88
0.1	0.98	0.79	0.89
0.05	0.99	0.78	0.89

To further investigate the classification accuracy bias towards truthful behavior, we conducted experiment C (Section 5.1) to identify the best compromise between sensitivity and specificity (and hence the *false positive* and *false negatives*) while varying the classifier's decision boundary (we call it *cutoff* value) between 0.05 to 1. These cutoff values represent the RF class prediction probabilities for deception (i.e. positive class in this case). Selection of a decision boundary to distinguish truthful and deceptive behavior has also been selected in previous related studies based on empirical experimentation. For instance, Borza (2018) recently used average blink rate per question to make decisions about whether the question is truthful or deceptive? However, selection of a decision boundary in terms of blink rate per question (to classify the question as truthful or deceptive), and/or number of deceptive questions per interview (to classify the entire interview as truthful or deceptive) don't reveal discrete level information in real time scenarios. More discrete level decision boundary analysis might be helpful to investigate the compromise between false positives/negatives (as shown in Table 5) at vector level (see section 4.2 for channel vector).

Table 5 indicates the best compromise with a 0.35 (i.e. 35%) decision threshold. This implies that a feature vector (i.e. slot) representing the facial micro-movements will be classified as deceptive if the classification probability crosses the threshold value of 0.35. In other words, if the classifier flags a test case as deceptive at the deceptive probability of 0.35, it will result an equal number of false positives and false negatives (i.e. 20%) across the test cases. This threshold might be helpful as a generic cutoff point in either scenario (i.e. frame level, question level, interview level) and may be useful for the human based judgments.

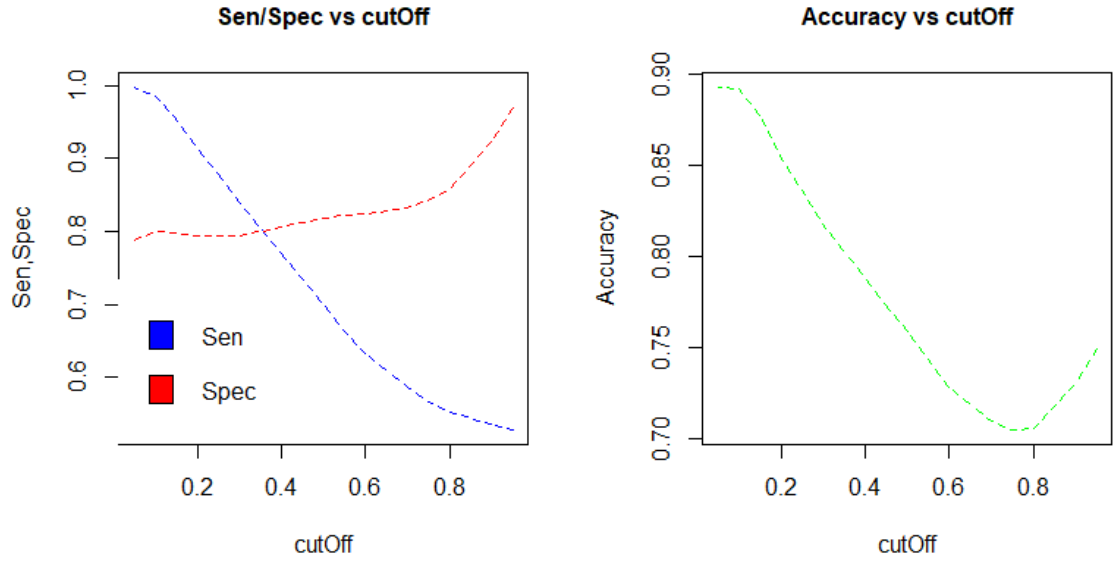


Fig. 5. Impact of varying deception decision boundary on the performance of RF classifier

Fig. 5 demonstrate the variations in sensitivity and specificity with varying cutoff values for RF class prediction probabilities. A random run (r) is chosen from Table 3 while leaving 10 random pairs (10 truthful, 10 deceptive) out for testing while training the RF classifier over the rest of dataset (40 truthful, 40 deceptive). It can be visualized that the accuracy graph varies w.r.t the cutoff thresholds while the sensitivity and specificity intersection around the 0.35 cutoff which indicates the best compromise between false positive, false negative and overall accuracy.

The aforementioned clustering outcomes and deception classifiers performance (with important features only) indicated the importance of eye-related micro-movements for distinguishing the truthful and deceptive behaviours. However, it would be an interesting aspect to perform a comparative analysis between the significance level (i.e. ranking) to assess the importance of the non-verbal channels within the individual groups (i.e. truthful and deceptive subjects). As described earlier (Section 4.4), PCA based attribute rotations are used to identify the top 20 non-verbal features in DT-Deception as shown in Table 6.

TABLE 6. IMPORTANCE SCORE FOR FACIAL AND EYE MICRO-MOVEMENTS WITHIN TRUTHFUL AND DECEPTIVE GROUPS

Feature Rank	Truthful Subjects		Deceptive Subjects	
	Features	**Importance Score**	**Features**	**Importance Score**
1	*lshift*	4.3	*lleft*	4.26
2	*rshift*	4.26	*lblink*	4.21
3	*rblink*	4.11	*rleft*	4.16
4	*lhleft*	4.09	*lright*	4.1
5	*rhleft*	4.08	*rhleft*	4.04
6	*lblink*	4.07	*lhleft*	3.94
7	*lleft*	3.95	*lhright*	3.94
8	*rhclosed*	3.91	*rright*	3.93
9	*lhclosed*	3.85	*lhclosed*	3.89
10	*rright*	3.8	*rshift*	3.85
11	*rhright*	3.63	*rhright*	3.82
12	*lright*	3.61	*lshift*	3.74
13	*lclosed*	3.58	*rblink*	3.73
14	*fblu*	3.57	*fmuol*	3.72
15	*rleft*	3.52	*rhclosed*	3.71

The outcomes indicate two aspects of feature importance. Firstly, it is clear that the top 15 most important NVB are related to eye movements (except fmuol and fblu) as compared to facial micro-movements which aligns with various related computational studies (Borza, 2018; Proudfoot et al., 2015; Pak & Zhou, 2011). In contrast, most of the psychological research studies disagree with this argument. For instance, Bond & DePaulo (2006) identified that audible information is more helpful for human judgments as compared to visual clues. However, these studies tend to focus on the differences between summary statistics across experimental groups of human judges, rather than the classification of individuals by machine intelligence in the present study using multiple facial micro-movements and not individual nonverbal indicator. Secondly, the top 15 features overlap in both groups (i.e. truthful, deceptive) however the ranking (i.e. importance) order is different in both groups. This finding may be helpful for human experts and professional investigators in the field. More specifically, the blinking eye feature (e.g. lblink, rblink, lhcloed, rhclosed) and eye movements (e.g. lleft, rleft, lright, rright) clearly overlap with some of the earlier described technological research findings.

7. CONCLUSIONS AND FUTURE DIRECTIONS

This paper has sort to investigate which specific nonverbal behaviours within the facial and eye micro-movements contains sufficient and dominant clues to distinguish the deceptive and truthful behaviours using intelligent computational models. Furthermore, the study performs detailed comparative analysis using multiple well-known clustering and classification algorithms to validate the research outcomes. A series of experiments has produced several deception classifiers that have been compared using various statistical metrics including accuracy, sensitivity and specificity. The most successful classifier (i.e. RF) achieved overall classification accuracy of 78% in 10 recursive runs while trained over 80 subjects (40 pairs) and tested over unseen random selection of 20 subjects (10 pairs). It also should be noted that the dataset is mixed in terms of gender and ethnicity, and that the coverage across these two factors is quite balanced in the dataset.

The aim of this study was not to investigate whether there was a difference in the NVB cues of participants of different gender and ethnicity from the perspective of automated deception detection, but to investigate which specific fine-grained eye and facial micro-movements contained distinguishing clues for the classification of deception in a general population. An initial study presented by (Crokett et al., 2020), investigated whether there was a difference between the non-verbal cues to deception generated by males and females. The evidence suggested that NVB cues are very similar between males and females but show some differences. The exploratory results indicated that there was a gender effect in that both genders appeared to be at disadvantaged when treated with a combined gender classifier than when a specific classifier tailored to each gender was used. However, as acknowledge in the paper, the results are not conclusive as a large sample size is required. The authors clearly understand and acknowledge the potential risks in using any kind of human-in-the-loop automated system that utilises

machine learning and the need for transparent decision making to avoid discrimination. Current work by the European Commission is looking at policy options to create an 'ecosystem of trust' through creation of regulatory framework for AI using a risk-based approach (EU Commission white paper, 2020). The White Paper (circulated for consultation in February 2020) envisaged "*Requirements to take reasonable measures aimed at ensuring that such subsequent use of AI systems does not lead to outcomes entailing prohibited discrimination. These requirements could entail in particular obligations to use data sets that are sufficiently representative, especially to ensure that all relevant dimensions of gender, ethnicity and other possible grounds of prohibited discrimination are appropriately reflected in those data sets;*". Until such regulatory frameworks are in place, it is the moral and ethical responsibility of researchers and those who apply such research to understand the implications of non-representative dataset on the research question they are addressing.

The results indicated that the most dominant features to distinguish NVB in truthful and deceptive subjects are related to eye micro movements which interestingly aligns with the several technological findings that focus on individual NVB. The study also indicated that the automated deception detection accuracy can be achieved by varying the classification decision boundary that might be helpful for the experts while making the decisions about the deceptive behaviour. The research outcomes also demonstrate that elimination of irrelevant NVB does not reduce the deception detection accuracy which also gives credence to the importance of eyes interactions as distinguishing clues for the automated deception detection. Furthermore, the feature ranking order within the truthful and deceptive subjects, supports some of the existing psychological and more specifically, computational studies which have reported the eye blink rate and eye gaze as significant clues for the deception detection. However, it is important to note that this study is based on machine modeling of multi-dimensional micro-movements and not the individual NVB. Further technical research will be carried out to seek to improve the classification accuracy using a multi-model approach for the DT-detection dataset and to investigate the psychological impact of simulated avatar emotions on the NVB of interviewee during the interview.

8. ACKNOWLEDGMENT

This work was supported by the funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 700626 as well as CfACS (Centre for Advanced Computer Science), Manchester Metropolitan University. The authors also wish to thank Silent Taker Ltd., for the cooperation in experimental design.

REFERENCES

Abouelenien, M., Pérez-Rosas, V., Mihalcea, R., & Burzo, M. (2017). Detecting deceptive behavior via integration of discriminative features from multiple modalities. *IEEE Transactions on Information Forensics and Security*, 12(05), 1042-1055, doi: 10.1109/TIFS.2016.2639344.

Abouelenien, M., Rosas, V. P., Mihalcea, R., & Burzo, M. (2014). Deception detection using a multimodal approach. 16th International Conference on Multimodal Interaction (ICMI '14). ACM, New York, NY, USA, 58-65, doi: https://doi.org/10.1145/2663204.2663229.

Aristoklis, D. A., George, D. M., & Michael, N. V. (2005). New globally convergent training scheme based on the resilient propagation algorithm," *Neurocomputing, Elsevier*, 64, 253-270, doi: https://doi.org/10.1016/j.neucom.2004.11.016.

Avlidis, I., & Levine, J. (2002). Thermal image analysis for polygraph testing. *IEEE Engineering in Medicine and Biology Magazine*, 21(06), 56-64.

Axe, L., Dougherty, D., & Cross, T. (1985). The validity of polygraph testing: Scientific analysis and public controversy, *American Psychologist*, 40(03), 355-366, doi: http://dx.doi.org/10.1037/0003-066X.40.3.355

Bashar, A., & Reyer, Z. (2014). Thermal Facial Analysis for Deception Detection. *IEEE Transactions on Information Forensics and Security*. 09(06), 1015-1023, doi: 10.1109/TIFS.2014.2317309.

Bond, C. F. J., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review*, 10(03), 214–234, doi: https://doi.org/10.1207/s15327957pspr1003_2.

Borza, D., Itu, R., & Danescu, R. (2018). In the Eye of the Deceiver: Analyzing Eye Movements as a Cue to Deception. *Journal of Imaging, MDPI*, 4(10), 1-20, doi: https://doi.org/10.3390/jimaging4100120.

Bradski, G. (2000). OpenCV Library. OpenCV.org. Retrieved from https://docs.opencv.org/master/d2/d42/tutorial_face_landmark_detection_in_an_image.html

Breiman, L. (2001). Random forests. *Machine learning*, 45(01), 5-32, doi: https://doi.org/10.1023/A:1010933404324.

Buckingham, F., Crockett, K., Bandar, Z., O'Shea, J. (December 2014). FATHOM: A Neural Network-based Non-verbal Human Comprehension Detection System for Learning Environments. IEEE SSCI, Florida, 403-409, doi:10.1109/CIDM.2014.7008696.

Chen, J., Chen, Z., Chi, Z., & Fu, H. (2018). Facial Expression Recognition in Video with Multiple Feature Fusion. *IEEE Transactions on Affective Computing*, 09(01), 38-50, doi: 10.1109/TAFFC.2016.2593719.
Cortes. C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(03), 273-297, doi: https://doi.org/10.1007/BF00994018.
Crockett, K. A., et al. (2017). Do Europe's borders need multi-faceted biometric protection, Biometric Technology Today. 07, 5-8, ISSN: 0969-4765.
Crockett, K., O'Shea. J., Khan, W. (2020). Automated Deception Detection of Male and Females from Non-Verbal Facial Micro-Gestures, *IEEE World Congress in Computational Intelligence – IEEE IJCNN*, July 2020, (accepted).
DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, *129*(1), 74–118. https://doi.org/10.1037/0033-2909.129.1.74
Dionisio, D. P., Granholm, E., Hillix, W. A., & Perrine, W. F. (2001). Differentiation of deception using pupillary responses as an index of cognitive processing. *Psychophysiology*, 38, 205-211.
Ekman, P. (2001). Telling Lies: Clues to Deceit in the Marketplace, Politics and Marriage. Norton, W.W. and Company, ISBN: 0393321886, 9780393321883 https://books.google.co.uk/books?id=7I_wDDfrwCgC.
Ekman, P. (2003). Darwin, deception, and facial expression. *Annals of the New York Academy of Sciences*, 1000(01), 205–221, 2003, doi: 10.1196/annals.1280.010.
Ekman. A., et al. (1991). Invited article: Face, voice, and body in detecting deceit. *Journal of nonverbal behavior*, 15(02), 125–135.
Eriksson, A., & Lacerda, F.)2007). Charlatanry in forensic speech science: A problem to be taken seriously. *International Journal of Speech, Language and the Law*, 14(02), 169-193.
European Commission White Paper (2020). On Artificial Intelligence - A European approach to excellence and trust. Online available: https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en. Accessed on: 24th May 2020.
Feldman, R. S., Jenkins, L., & Popoola, O. (19790. Detection of deception in adults and children via facial expressions. *Child development*, *Wiley*, 50(02), 350–355, doi: 10.2307/1129409.
Freire, A., Eskritt, M., & Lee, K. (2004). Are eyes windows to a deceiver's soul? children's use of another's eye gaze cues in a deceptive situation. *Developmental psychology*, 40(06), 1093-1104, doi: 10.1037/0012-1649.40.6.1093.
Frischen, A., Bayliss, A. P., & Tipper, S. (2007). Gaze cueing of attention: Visual attention, social cognition and individual differences. *Psychological Bulletin*, 133, 694–724. doi:10.1037/0033-2909.133.4.694.
Frosina, P., Logue, M., Book, A., Huizinga, T., Amos, S., & Stark, S. (2018). The effect of cognitive load on nonverbal behavior in the cognitive interview for suspects, *Personality and Individual Differences*, 130, 51-58, doi: https://doi.org/10.1016/j.paid.2018.03.012.
Fukuda, K. (2001). Eye blinks: new indices for the detection of deception. *International Journal of Psychophysiology*, 40, 239-245.
Happy, S. L., & Routray, A. (2015). Automatic facial expression recognition using features of salient facial patches. *IEEE Transactions on Affective Computing*, 06(01), 1-12, doi: 10.1109/TAFFC.2014.2386334.
Hervé, A., & Williams, L. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 02(04), 433-459,doi: 10.1002/wics.101.
Hirschberg, J., et al. (2005). *Distinguishing deceptive from non-deceptive speech*. Interspeech, Eurospeech, 1833-1836.
Ho, T. H. (2002). A data complexity analysis of comparative advantages of decision forest constructors. *Pattern Analysis & Applications*, 05(02), 102-112.
Howard. D., & Kirchhubel, C. (2011). Acoustic correlates of deceptive speech: an exploratory study. 9th international conference on engineering psychology and cognitive ergonomics, EPCE'11, Berlin, Heidelberg, 28-37.
Jones, M., & Viola, P. (2003). Fast multi-view face detection, Technical Report MERLTR2003-96, Mitsubishi Electric Research Laboratories.
Kleinberg, B., Mozes, Arntz, A., Verschuere, B. (2018). Using Named Entities for Computer-Automated Verbal Deception Detection. *Journal of forensic sciences*, 63(03), 714-723, doi: 10.1111/1556-4029.13645.
Kozel, F. A., et al. (2005). Detecting deception using functional magnetic resonance imaging. *Biological Psychiatry*, 58(08), 605–613.
Kumar, L. P. (2016). *Using eye-tracking to understand user behavior in deception detection system interaction*. Masters Theses, Missouri University of Science and Technology, Retrieved from https://scholarsmine.mst.edu/masters_theses/7605.
Larson, J. A., Haney, G. W., & Keeler, L. (1932). Lying and its detection: A study of deception and deception tests (p. 99). Chicago, IL: University of Chicago Press.
Levine, T. R. (2014). Active Deception Detection. *Policy Insights from the Behavioral and Brain Sciences.* 01(01), 122-128, doi: 10.1177/2372732214548863.
Li, S. Z., et al. (2002). Statistical learning of multi-view face detection. A. Heyden et al., (eds) Computer Vision-ECCV, Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2353, 67-81, doi: https://doi.org/10.1007/3-540-47979-1_5.
Lim, K. K., Friedrich, M., Radun, J., & Jokinen, K. (December 2013). Lying through the eyes: detecting lies through eye movements. 6th workshop on Eye gaze in intelligent human machine interaction: gaze in multimodal interaction, Sydney, Australia, 51-56, doi: 10.1145/2535948.2535954.
Lloyd, E. P., et al. (Feb, 2019) Miami University deception detection database. *Behavior Research Method*, 51(01), 429-439, doi: https://doi.org/10.3758/s13428-018-1061-4.
Mann, S., Ewens., S., Shaw, D., Vrij, A., Leal, S., & Hillman, J. (2013) Lying Eyes: Why Liars Seek Deliberate Eye Contact, *Psychiatry, Psychology and Law*, 20(03), 452-461, doi: 10.1080/13218719.2013.791218.
Mann, S., Vrij, A., & Bull, R. (2004). Detecting true lies: Police officers' ability to detect deceit. *Journal of Applied*

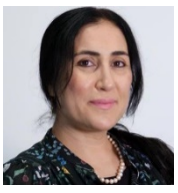
Psychology, 89, 137–149. doi:10.1037/0021-9010.89.1.137.
Mann, S., Vrij, A., Leal, S., Granhag, P., Warmelink, L., & Forrester, D. (2012). Windows to the soul? Deliberate eye contact as a cue to deceit. *Journal of Nonverbal Behavior*, 36, 205-215, doi: 10.1007/s10919-012-0132-y
Mann, S., Vrij, A., Shaw, D.J., Leal, S., Ewens, S., Hillman, J., Granhag, P.A. & Fisher, R.P. (2013), Two heads are better than one? How to effectively use two interviewers to elicit cues to deception. *Legal and Criminological Psychology*, 18, 324-340. doi:10.1111/j.2044-8333.2012.02055.x.
Marchak, F. M. (2013). Detecting false intent using eye blink measures. *Frontiers in psychology*, 04(736), doi:10.3389/fpsyg.2013.00736.
Max, A. L., Gael, V., Sohrab, S., Luca, L., Arun, J., David, C., Konrad, P. K. (2017). Using and understanding cross-validation strategies. Perspectives on Saeb et al., *GigaScience*, 06(05), doi: https://doi.org/10.1093/gigascience/gix020
Mendels, G., et al. (2017). *Hybrid Acoustic-Lexical Deep Learning Approach for Deception Detection.* Interspeech, 1472-1476, doi:10.21437/Interspeech.2017-1723.
Merla, A., & Romani, G. (2007). Thermal signatures of emotional arousal: A functional infrared imaging study. 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, (EMBS), 247-249.
Meservy, T. O., et al. (2005). Deception detection through automatic, unobtrusive analysis of nonverbal behavior. *Intelligent Systems, IEEE*, 20(05), 36–43.
Mihalcea, R., & Strapparava, C. (2009). The lie detector: Explorations in the automatic recognition of deceptive language. Association for Computational Linguistics, Suntec, Singapore, 309-312.
Nunamaker, J. F., Burgoon, K. J., & Giboney, J. S. (2016). Information Systems for Deception Detection. *Journal of Management Information Systems.* 33(02), 327-331, doi: 10.1080/07421222.2016.1205928.
O'Shea, J., Crockett, K., Wasiq, K., Kindynis, P., Antoniades, A., & Boultadakis, G. (2018). Intelligent Deception Detection through Machine Based Interviewing. IEEE International Joint conference on Artificial Neural Networks (IJCNN), Rio de Janeiro, 1-8, doi: 10.1109/IJCNN.2018.8489392.
Owayjan, M., et al. (December, 2012). The design and development of a lie detection system using facial micro-expressions. 2nd International Conference on Advances in Computational Tools for Engineering Applications (ACTEA), 33–38.
Pak, J., & Zhou, L. (August 2011). Eye Movements as Deception Indicators in Online Video Chatting. 17th Americas Conference on Information Systems, (AMCIS), Detroit, Michigan, USA, 1-9.
Pavlidis, I., & Levine, J. (2001). Monitoring of periorbital blood ow rate through thermal image analysis and its application to polygraph testing. 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 03, 2826-2829.
Pavlidis, I., Eberhardt, N. L., & Levine, J. A. (2002). Human behavior: Seeing through the face of deception. Thermal imaging offers a promising hands-off approach to mass security screening. *Nature, International Journal of Science*, 415(35), doi: https://doi.org/10.1038/415035a.
Pfister, T., & Pietikäinen, M. (January 2012). Electronic imaging & signal processing automatic identification of facial clues to lies. *SPIE Newsroom* http://spie.org/news/4095-automatic-identification-of-facial-clues-to-lies?SSO=1.
Pons, G., & Masip, D. (2018). Supervised Committee of Convolutional Neural Networks in Automated Facial Expression Analysis. *IEEE Transactions on Affective Computing*, 09(03), 343-350, doi: 10.1109/TAFFC.2017.2753235.
Proudfoot, J. G., Jenkins, J. L., Burgoon, J. K., & Nunamaker, J. F. (2015). Deception is in the eye of the communicator: Investigating pupil diameter variations in automated deception detection interviews. IEEE International Conference on Intelligence and Security Informatics (ISI), Baltimore, MD, 97-102. doi: 10.1109/ISI.2015.7165946.
Rosas, V., et al. (2015). Deception Detection using Real-life Trial Data. International Conference on Multimodal Interaction (ICMI '15), New York, NY, USA, 59-66, doi: https://doi.org/10.1145/2818346.2820758.
Roshni, D. T., & Bhavana, B. H. (2014). REVIEW: Previous Deception detection methods and New proposed method using independent component analysis of EEG signals. *Int. Journal of Engineering Research and Applications*, 04(12), 178-182.
Rothwell, J., Bandar, Z., O'Shea, J. and McLean, D. (2007). Charting the behavioural state of a person using a backpropagation neural network. *Neural Computing and Applications*, 16(4-5), 327-339.
Rothwell, J., Bandar, Z., O'shea, J., & Mclean, D. (2006). Silent talker: a new computer-based system for the analysis of facial cues to deception. *Applied cognitive psychology*, 20, 757-777.
Schuetzler, R. M. (2012). Countermeasures and Eye Tracking Deception Detection," in Faculty Proceedings & Presentations, University of Nebraska Omaha, https://digitalcommons.unomaha.edu/isqafacproc/28.
Sebanz, N., & Shiffrar, M. (2009). Detecting deception in a bluffing body: The role of expertise. *Psychonomic bulletin & review*, 16(01), 170–175.
Siering, M., Koch, J., & Deokar, A. V. (2016). Detecting fraudulent behavior on crowdfunding platforms: The role of linguistic and content-based cues in static and dynamic contexts. *Journal of Management Information Systems*, 33, 421-455.
Sporer, S. L., & Schwandt, B. (2007). Moderators of nonverbal indicators of deception: A meta-analytic synthesis. Psychology, Public Policy, and Law, 13(1), 1–34. https://doi.org/10.1037/1076-8971.13.1.1
Taylor, R., & Hick, R. F. (2007). Believed cues to deception: Judgements in self-generated serious and trivial situations. *Legal and Criminological Psychology*, 12, 321–332. doi:10.1348/135532506X116101.
Tian, Y., Kanade, T., & Cohn, J. (2005). *Facial expression analysis*. Handbook of Face Recognition, Springer New York, 247–275.
Viola, P., Jones, & M. J (2001). Rapid object detection using a boosted cascade of simple features. Computer Vision and Pattern Recognition, (CVPR '01), Los Alamitos, CA, USA, pp. 511-518.
Vrij, A. (2008). *Detecting Lies and Deceit: The Psychology of Lying and the Implications for Professional Practice*. (2nd Ed). Chichester: Wiley.
Vrij, A., et al. (2009). Outsmarting the liars: The benefit of asking unanticipated questions. *Law and Human Behavior*, 33(02), 159-166.

Vrij, A., Granhag, P. A., Mann, S., & Leal, S. (2011). Outsmarting the liars: Toward a cognitive lie detection approach. *Current Directions in Psychological Science*, 20, 28-32.
Vrij, A., Mann, S., Fisher, R., Leal, S., Milne, R., & Bull., R. (2008). Increasing cognitive load to facilitate lie detection: The benefit of recalling an event in reverse order, *Law and Human Behavior*, 32, 253-265, doi: 10.1007/s10979-007-9103-y.
Vrij, A., Mann, S., Leal, S., & Fisher,R. (2010). Look into my eyes: Can an instruction to maintain eye contact facilitate lie detection? *Psychology, Crime & Law*, 16, 327-348.
Walczyk. J., et al. (2013). 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*, 04(14), doi: 10.3389/fpsyg.2013.00014.
Zimmerman, L. (2016). Feature Deception Detection. *American psychological Association*, 47(03), http://www.apa.org/monitor/2016/03/deception.aspx.

Dr. Wasiq Khan is a Senior Lecturer in Artificial Intelligence (AI) & Data Sciences within the Department of Computer Science at Liverpool John Moores University, UK. Wasiq received his B.Sc. in Mathematics, Physics and M.Sc. in Computer Science from Pakistan. He further received an MSc in AI followed by a Ph.D. in AI & Speech Processing from Bradford University in 2015, UK. Wasiq is research active within the domain of AI, Deep/Machine learning, Speech processing, and Video/Image analysis. He has been working as a lead researcher/Co-Investigator on various large-scale research projects in collaborations with academia & industry. He has been publishing research outcomes in high impact Journals and conferences. He is an active reviewer of various top ranked Journals (including IEEE transactions and IEEE Access), UKRI/EPSRC research grants and chairing conference sessions. Along with the teaching roles and Ph.D supervisions, Wasiq has established academic citizenship within the domain of AI and Data Science while he is also Fellow of Higher Education Academy, UK and Member of IEEE, Computational Intelligence Society.

Dr. Keeley Crockett is a Reader in Computational Intelligence in the School of Computing, at Manchester Metropolitan University in the UK. She gained a BSc Degree (Hons) in Computation from UMIST in 1993, and a PhD in the field of machine learning from the Manchester Metropolitan University in 1998. She is A Senior Fellow of the Higher Education Academy. She leads the Computational Intelligence Lab that has established a strong international presence in its research into Adaptive Psychological Profiling including an international patent on "Silent Talker". She is currently a member of the IEEE Task Force on Ethical and Social Implications of Computational Intelligence. She has 22 PHD completions and externally examined 8 PhDs. Her main research interests include fuzzy decision trees, semantic text based clustering, conversational agents, fuzzy natural language processing, semantic similarity measures, and AI for psychological profiling. Currently the Principal Investigator (MMU) on the H2020 funded project iBorderCtrl: Intelligent Smart Border Control, CI on H2020 Grant "Populism and Civic Engagement: a fine-grained, dynamic, forward-looking response to the negative impacts of populist movements (PaCE)", and CI on UK KTP with Service Power. She is a member of the IEEE WIE Leadership committee, Chair of the IEEE CIS Webinars, Vice-Chair of the IEEE Women into Computational Intelligence. She is Student Activities co-chair for IEEE WCCI 2020. She has authored over 120 peer reviewed publications.

Dr. O'Shea holds concurrent affiliations as a Senior Lecturer in Computer Science at Manchester Metropolitan University. He offers 18 years of experience to his role as co-founder and consultant for the Silent Talker team. Dr. O'Shea helped develop, test, and patent the proprietary Silent Talker software. He has earned degrees in Chemistry and Artificial Intelligence. He is Editorial Board Member for the International Journal of Intelligent Defense Support Systems. He organized and chaired international conferences on Agent and Multi-Agent Systems. With more than 60 publications in international scientific Journals, book chapters and peer-reviewed conferences, Dr. O'Shea is a leading expert in adaptive psychological profiling, dialogue systems and computational intelligence. He is also an expert in applying these systems both in English and with resource-poor languages such as Arabic, Urdu, and Thai. Dr. O'Shea has led or contributed to projects funded by Horizon 2020. He is an enthusiastic science communicator including developing young scientists through Research Placements funded by the Nuffield Foundation. He is currently a co-investigator on the H2020 funded project iBorderCtrl – Intelligent Smart Border Control. He is a member of the IEEE and a lifetime gold member of the Knowledge Engineering Society.



Prof. Abir H. is a professor of Machine Learning and a member of the Applied Computing Research Group at the Faculty of Engineering and Technology. She completed her PhD study at The University of Manchester (UMIST), UK in 2000 with a thesis title Polynomial Neural Networks for Image and Signal Processing. She has published numerous referred research papers in conferences and Journal in the research areas of Neural Networks, Signal Prediction, Telecommunication Fraud Detection and Image Compression. She has worked with higher order and recurrent neural networks and their applications to financial, physical, e-health and image compression techniques. She has developed with her research students a number of recurrent neural network architectures. Her research has been published in a number of high esteemed and high impact journals such as the Expert Systems with Applications, PloS ONE, Electronic Letters, Neuro-computing, and Neural Networks and Applications. She is a PhD supervisor and an external examiner for research degrees including PhD and MPhil. She is one of the initiators and chairs of the Development in e-Systems Engineering (DeSE) series, most notably illustrated by the IEEE technically sponsored DeSE International Conference Series.

Dr. Bilal Khan. is senior researcher at the University of California Los Angeles. He also co-founded and currently serving as the chief analytics officer at Noria Water Technologies, a Los Angeles based technology company. Bilal has master's degrees in computer science (PK) and Pervasive Computing (Birmingham City University, UK), Ph.D. in cooperative vehicular networks using game theory and artificial intelligence (University of Bradford, UK, 2012) and the fellowship of the UK higher education commission. His current areas of expertise are the use of data analytics in the areas of membrane-based water treatment, heat exchangers, nanotechnology and environmental risk assessment. Dr. Bilal was recently nominated as 2019 young professional by the US Water and Wastes Digest for his IP protected work for digital advancements in water treatment and power industries. Bilal has active memberships of various prestigious scientific organizations (American Water Works Association, American Chemical Society, American Institute of Chemical Engineers, US-EU Nanotechnology Roadmap, US Nano Working Group). Dr. Bilal is also an active reviewer of reputed journals (including Nature Scientific Data, Environmental Science and Technology, Nanoscale) and has chaired technical sessions in the fields of Nanotechnology and Water treatment.