Intelligent Techniques for Deception Detection: A Survey and Critical Study

Haya Al-askar[1*], Zohra Sbaï[1,2], Wasiq Khan[3], Abir Hussain[3] and Arwa Alruis[1]

[1]Computer Science Department, College of Engineering and Computer Sciences, Prince Sattam Bin Abdulaziz University, Al-Kharj, 11942, Saudi Arabia.
[2]National Engineering School of Tunis, Tunis El Manar University, Tunis, 1068, Tunisia.
[3]Computer Science Department, Liverpool John Moores University, Liverpool, L33AF, UK.

*Corresponding author(s). E-mail(s): h.alaskar@psau.edu.sa;
Contributing authors: z.sbai@psau.edu.sa; w.khan@ljmu.ac.uk; a.hussain@ljmu.ac.uk; a.alrawais@psau.edu.sa;

Abstract

Machine intelligence methods originated as effective tools for generating learning representations of features directly from the data and have indicated usefulness in the area of deception detection. The success of machine intelligence based methods covers resolving multiple complex tasks that combine multiple low-level image features with high-level contexts, from feature extraction to classification. The goal of this paper, given this period of rapid evolution, is to provide a detailed overview of the recent developments in the domain of automated deception detection mainly brought about by machine intelligence based techniques. This study examines about 100 research papers that explores diverse areas of common deception detection through text, speech, and video data analysis. We performed a critical analysis of the existing techniques, tools and available datasets which have been used within the existing works, followed by possible directions for the future developments in this domain.

Keywords: Deception detection, deep learning, state of the art, verbal and non-verbal cues, datasets and libraries

1 Introduction

Around the world, the number of criminal cases increases every year. It is ethically and morally essential to only condemn guilty defendants and to set the innocent free. As the decision of each trial is focused more on the stakeholder hearings and facts (accused, witnesses, etc.), incorrect judgments could result if the stakeholders do not tell the truth. Thus, it is necessary to reliably detect deceptive activity to preserve law and order.

Additionally, social networking can be described as a simulated environment in which people communicate without a physical sensation and touch. It is easy not to disclose one's name and/or to impersonate another person on social media. Currently, cyberbullying is an extremely frequent problem among teenagers [Krishnamurthy et al. 2018, Smith et al. 2008]. Cyberbullying involves circulating gossip, making threats and sexual harassment. Cyberbullying negatively impacts victims and increases the possibility of negative emotional responses, such as reduced self-esteem and increased thoughts of suicide, anger

and depression, [Hinduja and Patchin 2015, Krishnamurthy et al. 2018].

Another application domain in which the identification of deception is extremely significant is the growing number of false reports on the Internet, including fake news or URL [Zhang and Kou 2022, Lie et al. 2020]. There are numerous real-world contexts, such as video advertising, airport previews, courtroom trials, and work interviews, in which deception detection could play a critical role, and video detection demands are enormous. State-of-the-art deception detection methods have been indicated to be unreliable by early theories and polygraphs [DePaulo and Morris 2004, Zimmerman 2016, Granhag and Stromwall 2004].

However, over a century of research and theorizing has seen physiognomic, and psychological models all attempted to find the best approach to tell if someone is lying. Two major types of signs have been studied in these approaches, including verbal (i.e., linguistic) indicators and nonverbal cues [DePaulo et al. 2003, Levitan et al. 2015]. Nonverbal features are classified into the following three classes: neurological, visual, and vocal indicators for example eye gaze, body movement, facial gestures, posture, etc. [Fitzpatrick et al. 2015]. Verbal features include psycholinguistic features that have been extracted from psycholinguistics lexicon, and syntactic features.

This research work summarizes the available approaches in the domain of deception detection. The main contributions of the proposed research work are:

- Exhaustive study and comparison of the different methods available for deception detection.
- Identification of different features/clues compulsory to classify deception detection techniques.
- Critical analysis of the different approaches and recommendations proposal.
- Report of the existing datasets in terms of their size, diversity, and how they are collected, leading thus to analyze their limitations.

The rest of this paper is organized as follows. Section 2 summarizes the different methods of deception detection. Section 3 classifies the deception detection approaches according to the clues used: verbal, non-verbal, or hybrid. Critical analysis and details of the different existing methods are exposed in section 4. Section 5 reports the datasets and libraries by classifying them in terms of well-chosen categories. Section 6 discusses some challenges faced in the domain of deception detection. Section 7 is reserved to conclude the paper and announce future directions.

2 Deception detection methods

Deception detection methods are classified in psychological, professional, and computational. The rest of this section exposes the details of each method, the state of the art approaches in each of which, as well as their main strengths and limitations.

2.1 Psychological methods

From a psychology perspective, researchers focused on nonverbal and verbal behaviours that an individual perceiver can use to detect deception. Researchers believe that deceit is accompanied with different psychological activity, and that some of the signs may seep through when individuals lie. As a result, individuals frequently attempt to detect deception based on nonverbal indicators like body movement and eye look, while ignoring or paying less attention to verbal signs. [An 2015]

For example, DePaulo and Morris conducted in [DePaulo and Morris 2004] a meta-analysis into the possible predictors of deception. They claimed that detecting deception is an inexact science, and there is a link between lying and an increase in pupil size, which is a sign of tension and concentration. Also, they discovered that people who listen to liars believe they appear more worried than those who tell the truth where their voices have a higher pitch. Liars are also more prone to press their lips together than truth-tellers. They did notice, however, that liars do not appear to be more fidgety, nor do they blink more or have less eyes standing in a casual manner. According to DePaulo and Morris, only when liars are more highly motivated and they are on stakes they seem unusually still and make notably less eye contact with listeners.

There are some psychological meaning units of measurement used to detect deception. Some

of these include touching, rubbing, etc., of various body parts which could be composed of hand, finger, and arm movements, but they were scored for theoretical rather than merely descriptive reasons. Other psychological meaning units of measurement include illustrators, which accompany speech to help keep the rhythm of the speech, emphasize a word, show direction of thought, or emblems which are gestures that have a speech equivalent such as a head nod meaning "yes", or facial emblems such as winking. Vocal tension, speech difficulties, negative remarks, contextual embedding, uncommon details, logical structure, unforeseen complexities, needless details, and self-doubt are examples of psychological meaning units. Other research works looked into behaviour at the most interpretative/impressionistic unit levels, which include fidgeting, involvement, body animation, posture, facial pleasantness, expressiveness, vocal immediacy and involvement, and spoken uncertainty [Adelson 2004].

Today's scholars are experimenting with new deception detection techniques. Rather of searching for visual indicators that someone is lying, such as lack of eye contact or fidgeting, psychologists are instead concentrating on devising proactive techniques that interviewers might use to elicit deceptive signs. So, the new approaches focus on strategic interplay between deceiver and observer. [Hartwig et al. 2014]

2.2 Professional investigations

Conventional police rehearses in trickery location originate from early hypotheses on lying that accept liars will show pressure based signals since they dread being gotten and feel regretful about lying. On the one hand, this hypothesis drove analysts to look for solid social pointers of double dealing. They inspected practices, for example, act shifts, gaze aversion, and foot and hand movements [Bond Jr and DePaulo 2006, Davis and Markus 2006, Lajevardi and Hussain 2012, Ekman 2009].

On the other hand, professionals claim that the accuracy of detecting deceptions can be due to the fact that observers depend more heavily on nonverbal actions when visual information is available and are immediately attracted to stereotypical signs of deception (e.g. gaze avoidance) to direct their judgments of deception [Bond Jr and DePaulo 2006, Davis and Markus 2006]. However, human capacity to distinguish deception without special assistance is limited to approximately 54%, as reported in [Bond Jr and DePaulo 2006].

According to the American psychological association, the human ability to spot lies is no more reliable than chance [Zimmerman 2016]. Therefore, methods to help detect deception are in high demand. In the past, numerous models for the detection of deceit have been suggested. Examples include physiological techniques such as the popular polygraph test [Gale 1988] or the latest magnetic resonance imaging (fMRI) tests [Rusconi and Mitchener-Nissen 2013]. The polygraph was considered as the most popular method used by forensic investigations on particular topics. However, these methods suffer from the following two drawbacks: they require a complex setup of facilities and a skilled operator. The applicability of these methods to real life is therefore fairly low [Turnip et al. 2017]. The popularly used polygraph method has not been objectively proven to be able to detect human deceit [Gupta et al. 2019]. The polygraph, a commonly used tool for detecting lies, is intrusive because it must be attached to the body of the participant during a session in which people know they are being observed and to which they can develop countermeasures [Lajevardi and Hussain 2012].

Physiological approaches are invasive, costly and not very successful. In addition, physiological approaches require skilled operators [Karimi 2018]. Creating an efficient and accurate lie detector can help to avoid and eliminate possible danger and harm. Lie detection based on nonverbal signals is unobtrusive, and it is less likely that countermeasures can be developed by the people being examined to lie detection based on nonverbal signals than to lie detection based on physiological approaches. This system depends on an autonomic nervous system response, which senses the subject's emotional reaction. There are some drawbacks of using polygraphs as lie detectors. A study was published by the national academy of sciences in 2003, posing several concerns regarding the legitimacy of the polygraph and the research behind it. The research was more important with respect to its use for career

screening and predicted potential wrongdoing. [Burgoon et al. 2015]

Studies identified unique movements of the hand that could correspond to the act of deception [Caso et al. 2006, Cohen et al. 2010]. [Hillman et al. 2012] concluded that deceit was correlated with increased speech urging gestures, while truthful behaviour was associated with increased rhythmic pulsing gestures. The taxonomy of hand movements produced for manipulation and social conduct by [Caso et al. 2006] is also related. Other nonverabl features has been addressed by number of studies of facial expression. For example [Ekman 2009] described micro-expressions as relatively short spontaneous expressions, which may be indicative of deceptive actions. In addition, [Ekman 2003] studied these expressions using measures of smoothness and asymmetry to further connect them to an act of deception.

Number of publications on verbal-based deception detection have investigated the detection of deceptive content in a range of contexts, including customer report websites [Li et al. 2014], online dating sites [Guadagno et al. 2012], forums [Warkentin et al. 2010], and social networks applications [Ho and Hollister 2013]. Studies have demonstrated the usefulness of text analysis characteristics, which also include simple linguistic representations such as n-grams and data on sentence counts [Strapparava and Mihalcea 2009]. Several scholars demonstrated that for the automated recognition of deception, the use of psycholinguistic knowledge is useful. They have also researched the association between text syntactic sophistication and deceit, pursuing the theory that deceivers may construct less complex sentences in an attempt to hide the facts and be able to remember their lies more quickly [Perez-Rosas et al. 2015b, Yancheva and Rudzicz 2013].

Study has shown that nonverbal indications are weak and unreliable compared to verbal indications of deception [Bond Jr and DePaulo 2006] A meta-analysis of the ability of observers to detect deception when detecting nonverbal and verbal indications of deception found that they performed badly (52%) when observers can only see liar than when they could only hear liar (63%). The same patterns of results occurred as police officers analysed victims' truths and lies in actual police interviews [Mann et al. 2008].

[Vrij et al. 2018] referred that this relative weakness in nonverbal signs of deception may be related to the techniques employed by truth tellers and liars when actually trying to make a persuasive impact on others. [Levine et al. 2011] claimed that reliance on body language signals (e.g., body expressions, eye gaze, and style of interaction) alone can lead to poor accuracy in distinguishing truths from lies. Both of them use similar nonverbal behaviour strategies in order to mask nervousness signs and try to substitute them with signs that show then more hones, such as looking into their eyes with conversation partners and resisting fidgeting (scratching head, wrists etc.) [Hartwig et al. 2010].

Truth tellers and liars employ different techniques for verbal behaviour. Truth tellers are straightforward and use a strategy of "tell it all," while liars use a strategy of "keep it simple" and avoid discussing incriminating information [Granhag and Hartwig 2008]. This indicates that it promotes segregation between truths and lies through access to verbal information.

Furthermore, the opinions of observers on the authenticity of verbal signs of deception are usually more reliable than those of nonverbal signs [Stromwall et al. 2004]. Hence, As a result, when visual information is ignored or missing, observers' detection accuracy may improve, allowing them to focus on the more diagnostic speech content and/or voice elements of the communication. According to the investigations of [Bradford et al. 2013], the validity of confessions is best established through audio recordings or written text-based remarks, as these appear to be less subject to the biases that might damage credibility evaluations based on visual evidence.

2.3 Computational models

Recent success in automated detecting deception (ADD) deployed the use of data mining and machine learning algorithms [Gogate et al. 2017, Jaiswal et al. 2016, Krishnamurthy et al. 2018]. Innovation in ADD could transform the performance of enforcement investigations by military/public/private/law enforcement agencies. For instance, recent works attempt to use ADD systems in the defense [Jupe and Keatley 2019],

border security [Khan et al. 2021] to recognize deception through machine intelligence. Researchers have proposed new strategies to assist investigative agencies and police in detecting deception [Gogate et al. 2017]. Scientists and researchers have used other techniques to detect deception by observing behaviour, testing sound stress [Al-tahri et al. 2022], detecting facial expressions, evaluating heart rate, measuring skin activity, and evaluating breathing rhythm, among other popular techniques [Fitzpatrick et al. 2015].

Researchers from a number of fields, including psychology, computer science, linguistics, and criminology, have been working to establish advanced tools for reliable deception detectinn [Kleinberg et al. 2019, Gogate et al. 2017, Gupta et al. 2019]. The focus of most current deception identification approaches is on detecting signs of deceit/truthfulness. One advanced deception detection methods uses machine learning algorithms in a number of different ways, such as speech , text analysis and micro-expression videos.

[Wang 2017] shows that a CNN trained on LIAR can classify fraudulent claims with a test-time accuracy of 27.4%. Whereas, [Upadhayay and Behzadan 2020] proposed a new architectures based on Bidirectional Encoder Representations from Transformer with CNN to trained on LIAR and they obtained 70% accuracy.

Recently, due to advances in computational power and resources, deep learning has been implemented in many related fields, such as facial recognition [Ding et al. 2019], signals analysis [Alaskar 2018] and speech recognition [Zhou et al. 2015a]. One of the advantages of deep learning is the automation in feature extraction from raw data as compared to conventional methods [Kou et al. 2022, Xiao et al. 2021a, Xiao et al. 2021b].

[Rahman et al. 2019] offer a new method for leveraging smartphones to detect lying. They provided a unique survey system with a carefully selected collection of questions that encourage participants to give both truth and deceptive answers. 47 people were participated on this experiment. In this experiment, machine learning algorithms were applied to analysis the data and discover that Random Forest can accurately distinguish true and false replies over the collected data with an accuracy of 83%.

3 Cues-based taxonomy of deception detection

This section introduces the different applications to detect deceptions based on different types of data. We propose to present this review in terms of used cues: verbal, non-verbal, and hybrid cues, thus reasoning based on different types of data such as text, video, audio, EEG-P300 signals, as well as multi models.

The most important and recent researches studied in this paper are first analyzed in terms of approach used, publisher, and year of publication as presented in figure 1. It is clear in the figure that all the approaches are equally used in the literature. However, a deep diving in the details and obtained results of each method will show the more promising cues.

The rest of this section reviews the studied works in terms of the cues used: verbal, non-verbal, and hybrid.

3.1 Verbal cues-based methods

This section reviews the deception detection methods based on verbal indicators. We analyze first the methods based on texts. Then, we discuss the speech-based approaches.

3.1.1 Text-based deception detection

Several experiments have been conducted to identify deceptive opinions using deep learning models for text analysis. Zhao et al. [Zhao et al. 2018] trained a CNN by inserting the characteristics of word order into its convolution and pooling layer; this allowed the model to be more effective in spotting deceptive opinions.

Ma et al. [Ma et al. 2016] proposed an approach in which continuous representations of microblog actions are learned to detect deceptive information. The suggested model is based on the gated recurrent unit (GRU) and LSTM to learn hidden representations to capture the variance of specific post-contextual knowledge over time. The detection efficiency of this model, however, suffers from the drawback of its set range of features.

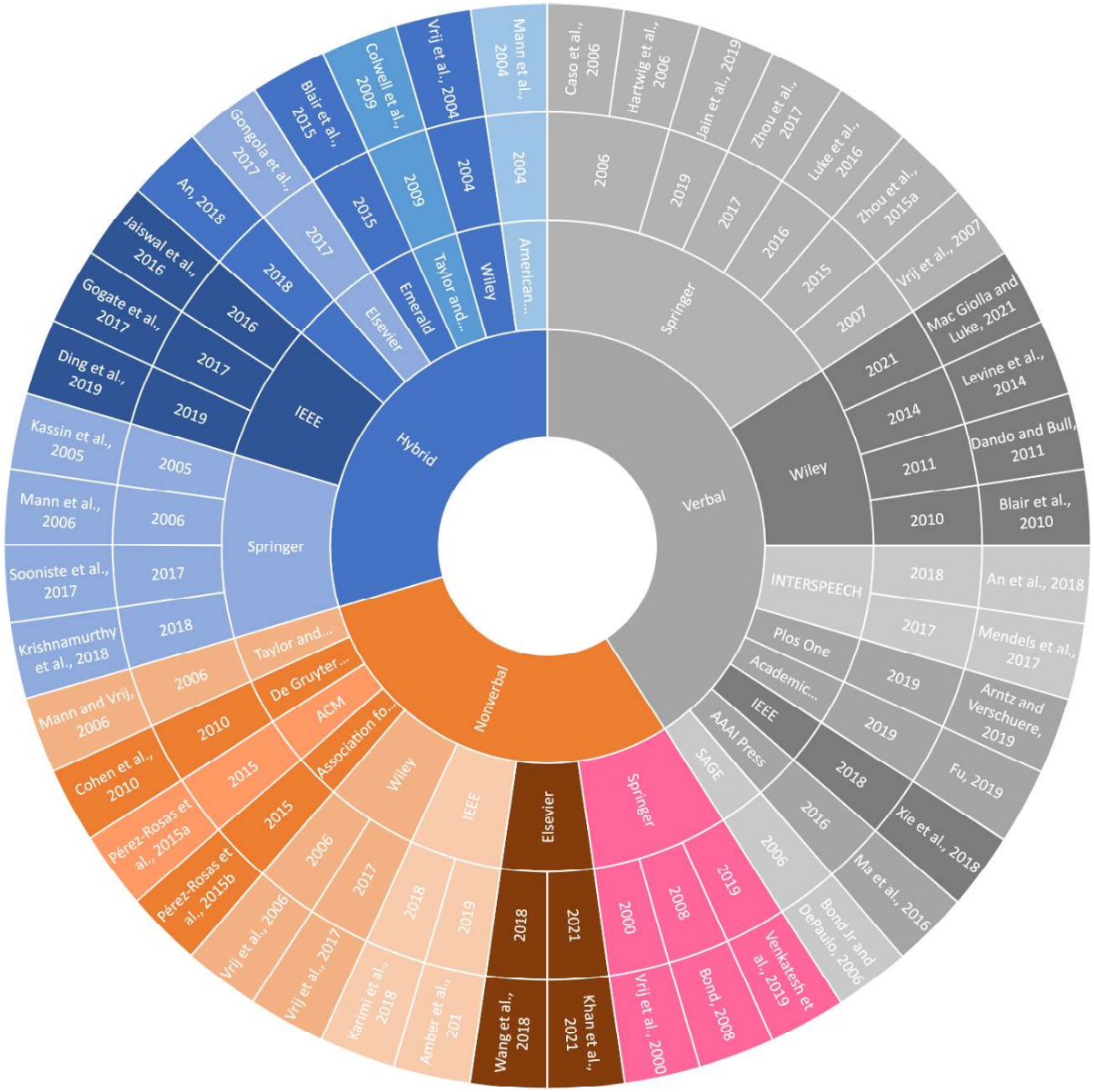

Fig. 1 Cues-based taxonomy of deception detection methods

Another study [Jain et al. 2019] suggested several deep neural network (DNN) approaches for identifying deceptive reviews. The authors used a hierarchical network model with CNN followed by a GRU to select text sequences with variable lengths as inputs. The authors evaluated approaches for the detection of deceptive reviews on five different datasets. The hierarchical CNN-GRU model with all datasets outperformed the CNN or GRU models by 1.6%. The experimental results indicated that CNN and GRU can handle very long review text and are better at identifying deception than other models.

One study was conducted for deceptive detection using LSTM [Fu 2019]. The author used a primary dataset called the Mafiascum dataset, which is a large-scale source of deceptive text of over 700 games of Mafia collected from an internet forum by Ruiter and Kachergis. The dataset includes over 9000 documents, each of which contains messages written in a single game by one player. LSTM, bidirectional LSTM, and CNN were used successfully to detect deception in this study. All three simulations were compared to baseline precision (81.6%), with results of 88.8% for LSTM, 94.7% for bidirectional LSTM, and 94.9% for CNN. These findings indicate that deep learning techniques can effectively classify linguistic signs that show deception.

3.1.2 Speech-based deception detection

Lie speech recognition requires not only focusing on surface details, such as sentences, symbols, and words, but also on features of the internal meaning structure. Therefore, deceptive information is embedded in the speech flow and is not easily identified.

[Ekman 2009] has shown that the fundamental volume of speech increases dramatically when a person lies . Compared to normal situations, people who lie typically experience minor differences in sound intensity, tone, speed, pause period and vocal organs [Kirchhuebel 2013, Zhou et al. 2015b], leading to improvements in some characteristic parameters of speech [Xie et al. 2018]. Therefore, by focusing on observing those changes, deception can be analysed and identified. In reality, in the past, few researchers have been able to identify deception based on speech characteristics.

Many researchers have been investigating the detection of speech and language deception by analysing various lexical and acoustic-prosodic features. Consequently, features must have contextual relationships of dependence, and the model should be able to understand this dependence relationship.

In [Zhou et al. 2015a], the authors proposed a deep belief network based on the K-singular value decomposition algorithm (K-SVD) for detecting lies in speed. This approach aimed to capture significant dynamic deep lie features in regard to time. The University of Arizona's deception database was used for evaluation. The authors achieved a good result, with a correct lie speech detection rate of 69.83%. The experimental findings showed that the deep features suggested in this paper has a higher identification rate than K-SVD sparse features and simple acoustic features.

Another interesting approach attempted to extract the deep lying characteristics of people [Zhou et al. 2017]. The authors described that to identify the characteristics of a lie, it is important to set up a chained layered model. As mentioned in [Zhou et al. 2017], the five chained layers include the language layer the physical layer, the acoustic layer, the pattern layer and the psychological perception layer. This paper suggests a deep learning algorithm, the stack sparse automatic encoder (SSAE), for the extraction of deep deceptive speech features. The SSAE has five layers, which is equivalent to the five speech chain layers. These deep features compensate for the lack of speech with simple acoustic features in deception. Their experiments indicated that the deception detection rate improved by 4% to 10% due to the addition of deep learning characteristics.

Despite the importance of acoustic features to detect speech deception, the temporal features of the speech share the importance of characterizing the speech RNN to capture temporal features. For example, bidirectional LSTM was used to align temporal characteristics with vector dimensions to learn contextual dependencies in speech.

Another paper [Mendels et al. 2017] examined multiple deep neural network based approaches. Experiments were conducted based on a comparison between three feature domains using different machine learning models. These domains

were spectral, acoustic-prosody and lexical feature sets. The study was based on four evaluated approaches, the first approach used lexical features to train (BLSTM). The second approach used Mel-Frequency Cepstral Coefficients (MFCC) features to train BLSTM. Their approach was designed to train DNN on two types of feature domains extracted by openSMILE features. The first domain was the Interspeech 2013 (IS13) ComParE Baseline feature set containing 6373 features from different low-level descriptor (LLD) practical computing contours. The second domain was the Interspeech 2009 (IS09) package of emotional challenge features, which includes 384 features. IS09 and IS13 features were selected since they are supportive of deception detection. These approaches were evaluated on a larger corpus, the Columbia X-Cultural Deception (CXD) Corpus, including over 120 hours of subject speech for deception detection. The last approach was a hybrid deep learning model based on the integration between DNN and LSTM, in which the output of the last hidden layer on DNN was trained on IS09 features and the output of LSTM. The last approach achieved the best performance, with an F1-score of 63.9

LSTM was also used in another study [Xie et al. 2018]. The authors proposed a method of integrating BiLSTM with frame-level acoustic features to detect deception. The proposed method (CovBiLSTM) is based on replacing the multiplication in LSTM with a convolution operation to extract time-frequency features. CovBiLSTM extracts variable length frame-level speech features from samples with different lengths and identifies the depth complexity features associated with deception using a deep neural network system. Frame-level features are derived to avoid temporal information loss, the measurements of which differ with the duration of the speech waveform.

For the purpose of evaluating the deception detection method, the Columbia-SRI-Colorado corpus (CSC) is used. In addition, a local corpus is also reported to validate this algorithm. The average accuracy of the Columbia-SRI-Colorado corpus experiment is 70.3%, whereas the accuracy of the local corpus is 83.6%. The experimental results prove that the proposed method is able to detect depth with a higher accuracy than the traditional machine learning tool SVM.

[An et al. 2018] extracted acoustic-prosodic low-level features, word category features from LIWC, and word scores for pleasantness, activation and imagery from the DAL. They also used the Gensim library [Rehurek and Sojka 2010] to extract two sets of word embedding features using Google's pretrained skip-gram vectors and GloVe vectors. LSTM uses pretrained GloVe vectors to classify data. They initialized a 300-dimensional word embedding layer with GloVe embeddings. An F1 of 0.69, achieved by the MLP-LSTM hybrid model, represents the best performance for deception detection without personality data. The authors compare the performance of MLP, LSTM, and hybrid models with multi-task learning and personality features in this study. They conclude that the detection of deception can be improved by integrating personality features into a deep deception classification model, indicating that deception detection can be enhanced by minimizing interpersonal differences. Speech is segmented in two separate ways in this study, turn and IPU. A turn is represented as speech separated by at least 500 ms of silence from a single speaker, and an IPU is represented as speech separated by at least 50 ms of silence from a single speaker.

3.2 Non-verbal cues-based methods

Nonverbal indicators of lying may be neurological, visual, or vocal such as facial gestures, eye gaze, micro-expressions, etc. Thus, it is possible to detect these signs by observing EEG signals or via video analysis.

3.2.1 P-300 signal-based deception detection

By observing an EEG signal, responses can be identified as deceptive, and it was discovered that the imagination generation portion of the brain is activated when people are being deceptive. If someone speaks the truth, then the brain section known to be responsible for memory management and memory is activated. The P300 signal is a type of EEG signal that has been frequently used by researchers to detect deception. This signal was evoked every time somebody saw particular objects or stimuli. When the subject encounters stimuli in the form of moving images, the P300 signal appears in the EEG recording system after 300 ms.

New deceptive detection methods using deep learning based on EEG-P300 signals is discussed. Accordingly, a lie detection method using the P300 signal has been introduced [Amber et al. 2019]. The authors used CNN for automated optimized feature learning. Thirty subjects were divided into lie and non-lie groups, and EEG signals were recorded from 14 electrodes. The developed model effectively achieved a high accuracy of 99.6%. The experimental results demonstrate that compared to other techniques, the proposed technique achieves the highest accuracy with less training and testing time and shows better performance. A NVIDIA GPU with 12 GB of RAM and a maximum capability of 3.5 was used for training in this study.

Recently, [Saini et al. 2021] attempted to detect deceptive based on EEG signals recorded using wearable sensor headset. They extracted 3 level discrete wavelet transform (DWT) and Morphological features include Amplitude, Absolute Amplitude (AAMP), Peak-to-peak value (P2P), Peak-to-peak time window (P2PT), and Peak-to-peak slope (P2PS) Frequency features include Mode frequency, Mean frequency, Median frequency. Entropy (ENT) and Average Power (AP) from on the recorded signals from each channel. Data co9llected from 30 subjects. The results show an accuracy of 83% obtained by combining PCA with SVM.

3.2.2 Video-based deception detection

In [Karimi et al. 2018], the authors proposed a method for capturing rich information into a coherent model and an end-to-end DEV framework for automatically detecting deceptive videos. CNN was employed to extract features from individual frames. LSTM was used to extract temporal information from deceptive videos. The authors used the actual trial videos provided in [Perez-Rosas et al. 2015a] to test the method and achieved an accuracy of 84.16%.

Furthermore, another study [Venkatesh et al. 2019] proposed a new approach based on CNN for the video-based deception technique. The proposed approach used the sequential input designed to capture spatiotemporal information from the video to obtain nonverbal behaviour. A pretrained GoogleNet CNN was used to extract the deep features from the frame set. The bidirectional LSTMs were linked to GoogleNet to learn the visual representation efficiently. The approaches were evaluated on 121 deceptive and truthful video clips representing a real-life situation carried out on a publicly accessible dataset [Perez-Rosas et al. 2015a]. The results demonstrate the excellent efficiency of the new approach as opposed to the four separate state-of-the-art techniques.

3.2.3 Micro-expression based deception detection

Facial micro-expressions are subtle and uncontrolled expressions that can disclose hidden emotions. Micro-expressions are an important source of information in a variety of areas, such as the identification of deception, mental wellbeing, and study of emotions [Verburg and Menkovski 2019]. Micro-expressions are the primary keys to identifying deception. The most notable characteristics of micro-expressions include short-term and low movement intensity.

Among the various nonverbal actions, micro-expressions are considered promising because they represent the hidden feelings of people and may expose their purpose to catch deception [Ekman 2009, Wang et al. 2018]. The challenge of using micro-expressions is their limited length, weak intensity, and fragmentary units of action. Even if there is controversy over the length of micro-expressions, the widely agreed limit is 0.5 s [Verburg and Menkovski 2019, Yan et al. 2013a].

[Patel et al. 2016] employed the pretrained ImageNet-VGG-f to extract deep features from each frame of videos. Then, an evolutionary search was used to identify the discriminative features of micro-expression identification. [Wang et al. 2018] employed long-term transfer of the CNN (TLCNN) on 560 micro-expression video clips. Deep CNN was trained on TLCNN to extract features from each micro-expression video frame. Then, these Learned features were passed to LSTM, which learned the micro-expression temporal sequence information.

Regardless of the limited sample size of micro-expression data, TLCNN uses the following two transfer learning steps: the first step is to transfer information from expression and the second step is to transfer information from a single frame of micro-expression video clips.

3.3 Hybrid methods

Automated identification of deception from real-life videos is a very challenging task. In fact, the face, audio responses, text responses and body provide valuable clues as to whether a subject is deceptive. Multimodal fusion has recently drawn the interest of many researchers, primarily because of its potential to outperform unimodal models. This section explores different approaches to data fusion.

In [Jaiswal et al. 2016], the authors extracted visual and vocal characteristics for the identification of deception. However, they found that vocal characteristics are not very successful relative to other sources.

In [Krishnamurthy et al. 2018], a basic but strong neural model for the identification of deceit was suggested to overcome multimodality. By integrating features from different modalities, such as film, audio, and text, with features of micro-expression, the authors demonstrate that detecting deceit in real-life videos can be more effective. The authors utilize 3D-CNN to extract visual features from the videos. 3D-CNN not only extracts features from each frame but also extracts spatiotemporal features from the entire video, helping to distinguish facial expressions such as smiling, anxiety or pain.

The authors employed CNN to derive features from a video transcript. First, a pretrained word2vec model was used to extract the vector representations in the transcript for each word. These vectors were convolved and passed to the CNN as an input vector. openSMILE was used to extract features from the input audio. Facial micro-expressions were also used to predict deceptive behaviour. Two techniques were used to fuse features from different modalities: concatenation and Hadamard. The first experiment was run using only using concatenation, in which a single feature vector was simply concatenated from the features from all the modalities. The second example combined the Hadamard and Concatenation methods, in which audio, visual and textual features were fused using the Hadamard software. Then, the features of micro-expression were concatenated with the element. The dataset used in this study contained 121 video recordings of court cases, 61 of which were of a deceptive nature, while the remaining 60 were of a truthful nature. The experimental results on a real-life deception video dataset showed that the MLP learned from features fused from the Hadamard and concatenation method performed better than existing deception detection methods, with a 96.14% accuracy. The authors infer that in the videos, there is a certain sequence that gives extremely important signals to allow such a specific classification. From these studies, the authors found that visual and textual features primarily led to accurate estimates accompanied by micro-expression features.

Another paper [Gogate et al. 2017] introduced a new, deep learning-driven multimodal integration for automatic deception detection, combining auditory signals along with visual and textual features for the first time. The audio features, including pitch and speech volume, were derived using openSMILE open source software. The derived features consisted of multiple low-level descriptors, such as Mel-frequency cepstral coefficients (MFCCs), speed, pitch, loudness and their statistics (e.g., mean, variation, skewness, root quadratic mean). Each utterance was presented as a vector of convolution of the constituent words. Each utterance was either trimmed with a 100-word window or padded with zero. Words were translated into vectors using the 300-dimensional GloVe word representation, which was educated on 840 billion web-crawling words. The CNN learned the abstract word representation with implied semantic knowledge, which stretched with each successive layer over the number of words and, finally, the entire utterance.

Visual features from videos were derived using a 3DCNN, and the simulation results of the proposed approaches, incorporating audio, visual and textual features, achieved the highest accuracies of 92% and 96% compared to the 82% predictive accuracy achieved by the state-of-the-art methods, in which only textual and visual indications are taken into account. In this study, we used in the experiment TensorFlow library and an NVIDIA Titan Xp GPU with 12GB of GDDR5X memory.

Automated detection of deception from real-life videos is an extremely difficult task. Specifically, this detection has to address the following two issues: (1) both the face and body provide useful clues as to whether a subject is deceptive. Thus, how to fuse the two effectively is

the key to the efficiency of an ADD model. (2) Real-life challenging samples are difficult to collect; learning with minimal training data thus threatens most models of deep learning-based ADD. All of these issues have been previously discussed [Ding et al. 2019]. Specifically, a face-focused cross-stream network (FFCSN) is proposed for face-body multimodal computing. The FFCSN model is based on extracting the temporal ambiguity between facial expressions and body motions for deceptive detection. FFCSN is based on using R-CNN to detect face expression (the spatial stream), in which, first, a CNN is used to localize face and, second, ResNet50 is used to extract face expression features. The temporal stream (i.e., body motion) acts on a stack of consecutive deformed optical flow fields to collect motion information. ResNet50 is used to extract temporal features. Two training approaches are introduced to train the FFCSN model, meta-learning and adversarial learning. These approaches have been used to address the issue of training data limitation. Extensive experiments show that the FFCSN model produces state-of-the-art outcomes in both identification of deceit and perception of emotions, with 57.8% accuracy. This study has mentioned that two Tesla K40 GPUs were used for experiments.

There are interesting approaches in which the cognitive side of deception and vision are examined and integrated to detect deception. The study in [Gupta et al. 2019] presented a new multimodal dataset that provides data for the detection of deception using various methods, such as video, audio, EEG and gaze data. The dataset includes a total of 325 records, including 162 lies and 163 truths. Local binary pattern features are extracted from video frames. In terms of audio, a number of frequency features are extracted, such as the zero-crossing rate, spectral centroid, spectral roll off, spectral bandwidth, chroma frequencies and MFCC. In EEG, CNN is used to extract the features of these signals. In Gaza data, fixation, eye blink and public size are extracted as features during the recording. Then, a random forest classifier is run to classify video features, EEG features, and gaze features, where the KNN classifier is used to classify audio features. The decision of all classifiers on all modalities has fusion. The performance score for level fusion has 66.17% accuracy.

Other researchers [An 2018] developed a model that integrates lexical and acoustic features in a deep neural network for personality recognition to recognize deceptive actions in American and Mandarin speakers. The evaluated-on (CXD) Corpus used the following five tools to extract features: low-level descriptor features, LIWC, DAL and word embedding features. Deep neural network with five fully connected layers were used with one output neuron that represents the probability of deception. Additionally, LSTM was trained on word embedding features that contained a 300-dimensional word embedding layer, and LST-Moutput referred to lexical contacts and contained a 256-dimensional vector. The last model (MLP-LSTM hybrid) incorporated the previous two models by holding the last hidden layer output in the deep neural model and integrating it with the 256-dimensional output of the LSTM. The output of the convolutional layer is passed to one output neuron that estimates the probability of deception. The best results for deception identification are an F1 of 0.69, achieved by the MLP-LSTM hybrid model.

4 Critical analysis

Tables 1, 2, and 3 compare the studied works in terms of the model used, the type of features extracted, the used dataset and the performance achieved. Many drawbacks are revealed in the existing works. On the one hand, many methods [Gupta et al. 2019, An 2018, Mendels et al. 2017, An et al. 2018, Wang et al. 2018, Ding et al. 2019] produce low performance results. On the other hand, the used datasets are not diverse and, in many cases, collected in simulated environments which is not realistic to train the deceptive and truthful behaviour. Also, sometimes, the dataset used contains only planned deceptive behaviours [Graciarena et al. 2006, Hirschberg et al. 2005, Howard and Kirchhübel2011, Levitan et al. 2015].

Table 1: State of the art of deception detection approaches - Verbal Methods

Reference	Model	Set of social context and statical features	Dataset	Performance
[Ma et al. 2016]	Multi-layer Gated Recurrent Unit (GRU)	Statistical features and social context features	Microblog datasets using Twitter Microblog datasets using Weibo (weibo.com)	Accuracy: 88% Deceptive: 85% Truth: 93% Accuracy: 91% Deceptive: 87% Truth: 95%
[Jain et al. 2019]	Hierarchical network model with CNN followed by a GRU to select text sequences with variable lengths as inputs	Text sequences	1. Deceptive Opinion Spam Corpus v1.4 2. Four-City Dataset 3. YelpZip Dataset 4. Large Movie Review Dataset 5. Drug Review Dataset	1. Accuracy: 91.9% Precision: 92% Recall: 91% 2. Accuracy: 84.7% Precision: 85% Recall: 85% 3. Accuracy: 66.4% Precision: 67% Recall: 65% 4. Accuracy: 88.9% Precision: 88% Recall: 89% 5. Accuracy: 83.8% Precision: 84% Recall: 83%
[Fu 2019]	1. LSTM 2. Bidirectional LSTM (BiLSTM) 3. CNN	Linguistic features	Mafiascum dataset contains 9000 documents	1. LSTM accuracy: 88.8% Precision: 87.4% Recall: 88% 2. BiLSTM accuracy: 94.7% Precision: 94.8% Recall: 94.5% 3. CNN accuracy: 94.9% Precision: 94.8% Recall: 94.8%

Continued on next page

Table 1: State of the art of deception detection approaches - Verbal Methods (Continued)

Reference	Model	Set of social context and statical features	Dataset	Performance
[Zhou et al. 2015a]	Deep belief network DBN based on the K-singular value decomposition algorithm (K-SVD)	sparse features	The University of Arizona's deception database	Accuracy: 69.83%
[Zhou et al. 2017]	The stack sparse automatic encoder (SSAE) with DNN	Acoustic features, deap feature set	The Soochow University Lie Database	Female deception: 75.3% Male deception: 76.5%
[Mendels et al. 2017]	Hybrid LSTM + DNN	OpenSMILE09 Word Embeddings	(CXD) Corpus	Accuracy: 63.9% Precision: 67.32% Recall: 60.80%
[Xie et al. 2018]	BiLSTM	Frame-level acoustic features, time-frequency features	CSD corpus A local corpus	Accuracy: 70.3% Deceptive: 59.6% Truth 77.2% Accuracy: 83.6% decptive: 74.8% Truth 86.9%
[An et al. 2018]	MLP-LSTM hybrid model	Acoustic-pros odic low-level descriptor features, LIWC (Linguistic Inquiry and Word Count), Dictionary of Affect in Language (DAL), word embedding features	(CXD) Corpus	Precision: 72.58 Recall: 72.98%

Table 2: State of the art of deception detection approaches - Non-verbal Methods

Reference	Model	Set of social context and statical features	Dataset	Performance
[Amber et al. 2019]	Deep learning based on EEG-P300 signals	P300 waves	EEG signals	Accuracy: 99.6%
[Karimi et al. 2018]	CNN to extract features from individual frames and LSTM to extract temporal information from deceptive videos	Deep features extracted from Video frames	Public dataset presented in [Perez-Rosas et al. 2015a]	Accuracy: 84.16%
[Venkatesh et al. 20	A pretrained GoogleNet used to extract the deep features from the frame set. The bidirectional LSTMs were linked to GoogleNet to learn the visual representation	Video frames based extracted features	Public dataset presented in [Perez-Rosas et al.	Accuracy: 100%
[Wang et al. 2018]	Long-term transfer of the CNN (TLCNN)	Features extracted from 560 micro-expression video frame	3 spontaneous micro-expression databases: SMIC , CASME and CASME 2	ccuracy: 60.11%
[Perez-Rosas et al.	Decision Trees (DT) and Random Forest (RF)	Verbal and non-verbal modalities	Videos collected from public court trials	Accuracy: 75%
[Khan et al. 2021]	Random forests (RF)	36 dimensional numerical features (facial and eye micro-movements)	Data were collected from 100 participants (50 truthful, 50 deceptive)	Accuracy: 78% Sensitivity 72% Specificity 84%

Table 3: State of the art of deception detection approaches - Hybrid Methods

Reference	Model	Set of social context and statical features	Dataset	Performance
[Jaiswal et al. 2016]	SVM	Extraction of visual and vocal characteristics (openFace)	dataset involved of 61 deceptive and 60 truthful videos obtained from various Youtube channels	Accuracy: 78.9% Truth: 81.10% Deceptive: 76.80%
[Krishnamurthy et al. 2018]	3D-CNN to extract visual features from videos frames and spatiotemporal features from the entire video	Features extracted from films, audios, and texts, with features of micro-expressions	121 video recordings of court cases	Accuracy: 96.14%
[Gogate et al. 2017]	A new, deep learning-driven multimodal integration for automatic deception detection, combining auditory signals along with visual and textual features	(MFCCs), speed, pitch, loudness and their statistics	Real-life Trial corpus	Accuracy: 96.42% Precision: 96% Recall 95%
[Ding et al. 2019]	A face-focused cross-stream network (FFCSN) is proposed for face-body multimodal computing	Facial expressions and body motions	Real-life dataset	Accuracy: 57.8%
[Gupta et al. 2019]	fusion with two modalities (RF used to classify video features, EEG features, and gaze features, and KNN used to classify audio features)	Features extracted from videos, audios, EEG and gaze data	35 university students 325 recordings	Accuracy: 66.17%

Continued on next page

Table 3: State of the art of deception detection approaches - Hybrid Methods (Continued)

Reference	Model	Set of social context and statical features	Dataset	Performance
[An 2018]	MLP-LSTM hybrid	Lexical and acoustic features	CXD Corpus	F1 of 0.69
[Karnati et al. 2021	CNN	Video + Audio features	1. Bag of lies 2. Real-life trail 3. Miami University Deception Detection	1. Accuracy of Bag-of-lies: 96.04% 2. Accuracy of RL: 97% 3. Accuracy: 98%
[Mohan and Seal 2021]	MLP	1. Video + Audio features 2. EEG + Video + Audio	1. Real-life trail 2. Bag of lies	1. Accuracy of RL: 76% 2. Accuracy of Bag-of-lies: 70%
[Mai et al. 2021]	CNN	temporal , frequence and time-frequence features	EEG-Based BCI system	
[Ascensión and Montero 2021]	LSTM	Gaze and speech features	Bag of Lies	Deceptive: 66.58% Truth: 70.99% Accuracy: 68.66%
[Baghel et al. 2020]	CNN	EEG Raw Data	Dryad dataset EEG lie detection dataset	Accuracy of Dryad dataset: 84.44% Accuracy of EEG lie detection dataset: 82.00%

5 Existing datasets and libraries

In the evolution of deceptive detection, datasets as well as tools and packages development have played a crucial role not only as a shared ground for testing and evaluating the success of competing algorithms but also in moving the field towards more complicated and difficult problems. This section first summarizes the existing dataset sources, then gives an overview of the developed libraries to support the techniques of deception detection.

5.1 Datasets

This section recalls the used datasets to support the existing approaches of deception detection. As shown in table 4, these datasets are classified into the following categories based on data types: audio, text, video, gaze, and EEG signals.

- Columbia-SRI-Colorado Corpus

The Columbia-SRI-Colorado Corpus (CSC) Deceptive Speech was recorded by the University of Columbia, SRI International and Colorado Boulder Universities. The CSC contains 32 hours of audio interviews from 32 Standard American English native speakers (16 male, 16 female) gathered from students and the community of Columbia University. The participants were told they were taking part in a communication project that aimed to recognize individuals who matched the profile of America's top entrepreneurs. The participants carried out tasks and replied to questions in six regions. They were later told that some of those areas had earned low scores and did not match the profile. The participants then took part in an interview in which they were told to persuade the interviewer that they had already earned high scores in all the areas and matched the profile in reality. The interviewer's job was to decide how he or she felt the participants had actually done. For each interviewer question, the participants were asked to identify whether the answer was true or whether it contained any false information by clicking one of two pedals hidden from the interviewer. The interviews were hand transcribed orthographically using the NIST EARS transcription instructions. [Hirschberg et al. 2005]

Table 4 Datasets

Dataset	Number of participants	Data Type	Data Size	Scenario
Columbia-SRI-Colorado Corpus (CSC)	32	Only Audio (1)	-	Hypothetical Scenario
Columbia Cross-Cultural Deception (CXD) Corpus	346			
ReLiDDB	40	Only Audio (1)	-	Hypothetical Scenario
Open Domain	512	Only Text (1)	7168	Crowdsourcing
EEG-P300	11	Only EEG (1)	88	Hypothetical Scenario
Real Life Trials	56	Video, Audio, Text (3)	121	Realistic Scenario
Multi Modal	30	Video, Audio, Thermal, Physiological (4)	150	Hypothetical Scenario
Bag-of-Lies	35	Video, Audio, EEG, Gaze (4)	325	Realistic Scenario
SMIC, CASME, and CASME 2	16 participants 26		164 195 247	
Facial micro expression			39 expressions	

- Columbia Cross-Cultural Deception Corpus

The 'fake resume' model was used in Columbia Cross-Cultural Deception (CXD) in which each participant was asked to provide true and false biographical information; this information represented the ground truth. The participants were separated and instructed to practice a lying game

with the other participants separately. The participants were asked to switch between being questioned by and questioning their partner to collect responses to a collection of 24 biographical issues. The game ensured no direct interaction between subjects; therefore, both participants were sitting opposite each other and were separated by a curtain. One participant played the role of the interviewer for the first half of the game, while the other replied to the biographical questions, truthfully answering half of the questions and lying for the other half. The roles of the participants were reversed for the second half of the game. [Nasriet al. 2016]

- ReLiDDB

ReLiDDB was obtained from 40 participants who were students at universities. In unconstrained acoustic environments, this speech corpus was documented. Every participant was asked to tell true and false stories. The period varied from person to person and from records of the same person. The range of length was approximately 1 minute 20 seconds per story. The data were recorded in a sound studio. Then, outdoor records were included for the sake of environmental impact analysis. [Nasriet al. 2016]

- Open domain

The Amazon Mechanical Turk was used for these data. Every employee was requested to tell seven lies and seven truths regarding topics of their own choice, each consisting of a single sentence. Additionally, personal data were obtained from the participants, including their gender, age, region of residence, and education level. The dataset consists of 7168 sentences from 512 participants. There are 3584 truths in the dataset and 3584 lies. Participants were aged from 18 to 72 years of age. [Perez-Rosas and Mihalcea 2015]

- EEG-p300

Data from twelve participants (10 men and 2 women) with an age of approximately 19±1 years were reported in an experiment. EEG data were obtained using a Mitsar 202 EEG device with five Ag/AgCl electrodes enclosed in an elastic hat. The frontal (Fz), core (Cz), parietal (Pz), and occipital (O1 and O2) electrodes were the electrodes used in this experiment. The subjects were educated on the experimental processes before the experiment. The subjects were split into the following two categories: truth and lies. When objects (P, T, and stimulus I) were shown, the subjects had to perform certain functions. [Turnip et al. 2017]

- Real Life Trials

In this dataset, deceptive clips were obtained from a suspect in a courtroom for guilty verdicts, and truthful clips were obtained from victims in the same courtroom. In other cases, deceptive recordings were obtained from a suspect admitting to a crime, and truthful clips were collected from the same suspect while answering questions about evidence that had determined to be factual by authorities. Testimonies confirmed by criminal reports were classified as facts by witnesses, while testimonies were classified as deceptive in favour of a convicted suspect. There are 121 videos in the final dataset, including 61 deceptive and 60 truthful trial videos. The videos in the dataset have an average length of 28.0 seconds. The total video lengths of the deceptive and truthful videos are 27.7 seconds and 28.3 seconds, respectively. The dataset includes 21 female speakers and 35 male speakers, with ages ranging from 16 to 60 years. [Perez-Rosas et al. 2015a]

- Multimodal

In Multimodal dataset [Perez et al. 2014], the following four types of data were gathered: video, audio, thermal, and physiological. Two separate stations were identified; in the laboratory, one station was a private area and the other station was the experimental device. Each participant was asked before each recording session to sit at the recording station. Because the equipment was vulnerable to movement, the participants were often asked to stop making any repetitive gestures of their head or hands. The purpose of the movement restrictions was to collect high quality data from the cameras and the physiological sensors, which are especially important for temperature and skin conductance measurements since they are collected by means of wired sensors that must be in direct contact with the skin. Based on the scenario being run, each participant was required to react either truthfully or deceptively.

- Bag-of-Lies

The bag-of-lies [Gupta et al. 2019] contains several modalities, such as visual, audio, EEG and eye gaze modalities. A standard mobile camera and microphone were used to capture video and audio samples from the subjects, which means that the dataset fits practical situations, such as a YouTube video blog or CCTV footage, for which high-quality videos might not be available. For EEG data collection and Gazepoint GP3 Eye Sensor gaze data collection, a 14-Channel Emotiv EPOC+EEG headset was used.

- SMIC database

This dataset was collected by [Li et al. 2013] and includes the following three datasets: high-speed camera (HS), near-infrared camera (NIR), and normal visual camera (VIS). The HS dataset was captured using a 100-fps camera with 50 frames in the longest micro-expression clips. A 25-fps camera captured the VIS and NIR datasets, and the longest micro-expression clips included 13 frames. A total of 164 micro-expression image sequences from 16 participants are included in the HS dataset. In the SMIC database, the micro-expression picture sequences are categorized into the following three categories: positive (happiness), negative (sadness, disgust and fear), and surprise. The series of images begins from a neutral frame, and the second frame is the facial expression's starting point and eventually ends when the facial expression returns to neutral.

- CASME database

Yan et al. in [Yanet al. 2013b] collected the CASME database. This dataset is composed of 195 videos, with 50 frames in the longest micro-expression clips. In the CASME database, the micro-expression picture sequences are grouped into the following four categories: positive (happiness), negative (sadness, fear and disgust), surprise and tense (tense, repression). The CASME database is divided into the following two classes: A and B. A BenQ M31 camera captured the samples in Class A at 60 fps. A Point Grey GRAS03K2C camera captured the samples in Class B at 60 fps. We used the image sequence from class A in this analysis.

- Facial micro-expression

This work introduced a novel dataset consisting of 121 deceptive and truthful video clips from real court trials. Transcriptions of these videos were used to extract several linguistic features, and the videos were manually annotated for the presence of multiple gestures that were used to extract nonverbal features. Moreover, a system that jointly used the verbal and nonverbal modalities was developed to automatically detect the presence of deception. The performance of the system was compared to that of human annotators. The data include 121 videos, containing 61 deceptive and 60 truthful clips. The videos in the dataset have an average length of 28.0 seconds. The data consist of 21 unique female speakers and 35 unique male speakers, with a distribution of ages between 16 and 60 years. The video clips were labelled deceptive or truthful based on guilty verdict, not guilty verdict, and exoneration. Examples of famous trials included in the dataset are the trials of Jodi Arias, Donna Scrivo, Jamie Hood, Andrea Sneiderman, Mitchelle Blair, Amanda Hayes, Crystal Mangum, Marissa Devault, Carlos Miller, Michael Dunn, Bessman Okafor, Jonathan Santillan, among other trials. [Perez-Rosas et al. 2015a]

5.2 Libraries and tool packages

This section summarizes the existing libraries and tool packages, as shown in table 5. These packages are classified based on the following categories based on data types: audio, text, and video.

- OpenSMILE

OpenSMILE is an open-source toolkit used to capture high-dimensional audio features and to evaluate other modalities, including visual physiological signals and other physical sensors. Such features include low-level descriptor functions contours (LLDs). Specifically, IS13-ComParE opens the SMILE interface to obtain 6373 dimensional features for each audio input. In addition, the Interspeech 2009 Emotion Challenge feature set (IS09) can be extracted to represent emotion recognition, and it contains 32 descriptors.

- Global Vectors for Word Representation (GloVe)

The Global Vectors for Word Representation (GloVe) is an unsupervised learning algorithm developed at Stanford to generate the embedding of words. The embedding of words is created by

Table 5 Tool packages

Package	Type of data	Description
openSMILE	audio	Extract features from audio file
Global Vectors for Word Representation) (GloVe)	word	
Word2Vec	text	Extract vector representation for every word in the transcript
Linguistic Inquiry and Word Count (LIWC)	text analysis software program	
N-gram	text	
Dictionary of Affect Features	lexical analysis	Extract emotion content of speech
OpenFace	facial behaviour features	

using statistics obtained from the co-occurrence of global words in a corpus. This is a type of bilinear regional log regression using global matrix factorization as well as local context window models. GIoVe has been integrated with other modalities to analyse and detect deceptive videos [Gogate et al. 2017].

- Word2Vec

Word2Vec is one of the most common strategies for learning shallow neural network word embeddings. By using a corpus as the input, a word2vec model is trained. Word2Vec was created in 2013 by Google's Tomas Mikolov. The features extracted from word2vec have been used for deceptive detection [Krishnamurthy et al. 2018]. Two strategies include Skip Gram and Common Bag Of Words (CBOW). This approach takes each word's context as an input and aims to predict the word that corresponds to the context.

- N-gram

One form of the language model (LM) is the N-gram model, which is constructed by counting how many word sequences appear in the text of a corpus and then calculating the probabilities. An N-gram model estimates a word's frequency dependent on the frequency of the previous N-1 terms.

- Linguistic Inquiry and Word Count (LIWC)

LIWC is a software tool used in textual studies. LIWC tests the degree to which people use multiple types of words that can quantify the degree of positive or negative emotions, self-references, and explanatory sentences. LIWC contains 70 language metrics in each text. For psychologists, the design of LIWC has made it a favourite. Finding psychological properties is easier using LIWC than other tools. From language, LIWC can extract perceptual, behavioural, and emotional components.

- Dictionary of Affect in Language

Dictionaries of affect provide great promise for lexical influence detection, as they contain information on the affective features of single words or phrases that can be used to measure the emotional sound of the corresponding turn of a conversation. Dictionaries of affect are typically composed of human common-sense information or use rating tools, such as semantic differential scales, to influence words.

DAL is composed of 8,742 distinctly inflected words that are distinguished by three aspects of their emotional connotation: appraisal, activation and imagery. Evaluation scores vary from 1 (unpleasant) to 3 (pleasant), activation scores range from 1 (passive) to 3 (active), and imaging range scores range from 1 (difficult to build a mental image of this word) to 3 (easy to create a mental image). Human judgement is calculated as the total rating.

Affect dictionaries have also been found to be helpful in distinguishing non-deceptive expressions that are deceptive. For example, to extract features from expression, Hirschberg et al. [Hirschberg et al. 2005] made use of the LIWC and DAL dictionaries. They noted that both the pleasantness score and the frequency of words of optimistic sentiment appeared to be promising elements in the prediction of deceit. Deceptive expression appears to have a higher level of pleasantness and a higher percentage of words of positive emotion than real speech.

- OpenFace

OpenFace is the first toolkit capable of detecting facial landmarks, estimating head positions,

understanding facial action units, and estimating eye gaze with a usable source code for both model running and instruction. In all of the tasks listed above, the computer vision algorithms that form the heart of OpenFace have state-of-the-art outcomes. In addition, OpenFace can be used in real-time and is able to run without any specialist hardware other than a basic webcam. OpenFace can be used to analyse the movement of facial features to detect deception [Jaiswal et al. 2016].

6 Challenges of deception detection

Unlike physiological devices, videos are non-invasive and cost-effective and do not require skilled operators. Furthermore, videos provide rich sources of evidence to help assess deception detection, i.e., visual sources (e.g., frowned eyebrows), verbal sources (e.g., silent pauses), and linguistic sources (e.g., psycholinguistic features). [Karimi 2018] indicates that videos offer excellent opportunities to detect deception. In addition, temporal signs may be predictive of deception in a series of video clips. For example, eyebrow raising can be identified as a deceptive sign when processing a series of frames of a person's face [Wu et al. 2018].

Regarding the role of deception identification, certain nonlinear features are related, such as auditory perception and psychological acoustics, which are difficult to extract and are extremely complex to compute [Zhou et al. 2015a].

On the one hand, there are three enormous obstacles to detecting deception in videos. First, video data are intrinsically multimodal and involve diverse and inter-linked sources (i.e., visual, auditory, and textual sources) as well as temporal dependencies; additionally, the fact that videos actually contain complicated temporal correlations makes them inherently complex and naturally multimodal. The inherent nature of the deception itself makes the detection of deceit much more difficult in videos. Second, to further detect deception in videos, it is necessary to identify deception-related signs. The third challenge includes limited accessible labeled real-world videos. Earlier research on video-based deception detection is constrained by datasets containing only planned deceptive behaviours [Graciarena et al. 2006, Hirschberg et al. 2005, Howard and Kirchhübel2011, Levitan et al. 2015].

Furthermore, in real-world applications, most videos do not include deceptive labels. The labelling of deceptive videos requires experts from the domain and is quite expensive [Karimi et al. 2018], and the task is beyond human skill. Furthermore, while there have been many attempts to synthesize deceptive videos in the laboratory, such videos are not only unethical but are also constrained in scale due to technical limitations [Sen et al. 2018].

Hence, the effectiveness of using video to identify deception in the real world is uncertain. The first attempted deception detection using real-life videos was initially introduced in previous research [Perez-Rosas et al. 2015a, Perez-Rosas et al. 2015b], in which deceit recognition of claims made by witnesses and defendants was conducted using a corpus obtained from proceedings in Italian courts. Modern multimodal deception data collection was first implemented for real-life videos from court trials, and a mixture of features derived from various modalities were used to identify deceptions. Consequently, more advanced deception decision methods have been proposed [Abouelenien et al. 2016, Jaiswal et al. 2016, Wu et al. 2018].

To summarize the outcome of this research work and notice the challenges when working on deception detection, we mention the following points:

1. In term of accuracy, lots of reviewed studies produce low accuracy that is less than 70%.
2. For the available datasets, which are not diverse, first they are collected in simulated environments which is not realistic to train the deceptive and truthful behaviour. Second, limited are the subjects used in datasets. Finally, there still static questions asked to participants.
3. There is contradiction of accuracy and outcomes: for example psychological studies summarize that verbal cues are more important than visual ones. However professional investigators reported otherwise. Likewise, deep learning models find visual cues as well.

7 Conclusion

Psychological methods of deception detection pay less attention to verbal signs [An 2015]. Instead of looking for visual indicators of lying, such as lack of eye contact or fidgeting, psychologists are concentrating on devising proactive techniques that interviewers might use to elicit deceptive [Hartwig et al. 2014]. However, physiological approaches are invasive and costly. In addition, physiological approaches require skilled operators [Karimi 2018].

Professional investigations accentuated the importance of other indicators, such as act shifts, foot and hand movement, as well as gaze aversion [Bond Jr and DePaulo 2006, Davis and Markus 2006, Lajevardi and Hussain 2012, Ekman 2009]. Some professionals claim that ignoring visual information may improve the accuracy. According to the investigations of [Bradford et al. 2013], the validity of confessions is best established through audio recordings or written text-based remarks, as these appear to be less subject to the biases that might damage credibility evaluations based on visual evidence.

Non-verbal indicators as well as mixing verbal and non-verbal cues together have made it possible to obtain more accurate results when looking for deception detection [Venkatesh et al. 2019, Krishnamurthy et al. 2018, Gogate et al. 2017, Amber et al. 2019]. The cited papers have used machine learning techniques towards automatic deception detection, by selecting the features with the most added value.

Last and most critically, an integral desideratum of a system for the identification of deception provides an interpretable approach that describes whether a video is misleading or truthful in a humanly understandable manner. The problem here is that the effectiveness of using video to identify deception in the real world is uncertain.

Automatic detection of micro-expressions from Web cameras or off-line videos in the setting of interrogation interviews can significantly assist intelligence officers in identifying normal or even trickery signs of suspects. Computer vision technologies, such as deep learning models, can be used in rapid security scanning without the need for professional staff or physical interactions. These technologies are able to produce reasonable and acceptable results and have received much scientific interest in recent years because of the close relationship among object recognition, video processing and image comprehension.

Furthermore, multimodal fusion has recently attracted the attention of many researchers, mainly because of its ability to surpass unimodal models. Therefore, providing deep learning models with sufficient and unique datasets, such as face, audio, text and body expression, can enhance their performance.

However, there may be many confounding variables, such as culture, socioeconomic status, and personal beliefs, that can influence a person's language rather than their intent for deceit.

Declarations

- Ethics approval
  The paper is not currently being considered for publication elsewhere.
- Funding details
  This project was supported by the Deanship of Scientific Research at Prince Sattam Bin Abdulaziz University under the research project No 2020/01/11744.
- Conflict of interest
  The authors declare that they have no conflict of interest.
- Informed consent
  Not applicable.
- Authors' contribution
  The authors contributed equally to this work.

References

[Abouelenien et al. 2016] Abouelenien M, Pérez-Rosas V, Mihalcea R, Burzo M (2016) Detecting deceptive behavior via integration of discriminative features from multiple modalities. IEEE Transactions on Information Forensics and Security 12, 1042–1055.

[Adelson 2004] Adelson R (2004) Psychological sleuths–Detecting deception [WWW Document]. https://www.apa.org. URL https://www.apa.org/monitor/julaug04/detecting (accessed 1.2.21).

[Alaskar 2018] Alaskar H (2018) Deep Learning-Based Model Architecture for Time-Frequency

Images Analysis. International Journal of Advanced Computer Science and Applications 9.

[Al-tahri et al. 2022] Al-tahri M, Al-tamimi N, Al-harbi S, Abduallah A, Alaskar H, Sbai Z (2022) Deceptive Detection based on Spectrum Analysis using Deep Learning. 2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), 2022

[Amber et al. 2019] Amber F, Yousaf A, Imran M, Khurshid K (2019) P300 Based Deception Detection Using Convolutional Neural Network, in: 2019 2nd International Conference on Communication, Computing and Digital Systems (C-CODE). IEEE, pp. 201–204.

[An 2018] An G (2018) Personality Recognition For Deception Detection.

[An 2015] An G (2015) Literature review for Deception detection. Dr. Diss. City Univ. New York.

[An et al. 2018] An G, Levitan S.I, Hirschberg J, Levitan R (2018) Deep Personality Recognition for Deception Detection., in: INTERSPEECH. pp. 421–425.

[Andersson 2018] Andersson G (2018) Classification of Heart Sounds with Deep Learning.

[Ascensión and Montero 2021] Ascensión G-A and Montero J M (2021) "Detecting Deception from Gaze and Speech Using a Multimodal Attention LSTM-Based Framework" Applied Sciences 11, no. 14: 6393. https://doi.org/10.3390/app11146393

[Kleinberg et al. 2019] Kleinberg B, Arntz A, Verschuere B (2019) Being accurate about accuracy in verbal deception detection. PloS one, 14(8), e0220228. https://doi.org/10.1371/journal.pone.0220228

[Baghel et al. 2020] Baghel N, Singh D, Dutta M.K, Burget R, and Myska V (2020) "Truth Identification from EEG Signal by using Convolution neural network: Lie Detection," 43rd International Conference on Telecommunications and Signal Processing (TSP), 2020, pp. 550-553, doi: 10.1109/TSP49548.2020.9163497.

[Bond Jr and DePaulo 2006] Bond Jr C.F, DePaulo B.M (2006) Accuracy of deception judgments. Personality and social psychology Review 10, 214–234.

[Bradford et al. 2013] Bradford D, Goodman-Delahunty J, Brooks K.R, (2013) The impact of presentation modality on perceptions of truthful and deceptive confessions. Journal of Criminology.

[Burgoon et al. 2015] Burgoon J, J Mayew W, Giboney J.S, C Elkins A, Moffitt K, Dorn B, Byrd M, Spitzley L (2015) Which Spoken Language Markers Identify Deception in High-Stakes Settings? Evidence From Earnings Conference Calls, Journal of Language and Social Psychology

[Carpenter et al. 2018] Carpenter K.A, Cohen D.S, Jarrell J.T, Huang X (2018) Deep learning and virtual drug screening. Future medicinal chemistry 10, 2557–2567.

[Caso et al. 2006] Caso L, Maricchiolo F, Bonaiuto M, Vrij A, Mann S (2006) The impact of deception and suspicion on different hand movements. Journal of Nonverbal behavior 30, 1–19.

[ChoAl2014] Cho K, Van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y (2014) Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078.

[Cohen et al. 2010] Cohen D, Beattie G, Shovelton H (2010) Nonverbal indicators of deception: How iconic gestures reveal thoughts that cannot be suppressed. Semiotica 2010, 133–174.

[Davis and Markus 2006] Davis M, Markus K.A (2006) Misleading cues, misplaced confidence: An analysis of deception detection patterns. American Journal of Dance Therapy, 28, 107-126.

[deVos et al. 2016] de Vos B.D, Wolterink J.M, de Jong P.A, Viergever M.A, Isgum I (2016) 2D image classification for 3D anatomy localization: employing deep convolutional neural networks, in: Medical Imaging 2016: Image Processing. International Society for Optics and Photonics, p. 97841Y.

[Deng and Yu 2014] Deng L, Yu D (2014) Deep learning: methods and applications. Foundations and Trends in Signal Processing 7, 197–387.

[DePaulo et al. 2003] DePaulo B.M, Lindsay J.J, Malone B.E, Muhlenbruck L, Charlton K, Cooper H (2003) Cues to deception. Psychological bulletin 129, 74.

[DePaulo and Morris 2004] DePaulo B.M, Morris W.L (2004) Discerning lies from truths: Behavioural cues to deception and the indirect pathway of intuition.

[Granhag and Stromwall 2004] Granhag P, Stromwall L (Eds.) (2004) The Detection of Deception in Forensic Contexts. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511490071.

[Ding et al. 2019] Ding M, Zhao A, Lu Z, Xiang T, Wen J.R (2019) Face-focused cross-stream network for deception detection in videos, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7802–7811.

[Ekman 2009] Ekman P (2009) Lie catching and microexpressions. The philosophy of deception 1, 5.

[Ekman 2003] Ekman P (2003) Darwin, deception, and facial expression. Annals of the New York Academy of Sciences 1000, 205–221.

[Fitzpatrick et al. 2015] Fitzpatrick E, Bachenko J, Fornaciari T (2015) Automatic detection of verbal deception. Synthesis Lectures on Human Language Technologies 8, 1–119.

[Fu 2019] Fu L (2019) Deception Detection in Online Mafia Game Interactions.

[Gale 1988] Gale A (Ed.) (1988) The polygraph test: Lies, truth and science. Sage Publications, Inc; British Psychological Societ

[Gogate et al. 2017] Gogate M, Adeel A, Hussain A (2017) Deep learning driven multimodal fusion for automated deception detection, in: 2017 IEEE Symposium Series on Computational Intelligence (SSCI). IEEE, pp. 1–6.

[Graciarena et al. 2006] Graciarena M, Shriberg E, Stolcke A, Enos F, Hirschberg J, Kajarekar S (2006) Combining prosodic lexical and cepstral systems for deceptive speech detection, in: 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings. IEEE, p. I–I.

[Granhag and Hartwig 2008] Granhag P.A, Hartwig M (2008) A new theoretical perspective on deception detection: On the psychology of instrumental mind-reading. Psychology, Crime Law 14, 189–200.

[Graves et al. 2005] Graves A, Fernandez S, Schmidhuber J (2005) Bidirectional LSTM networks for improved phoneme classification and recognition, in: International Conference on Artificial Neural Networks. Springer, pp. 799–804.

[Guadagno et al. 2012] Guadagno R.E, Okdie B.M, Kruse S.A (2012) Dating deception: Gender, online dating, and exaggerated self-presentation. Computers in Human Behavior 28, 642–647.

[Gupta et al. 2019] Gupta V, Agarwal M, Arora M, Chakraborty T, Singh R, Vatsa M (2019) Bag-of-lies: A multimodal dataset for deception detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 0–0.

[Hartwig et al. 2014] Hartwig M, Granhag P.A, Luke T, (2014) Strategic use of evidence during investigative interviews: The state of the science. Credibility assessment 1–36.

[Hartwig et al. 2010] Hartwig M, Granhag P.A, Stromwall L.A, Doering N (2010) Impression and information management: On the strategic

self-regulation of innocent and guilty suspects. The Open Criminology Journal 3.

[Hillman et al. 2012] Hillman J, Vrij A, Mann S (2012) Um ... they were wearing ...: The effect of deception on specific hand gestures. Legal and Criminological Psychology, 17(2):336–345, 2012.

[Hinton et al. 2012] Hinton G, Deng L, Yu D, Dahl G.E, Mohamed A, Jaitly N, Senior A, Vanhoucke V, Nguyen P, Sainath T.N (2012) Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. IEEE Signal processing magazine 29, 82–97.

[Hinduja and Patchin 2015] Hinduja S and Patchin J. W (2015). Bullying Beyond the Schoolyard: Preventing and Responding to Cyberbullying (2nd Ed.). Thousand Oaks, CA: Sage Publications (978-1483349930)

[Hirschberg et al. 2005] Hirschberg J.B, Benus S, Brenier J.M, Enos F, Friedman S, Gilman S, Girand C, Graciarena M, Kathol A, Michaelis L (2005) Distinguishing deceptive from non-deceptive speech.

[Ho and Hollister 2013] Ho, S.M and Hollister J.M (2013) Guess who? An empirical study of gender deception and detection in computer-mediated communication. Proceedings of the American Society for Information Science and Technology 50, 1–4.

[Hochreiter and Schmidhuber 1997] Hochreiter S, Schmidhuber J (1997) LSTM can solve hard long time lag problems, in: Advances in Neural Information Processing Systems. pp. 473–479.

[Howard and Kirchhübel2011] Howard D.M and Kirchhübel C (2011) Acoustic correlates of deceptive speech–an exploratory study, in: International Conference on Engineering Psychology and Cognitive Ergonomics. Springer, pp. 28–37.

[Jain et al. 2019] Jain N, Kumar A, Singh S, Singh C, Tripathi S (2019) Deceptive Reviews Detection Using Deep Learning Techniques, in: International Conference on Applications of Natural Language to Information Systems. Springer, pp. 79–91.

[Jaiswal et al. 2016] Jaiswal M, Tabibu S, Bajpai R (2016) The truth and nothing but the truth: Multimodal analysis for deception detection, in: 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW). IEEE, pp. 938–943.

[Jupe and Keatley 2019] Jupe L and Keatley D.A (2019) Airport Artificial Intelligence Can Detect Deception – Or am I Lying. Security Journal

[Karimi 2018] Karimi H (2018) Interpretable Multimodal Deception Detection in Videos, in: Proceedings of the 20th ACM International Conference on Multimodal Interaction. pp. 511–515.

[Karimi et al. 2018] Karimi H, Tang J, Li Y (2018) Toward end-to-end deception detection in videos, in: 2018 IEEE International Conference on Big Data (Big Data). IEEE, pp. 1278–1283.

[Karnati et al. 2021] Karnati M, Seal A, Yazidi A, and Krejcar O (2021) "LieNet: A Deep Convolution Neural Networks Framework for Detecting Deception," in IEEE Transactions on Cognitive and Developmental Systems, doi: 10.1109/TCDS.2021.3086011.

[Khan et al. 2021] Khan W, Crockett K, O'Shea J, Hussain A, Khan B.M (2021) Deception in the eyes of deceiver: A computer vision and machine learning based automated deception detection. Expert Systems with Applications 169, 114341.

[Kirchhuebel 2013] Kirchhuebel C (2013) The acoustic and temporal characteristics of deceptive speech (PhD Thesis). University of York.

[Kou et al. 2022] Kou G, Yi K, Xiao H, Peng R (2022) Reliability of a Distributed Data Storage System Considering the External Impacts. IEEE Transactions on Reliability, doi: 10.1109/TR.2022.3161638.

[Krishnamurthy et al. 2018] Krishnamurthy G, Majumder N, Poria S, Cambria E (2018) A deep learning approach for multimodal deception detection. arXiv preprint arXiv:1803.00344.

[Lajevardi and Hussain 2012] Lajevardi and Hussain (2012) Automatic facial expression recognition: Feature extraction and selection

[Levine et al. 2011] Levine T.R, Serota K.B, Shulman H, Clare D.D, Park H.S, Shaw A.S, Shim J.C, Lee J.H (2011) Sender demeanor: Individual differences in sender believability have a powerful impact on deception detection judgments. Human Communication Research 37, 377–403.

[Levitan et al. 2015] Levitan S.I, An G, Wang M, Mendels G, Hirschberg J, Levine M, Rosenberg A (2015) Cross-cultural production and detection of deception from speech, in: Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection. pp. 1–8.

[Li et al. 2014] Li J, Ott M, Cardie C, Hovy E (2014) Towards a general rule for identifying deceptive opinion spam, in: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 1566–1576.

[Li et al. 2013] Li X, Pfister T, Huang X, Zhao G, Pietikäinen M (2013) A spontaneous micro-expression database: Inducement, collection and baseline, in: 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (Fg). IEEE, pp. 1–6.

[Ma et al. 2016] Ma J, Gao W, Mitra P, Kwon S, Jansen B.J, Wong K.F, Cha M (2016) Detecting rumors from microblogs with recurrent neural networks.

[Mai et al. 2021] Mai N, Nguyen T, Chung W (2021) Deception Detection Using a Multichannel Custom-Design EEG System and Multiple Variants of Neural Network In book: Intelligent Human Computer Interaction (pp.104-109) DOI:10.1007/978-3-030-68449-5$_1$1

[Mann et al. 2008] Mann S.A, Vrij A, Fisher R.P, Robinson M (2008) See no lies, hear no lies: Differences in discrimination accuracy and response bias when watching or listening to police suspect interviews. Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition 22, 1062–1071.

[Mendels et al. 2017] Mendels G, Levitan S.I, Lee K.Z, Hirschberg J (2017) Hybrid Acoustic-Lexical Deep Learning Approach for Deception Detection., in: INTERSPEECH. pp. 1472–1476.

[Mohan and Seal 2021] Mohan K, Seal A (2021) Deception Detection on "Bag-of-Lies": Integration of Multi-modal Data Using Machine Learning Algorithms. In: Prateek M., Singh T.P., Choudhury T., Pandey H.M., Gia Nhu N. (eds) Proceedings of International Conference on Machine Intelligence and Data Science Applications. Algorithms for Intelligent Systems. Springer, Singapore. https://doi.org/10.1007/978-981-33-4087-9$_3$8.

[Nasriet al. 2016] Nasri H, Ouarda W, Alimi A.M (2016) ReLiDSS: Novel lie detection system from speech signal, in: 2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA). IEEE, pp. 1–8.

[Patel et al. 2016] Patel D, Hong X, Zhao G (2016) Selective deep features for micro-expression recognition, in: 2016 23rd International Conference on Pattern Recognition (ICPR). IEEE, pp. 2258–2263.

[Perez-Rosas et al. 2015a] Pérez-Rosas V, Abouelenien M, Mihalcea R, Burzo M (2015) Deception detection using real-life trial data, in: Proceedings of the 2015 ACM on International Conference on Multimodal Interaction. pp. 59–66.

[Perez-Rosas et al. 2015b] Pérez-Rosas V, Abouelenien M, Mihalcea R, Xiao Y, Linton C.J, Burzo M (2015) Verbal and nonverbal clues for real-life deception detection, in: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. pp. 2336–2346.

[Perez-Rosas and Mihalcea 2015] Pérez-Rosas V, Mihalcea R (2015) Experiments in open domain deception detection, in: Proceedings of the 2015

Conference on Empirical Methods in Natural Language Processing. pp. 1120–1125.

[Perez et al. 2014] Pérez-Rosas V, Mihalcea R, Narvaez A, Burzo M (2014) A Multimodal Dataset for Deception Detection., in: LREC. pp. 3118–3122.

[Rahman et al. 2019] Rahman Md.M, Shome A, Chellappan S, and Al Islam A (2019) How Smart Your Smartphone Is in Lie Detection?. In 16th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (MobiQuitous), November 12–14, 2019, Houston, TX, USA. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3360774.3360788.

[Rehurek and Sojka 2010] Rehurek R, Sojka P (2010) Software framework for topic modelling with large corpora, in: In Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks. Citeseer.

[Ren and Ji 2019] Ren Y, Ji D (2019) Learning to detect deceptive opinion spam: A survey. IEEE Access 7, 42934–42945.

[Rusconi and Mitchener-Nissen 2013] Rusconi E, Mitchener-Nissen T (2013) Prospects of functional magnetic resonance imaging as lie detector. Frontiers in Human Neuroscience, 7

[Saini et al. 2021] Saini N, Bhardwaj S, Agarwal R, Chandra S (2021), "Information Detection in Brain Using Wavelet Features and K-Nearest Neighbor", Communication and Electronics Systems (ICCES) 2021 6th International Conference on, pp. 1704-1709, 2021.

[Sen et al. 2018] Sen T, Hasan M.K, Teicher Z, Hoque M.E (2018) Automated dyadic data recorder (ADDR) framework and analysis of facial cues in deceptive communication. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 1, 1–22.

[Strapparava and Mihalcea 2009] Strapparava C and Mihalcea R (2009) The Lie Detector: Explorations in the Automatic Recognition of Deceptive Language. Proceedings of the ACL-IJCNLP 2009 Conference Short Papers, Singapore, 4 August 2009, 309-312.

[Stromwall et al. 2004] Stromwall L, Granhag P.A, Hartwig M (2004) 10 Practitioners' beliefs about deception. The detection of deception in forensic contexts 229.

[Smith et al. 2008] Smith P.K, Mahdavi J, Carvalho M, Fisher S, Russell S, Tippett N (2008) Cyberbullying: Its Nature and Impact in Secondary School Pupils. Journal of Child Psychology and Psychiatry, 49, 376-385.

[Turnip et al. 2017] Turnip A, Amri M.F, Fakrurroja H, Simbolon A.I, Suhendra M.A, Kusumandari D.E (2017) Deception detection of eeg-p300 component classified by svm method, in: Proceedings of the 6th International Conference on Software and Computer Applications. pp. 299–303.

[Upadhayay and Behzadan 2020] Upadhayay B, Behzadan V, "Sentimental LIAR: Extended Corpus and Deep Learning Models for Fake Claim Classification," 2020 IEEE International Conference on Intelligence and Security Informatics (ISI), 2020, pp. 1-6, DOI: 10.1109/ISI49825.2020.9280528.

[Venkatesh et al. 2019] Venkatesh S, Ramachandra R, Bours P (2019) Video Based Deception Detection Using Deep Recurrent Convolutional Neural Network, in: International Conference on Computer Vision and Image Processing. Springer, pp. 163–169.

[Verburg and Menkovski 2019] Verburg M, Menkovski V (2019) Micro-expression detection in long videos using optical flow and recurrent neural networks, in: 2019 14th IEEE International Conference on Automatic Face Gesture Recognition (FG 2019). IEEE, pp. 1–6.

[Vrij et al. 2018] Vrij A, Leal S, Fisher R.P (2018) Verbal deception and the model statement as a lie detection tool. Frontiers in psychiatry 9, 492.

[Wang et al. 2018] Wang S.J, Li B.J, Liu Y.J, Yan W.J, Ou X, Huang X, Xu F, Fu X (2018) Micro-expression recognition with small sample size by transferring long-term convolutional neural network. Neurocomputing 312, 251–262.

[Warkentin et al. 2010] Warkentin D, Woodworth M, Hancock J.T, Cormier N (2010) Warrants and

deception in computer mediated communication, in: Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work. pp. 9–12.

[Ekman 2009] Ekman P (2009) Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage. New York: Norton. Taylor Francis.

[Wang 2017] Wang W.Y (2017) "liar liar pants on fire": A new benchmark dataset for fake news detection. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). DOI:10.18653/v1/P17-2067

[Wu et al. 2018] Wu Z, Singh B, Davis L.S, Subrahmanian V.S (2018) Deception detection in videos, in: Thirty-Second AAAI Conference on Artificial Intelligence.

[Xie et al. 2018] Xie Y, Liang R, Tao H, Zhu Y, Zhao L (2018) Convolutional bidirectional long short-term memory for deception detection with acoustic features. IEEE Access 6, 76527–76534.

[Xiao et al. 2021a] H. Xiao, K. Yi, R. Peng and G. Kou (2021) Reliability of a Distributed Computing System With Performance Sharing. IEEE Transactions on Reliability, doi: 10.1109/TR.2021.3111031.

[Xiao et al. 2021b] H. Xiao, Y. Yan, G. Kou and S. Wu (2021) Optimal Inspection Policy for a Single-Unit System Considering Two Failure Modes and Production Wait Time. IEEE Transactions on Reliability, doi: 10.1109/TR.2021.3125963.

[Lie et al. 2020] Tie Li, Gang Kou, Yi Peng (2020) Improving malicious URLs detection via feature engineering: Linear and nonlinear space transformation methods. Information Systems 91.

[Yan et al. 2014] Yan W.J, Li X, Wang S.J, Zhao G, Liu Y.J, Chen Y.H, Fu X (2014) CASME II: An improved spontaneous micro-expression database and the baseline evaluation. PloS one 9, e86041.

[Yan et al. 2013a] Yan W.J, Wu Q, Liang J, Chen Y.H, Fu X (2013) How fast are the leaked facial expressions: The duration of micro-expressions. Journal of Nonverbal Behavior 37, 217–230.

[Yanet al. 2013b] Yan W.J, Wu Q, Liu Y.J, Wang S.J, Fu X (2013) CASME database: a dataset of spontaneous micro-expressions collected from neutralized faces, in: 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG). IEEE, pp. 1–7.

[Yancheva and Rudzicz 2013] Yancheva M, Rudzicz F (2013) Automatic detection of deception in child-produced speech using syntactic complexity features, in: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 944–953.

[Zhao et al. 2018] Zhao S, Xu Z, Liu L, Guo M, Yun J (2018) Towards accurate deceptive opinions detection based on word order-preserving CNN. Mathematical Problems in Engineering 2018.

[Zhou et al. 2015a] Zhou Y, Zhao H, Pan X (2015) Lie Detection from Speech Analysis Based on K–SVD Deep Belief Network Model, in: International Conference on Intelligent Computing. Springer, pp. 189–196.

[Zhou et al. 2015b] Zhou Y, Zhao H, Pan X, Shang L (2015) Deception detecting from speech signal using relevance vector machine and nonlinear dynamics features. Neurocomputing 151, 1042–1052.

[Zhou et al. 2017] Zhou Y, Zhao H, Shang L (2017) Lying Speech Characteristic Extraction Based on SSAE Deep Learning Model, in: International Conference on Intelligent Computing. Springer, pp. 672–681.

[Zimmerman 2016] Zimmerman L (2016) Deception detection [WWW Document]. https://www.apa.org. URL https://www.apa.org/monitor/2016/03/deception (accessed 11.4.20).

[Zhang and Kou 2022] Hegui Zhang, Gang Kou (2022) Role-based Multiplex Network Embedding. 39th International Conference on Machine Learning, PMLR 162:26265-26280, 2022