International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 14, No.1 (Dec-23)

http://dx.doi.org/10.12785/ijcds/1401119

A Comprehensive Study of Various Modalities Used for Deception Detection

Maheta Hardikkumar Harishbhai[1] and Dr. Dineshkumar B. Vaghela[2]

[1]*Research Scholar, Gujarat Technological University, Ahmedabad, Gujarat, India*
[2]*Gujarat Technological University, Ahmedabad, Gujarat, India*

Received 30 Jan. 2023, Revised 16 Sep. 2023, Accepted 31 Oct. 2023, Published 28 Dec. 2023

Abstract: Deception detection is an investigative method to determine if someone is telling the truth or fabricating information. It has attracted a lot of study interest because of its potential to be helpful in a variety of real-life problems, including healthcare, law enforcement, internet fraud, criminal investigation, and national security systems. Conventional methods such as the polygraph, demeanor observation, electroencephalogram, and functional magnetic resonance imaging (fMRI) are available to detect deception. These methods are unreliable because they require human interaction and training. They are also time-consuming and costly. Therefore, researchers developed machine learning-driven algorithms to remove human dependency. They have explored thermal imaging, acoustic analysis, eye tracking, facial micro-expression processing, and linguistic analysis to detect deception using machine learning. These techniques may produce better results because they are human-independent and unaffected by race or ethnicity. One can achieve a more reliable automatic deception detection system using features from multiple modalities. This study investigates the feasibility of using linguistic, speech, thermal, and video modalities for automatic deception detection. This paper intends to present a detailed analysis of various deception detection data sets, modalities, and possible directions for the field's development in the future.

Keywords: Deception detection, Facial micro-expression, Acoustic analysis, Brain activity mapping, Linguistic analysis, Eye movement monitoring

1. INTRODUCTION

Deception is an intentionally successful or unsuccessful attempt to mislead others [1]. It happens by providing inaccurate or misleading information and leaving out pertinent details [1]. In most interactions between people, there is some element of deception, sometimes with disastrous outcomes. A deceptive act can range in severity from a minor transgression to a substantial security concern. Real-life deception scenarios fall into two main categories. The first scenario involves a high-stakes deception scenario, such as in courtroom proceedings, police interrogation, or airport screening, where the user's deception has a significant impact on a decision [2]. The second scenario involves low-stakes deception detection, where a user's deceptive behavior indirectly affects many individuals, such as fake customer reviews affecting customers' purchasing decisions, contents of social media sites, job interviews, and counseling. A literature survey states that a human's ability without sophisticated instruments is slightly better than a chance to spot deception [1], [3]. Hence, automatic deception detection is increasingly in demand in fields such as law enforcement, criminal investigations, healthcare, and national security [4], [5]. One of the pioneering approaches to deception detection employed physiological sensors like polygraph methods that extract physiological parameters including body temp, pulse rate, breathing rates, and galvanic skin response [4]. A decision about deceptive behaviors is taken based on data provided by these physiological sensors combined with human expertise [6]. In addition, trained experts analyzed facial patterns and body language clues frame-by-frame [7]. The two methods described above require the input of human experts, whose judgment is biased and whose training is costly and excessive [3], [8]. The data-driven methodologies and machine learning (ML) have developed many novel deception detection techniques. Researchers examined psychological, physiological, visual, linguistic, auditory, and thermal modalities to uncover discriminative features to recognize deceptive behavior. There are physiological features-based techniques for detecting deception, such as polygraphs, electroencephalograms (EEGs), and functional magnetic resonance imaging (fMRI). The issue with these techniques is that they necessitate expensive gear, a proper setting, and human expertise. They are invasive, and the participant must consent to have electrodes and other monitoring equipment affixed to their body [9], [10]. Verbal (acoustic) non-linguistic features like pitch, speaking rate, vocal sound, and voice stress analysis are good indicators to identify deceit using var-

E-mail address: phd.hardikmaheta@gmail.com, dineshvaghela28@gmail.com

ious machine learning-based techniques [11]. Researchers employed thermal imaging to detect variations in blood outpours in a specific facial area to find thermal facial patterns for deception detection [12], [13]. The linguistic analysis employs various linguistic features extracted from a language, such as word usage, word count analysis, and the mean length of sentences to identify deceptive behavior [14], [15]. Facial expression analysis and eye interaction are two vision-based techniques used in addition to the methods mentioned above to identify deception. The geometric facial features and facial micro-expressions are considered distinctive features for deception detection [2], [16], [17]. Facial emotion analysis plays a crucial role in identifying deceptive behavior. Facial Action Unit (FAU) refers to a set of facial muscle movements that correspond to a specific emotion and is used to identify deception [18], [19], [20]. Eye movement, gaze direction, size of the pupil, and eye blink frequency have been used as features in various machine learning-based approaches to identify deceptive behavior [18], [21], [22]. Multi-modal analysis has recently gained a significant amount of interest due to its improved performance compared to individual modality analysis [10], [14], [23]. In multi-modal analysis, one must select the most significant features from each modality and eliminate any irrelevant features to be effective. In this paper, we review all the methods described above in detail.

We structure this paper as follows: We explored the features of each deception detection modality in Section 2. In section 3, we addressed the various famous datasets used in mainstream scientific research studies. We analyzed different machine learning models used for deception detection in section 4. We discussed challenges, future scope, and direction in deception detection in section 5. We concluded this paper in section 6.

2. DECEPTION DETECTION MODALITIES

A. *Deception detection Using Physiological Measurements*

Physiological clues are crucial for detecting changes in human behavior. Larson et al. developed the Polygraph method in 1932 [24]. A polygraph instrument is widely utilized equipment to spot deception. It evaluates physiological alterations brought on by stress in a person's body. It captures physiological changes such as pulse rate, breathing rate, and galvanic skin response through sensors connected to diverse body parts [4]. Psychologists evaluate whether someone is truthful or deceiving based on information derived from a polygraph instrument. They have also utilized brain waves collected using MRI scanners as a deception indicator [25], [26]. When the individual shifted from truthful responses to deceptive responses, there were variations in the results from the MRI sensors. EEG (electroencephalogram) records the electrical potential from every brain neuron. EEG helps to analyze the internal responses of brain signals [27]. The thinking portion of the brain is more engaged when a person is deceptive. The evaluation of the part of the brain that assists in cognitive activities is crucial to understand the deceptive behavior of the human being. The strategies discussed above have several faults, including the ability to falsely accuse innocent people of crimes and acquit those guilty persons [28], [29]. Suspects can take control of their physiological signals and influence the results if they use the proper countermeasures [1]. The scientific community made improvements in polygraph tests by using the Guilty Knowledge Test (GKT) instead of the Control Question Test (CQT) [30]. GKT is a multiple-choice question bank that aims to uncover any hidden truth that an accused tries to cover up. One must collect a sizable amount of background knowledge about the subject to develop a robust set of control questions in a CQT before the test. All the techniques mentioned here use sensors and are invasive contact-based techniques, making it impossible to conduct covert operations. An individual's behavior can change during the deception detection test to decrease the accuracy of these techniques. The techniques based on physiological clues necessitate expensive technology, a suitable environment, and human experts. Therefore, these techniques are highly challenging to implement on a large scale in the public domain, such as inside airports.

B. *Deception detection Using Eye related features*

Studies show that the eye represents the actual state of any human being. During a deceptive scenario, cognitive load increases on the person. The change in the eye-related features such as eye blink frequency, eye gaze pattern, and pupil dilation indicates increases in cognitive load on the person [31]. Numerous research studies have shown that oculometric behavior provides excellent accuracy for detecting deceptive scenarios [10], [18], [21], [22]. In deception detection research, eye gaze proved to be a valuable and distinguishing feature [21], [22], [32], [33]. There is less eye movement in deceptive persons compared to truth-tellers [22], [33]. Vrij et al. explored that deception causes more long-term memory searches compared to truth [34]. In the findings, They observed that fewer saccades are there in truth-telling compared to deceptive speeches. Various research studies indicate that eye gaze dwell time and eye gaze fixation are valuable distinctive features for deception detection [33], [35], [36]. Eye gaze dwell time is more for deceivers than truth-tellers [35]. The eye blink frequency and eye blink intervals are also used in many kinds of research to identify deceptive behavior [22], [37], [38]. Borza et al. analyzed eye gaze and eye blink intervals as features for deception detection [22]. They developed a new metric called NBRD (Normalized Blink Rate Deviation). The normalized average blink rate of the person is called NBRD. They showed that NBRD was a distinctive feature for deception detection. According to earlier research, when people are deceptive, their frequency of eye blinks decreases, but when they are telling the truth, the frequency and rate of eye blink remain unchanged [38], [39]. Some research also showed that once the deceptive scenario is over, there is a sudden increase in eye blink rate to counter the cognitive load that a person has to feel during a deceptive scenario [31]. Some research also suggests that the eye blink rate increases during deception [37]. Some studies show eye blink rate depends on the type of interviewer. In the case of the avatar-based interview, there are minor changes in the eye blink rate in the deceptive scenario [40]. According to certain research findings, while responding to inexplicable questions, truthful people respond more slowly than deceptive people. Long response time increases their blink rate [41]. Researchers used pupil diameter in various research to identify deceptive behavior [22], [33], [42], [43]. Pupil dilation usually indicates an increase in cognitive load. Liars typically experience increases in cognitive load. Pupil dilation is more when lying compared to truth-telling [42], [43]. Proudfoot et al. show that pupil dilation decreases over time after the completion of deceptive answers [35]. They have used latent growth curve modeling on pupil size to see the slope in a change of pupil dilation. Fang et al. observed that eye-tracking indicators of deception are helpful for both instructed and spontaneous deception detection. They observed pupil size and gaze behaviors are significant features to identify lies during surveys. They found that respondents have larger pupil sizes while lying [33]. Seymour et al. claim that a combination of eye-related features, such as pupil dilation rate, eye blink rate, and eye gaze, offers greater accuracy in contrast to employing an individual eye-related feature [42]. Noman et al. employed a template-based technique to catch the eye blinking from a real-time mobile video frame [44]. They used an image dataset of an open and closed eye to correlate it with the current eye image. They used correlation coefficient value for template matching to decide a blink or non-blink image of the eye. Soukupova and Cech employed the eye-aspect ratio (EAR). In an eye image, the EAR indicates the proportion between the eye height and the eye width. They used eye landmarks to find the eye-aspect ratio (EAR). The EAR that is near zero indicates a blink of an eye. Borza et al. used a combination of the EAR and a CNN to find an eye blink in the video frame. They achieved good accuracy by removing false positive results. Eye gaze is the angle between a person's eye center and the pupil or iris center. Nurcin et al. have segmented the pupil and iris from the image of the eye using image processing techniques. They used otsu's thresholding and hough transform method to segment the pupil and iris. They considered the ratio of pupil to iris as a feature in their research. So, classification is not affected by distance variations of eye images. Schuetzler discussed various countermeasures that a deceiver can perform to counter the accuracy of deception detection through eye-tracking technologies [45].

C. Deception detection Using Thermal Modality

The changes in blood circulation because of the nervous system response can be detected using thermal image analysis. The thermal image analysis approach is considered an alternative approach that removes human dependency and the invasive character of a polygraph. Initially, Pavlidis et al. applied thermodynamic modeling on thermal images [46]. Their method involves transforming periorbital thermal data into blood circulation rate to detect deception. They perform analysis on the periorbital facial area of the subject. They indicated in their research that the periorbital facial area is affected largely by the redistribution of blood circulation in stressed conditions. They achieved up to 78% classification accuracy in an extremely challenging mock crime dataset using the K-nearest neighbor (KNN) classifier. They discovered that when people act deceptively, there is an increased influx of blood around the eyes. Tsiamyrtzis et al. acquired thermal characteristics from the periorbital area of the face while allowing the subjects' face movements to recognize deception. They did this by using continuous imaging and noise removal techniques. [47]. Pollina et al. collected the lowest and highest heat values from video frames of persons in various stages of deception and honesty [48]. They concentrated on the region around the eyes of the face. They discovered a substantial change in the temperature of a layer of the skin between the two states. They have used two different concealed information tests. Jain et al. combined a thermal camera and facial landmark detection algorithms to recognize and follow landmarks in the regions of focus in the face [49]. They estimated the average temperature of the 10% hottest pixels to discriminate between deceptive and truthfulness. Warmerlink et al. tested the efficacy of facial thermal image processing on 51 travelers in an international airport departure [50]. They found that during the interview, the body temp of the deceptive people raised while it stayed the same for the truthful people. This physiological and technological evidence sufficiently supports the relationship between temperature changes and anxiety levels. However, previous studies have reported inconsistent and contradictory classification accuracy rates. Bashar et al. utilized thermal image processing to monitor the thermal fluctuations of the facial periorbital region [12]. The author employed 492 thermal responses (249 deceptive and 243 truthful) from 25 subjects. They employed a high-dimensional feature vector collected from each subregion of the periorbital region and a KNN classification approach. They discovered that, in addition to deceit, other factors such as facial expressions, body synergy, variation in underlying muscle-related thermal activity, the ambient environment's thermal fluxes, and disease modify skin surface temperature. As a result, it's critical to account for such impacts by considering each individual's varying initial baseline temperatures. Based on a within-person technique, they declared an 87 percent capability to anticipate lie-truth responses. However, their outcomes demonstrate that the between-person deception modeling approach does not readily adapt to the training data. Abouelenien et al. the maximum, minimum, mean, and range of the maximum to minimum temperature value of a face [51]. After that, they produce a facial region histogram with 120 bins all over the image pixels. They got a total of 124 thermal properties for each image and took 200 frames from each video clip for sampling. They averaged the obtained features for each participant over these many frames. When these statistics and the histogram are combined, they result in a detailed heat map. The heat map displays the distribution of heat across the face. They tried to spot differences in this heat map whenever one of the participants tried to deceive them. They used the first 50 seconds of each video clip as the baseline for each participant to address the issue of individual variances in base temperature, ensuring that everyone received the same treatment. Each participant sat in a chair for a fixed period of the session, doing nothing or responding verbally. The same set of 124 features was collected and averaged during this period. This normalization method determines whether there is a movement in the thermal map when responding deceptively regardless of inter-personal temperature. Park et al. found that functional discriminant analysis revealed that responding to questions related to crime items can be used to discriminate victims from innocent participants [52]. The results of their analysis indicate that the measurement of temperatures in the periorbital region of the face while performing the CIT may identify a deceptive person during a culpable condition. Derakhshan et al. have collected a dataset by measuring the facial thermography of 41 subjects in two different procedures (mock crime and best friend) [53]. They have concentrated on five parts of the face, namely the chin, periorbital, forehead, cheeks, and perinasal. They tracked the head movement using corner points of the face. They have used six statistical techniques and four feature reduction strategies to determine the optimal features from the most promising ROI. They have used Four different classifiers, including SVM, Linear Discriminant Analysis (LDA), KNN, and Decision Tree. They deduced that the body's perinasal and chin regions are the most strongly connected to deceptive anxiety. Hence, They could be an effective sign of deception.

D. Deception detection Using Linguistic Modality

Numerous research studies have investigated detecting deception using linguistic features from the textual content [54]. Many research studies have shown a strong correlation between the deceptive behavior of a person and his linguistic preferences [54], [55]. Newman et al. have identified that deceivers displayed less cognitive complexity, made fewer self- and other-references, and expressed higher negativity in words compared to truth-tellers [54]. Bachenko et al. selected twelve linguistic features of deception [56]. They have taken phrase tense variation, negative expressions, conflicting verb, and noun forms, and statements that lack commitment. They have utilized a textual repository of court testimonials, police interrogations, and criminal comments to extract and examine the effect of these indications on deception. Hauch et al. performed a meta-analysis that included 44 research and 79 verbal deception cues [57]. Their meta-analyses showed that liars had a higher mental workload, displayed higher negative emotions, released themselves from activities, expressed fewer feelings and words, and referred to cognitive operations less frequently than truth-tellers. Zhou et al. used 27 different linguistic clues and clustered them into nine different linguistic constructs [58]. They discovered that deceivers used more phrases, verbs, noun words, and statements in their research. The communication of deceivers was more expressive than truth-tellers and appeared more casual with typographical errors. They found that deceitful participants had less lexical and content diversity than truth-tellers. They have considered linguistic features such as diversity, complexity, expressiveness, and informality in sentences. They discovered that the above linguistic constructs are useful in distinguishing between honest and deceptive communications. Toma et al. conducted linguistic analyses using LIWC of online dating profiles and discovered many self-references, negations, and lower degrees of word usage in fake profiles [59]. Yancheva et al. have examined the relation between speech syntactic complexity and the age of children [60]. In their research, they examined many linguistic features to assess changes in the complexity of a child's language, whether lying versus telling the truth. They examined the fluency index of verbal responses, sentence complexity based on T-unit research, and the use of passive expressions. The results revealed a direct link between the complexity of deceptive speech and the age of the children. Text analysis features commonly contain fundamental linguistic representations such as n-gram models and word frequency data. The semantic features are derived using the LIWC dictionary. Syntactic CFG trees and Part-of-Speech (POS) tags derive complicated linguistic syntactic features. Text analysis, semantic, and syntactic features are considered useful in research studies to detect deception [61], [62]. Some studies used the Linguistic Inquiry and Word Count (LIWC) lexicon to develop deception models using ML techniques and found that utilizing cognitive science information was effective for the automatic detection of deception [61], [63], [64]. Several studies have investigated the relationship between the syntactic complexity of a text and deception [60]. They postulated that deceivers might produce simpler words to hide the truth and have a quicker memory of their lies. Gogate et al. have used CNN architecture applied to the transcript to detect deception [65]. In their research, Words are transformed into vectors using a 300-dimensional Global Vectors for Word Representation (GloVe) trained on 840 billion phrases from authoritative web crawling. Then, using concatenated word representations, They constructed CNN architecture. Krishnamurthy et al. used a Word2Vec [66] model to obtain the vector embeddings for every phrase in the text data [15]. Then, these vectors are merged and supplied to the CNN as an input vector. The deceit detection problem has also been the subject of numerous efforts to explore it in languages other than English. Almela et al. used an SVM with linguistic classes from the Spanish version of the LIWC dictionary to handle the deception detection task in Spanish essays [64]. Fornaciari et al. investigated deception in Italian court proceedings [67]. They have examined various methods for detecting deceptive clues, including utterance duration, LIWC features, lemmas, and POS patterns. Perez-Rosas et al. presented a study to identify the cultural differences between deceitful and genuine articles documented in English, Spanish, and Romanian languages [68]. The authors developed classifiers specifically for each culture and then conducted numerous trials across cultures to address the challenge of deception detection. The findings indicate major cultural differences and the likelihood of using semantic data as a link when misleading information for a specific language is not readily available. They discovered that all cultures share basic deceptive patterns, including affirmation usage, negative emotions, and references to others. There are extensive public corpora accessible for linguistic study in the English language. Mihalcea et al. compiled a database of deceptive and truthful essays [61]. Ott et al. gathered data on fictitious hotel reviews from Trip Advisor [69]. Li et al. used Mechanical Turk to collect a fake product review dataset [70]. A few deception datasets are also available for languages other than English. The German deception corpus contains purchased item reviews [71]. The Spanish and Romanian essay dataset [51] contains perspectives on various themes like the death penalty and abortion. Overall, the n-grams technique and word statistics like phrase size, word type proportion, and word variety utilized for deception detection frequently. The grammatical configuration of a text is useful in identifying syntactical linguistic forms associated with deceit. Semantic data has also proven to be a useful resource for comprehending knowledge concerning the thought patterns of the deceiver. LIWC and Wordnet are useful resources for analyzing the word usage of deceivers in this category.

E. *Deception detection Using Acoustic Modality*

Numerous research in the literature has employed acoustic features to identify deception [72]. The acoustic features provide a non-intrusive experience to detect deception. Acoustic modality-based analysis systems are inexpensive to build, simple to use, installable on portable devices, and used at any time to examine voice recordings. Researchers have studied various methods to analyze and identify the characteristics and properties of the audio signal to detect deception for a long time. Spectral energy characteristics and Cepstral features are the two main categories of speech signal features used to detect falsehood in speech. The development of the Spectral energy characteristics takes advantage of the psychoacoustic masking property of how humans hear speech transmissions [73]. The irrelevant speech conveys information that the human ear usually misses. The psychoacoustic masking properties identify irrelevant speech signal information. The Mel Frequency Cepstral Coefficients (MFCC) elements are the basis for the foundation of cepstral features. One can understand human sentiment and deception through speech using cepstral features [74]. Increased loudness, shorter speech time, and higher fundamental frequency are among the cepstral features [73], [74]. Srivastava and Dubey employed fundamental frequency (F0), Zero Crossing Rate (ZCR), and energy features to detect deception from the voice signal data from an interview they performed in a remote setting [75]. They used SVM and ANN Classifiers in their research work. Ullah and Gopalan retrieved the Bark energy and other energy features from stressful speech signals to detect deception from a criminal questioning database [76]. Desai et al. used ZCR, energy features, the entropy of energy, Spectral Flux (SF), Spectral Roll-off (SR), Chroma Vector (CV), Chroma Deviation (CD), and MFCC feature to detect deception from the audio signal [77]. They have used Columbia University SRI-Colorado State University (CSC) Deceptive Speech Dataset [78]. They used the Recurrent Neural Network classifier and got 62.59% accuracy from the audio modality. Tao et al. utilized the F0, ZCR, and energy features to explore how it concerns deception detection [79]. They have used the IDIAP WOLF data set from the Swiss Research Institute [80]. The experimental outcomes on the dataset show that the recognition accuracy can increase to more than 80% using the SVM classifier. Xie et al. have extracted frame-level acoustic features. They have used ZCR, a Root Mean Square (RMS) of the frame energy, F0, a center of gravity and the second central moment of the spectrum, the MFCCs, and its first-order and second-order delta and linear prediction coefficients as feature set [81]. A frame represents the temporal information in the speech waveform that brought all the elements together. Each frame contains 60-dimensional data. The author employed the convolutional bidirectional LSTM model on Columbia–SRI–Colorado dataset [78]. They achieved 70.3% accuracy. Fan et al. have extracted Short Time Energy, Pitch, Formant, and Duration feature from speech signals to detect deception [82]. They used the Chinese Deception Detection dataset based on the Chinese language [82]. They used a variety of classifiers with transfer learning approaches to identify cross-gender deception. Fernandes and Ullah have extracted four distinct types of unique features for deception detection [83]. They retrieved cepstral features and spectral energy features. They applied the Levenberg-Marquardt classification method and the LSTM classification method in their research. They employed PCA to reduce the dimensionality of the retrieved features for improvement. LSTM classification method with time-difference spectral energy features following the PCA shows the highest accuracy rate. Zhou et al. have Extracted prosodic and non-linear dynamics (NLD) feature sets from the voice signal and applied the relevance vector machine (RVM) classification technique [84]. They used the deception corpus of Soochow University. They used 30 prosodic and 18 NLD features with the RVM Classification model built on sparse Bayesian learning. RVM technology is more reliable than the SVM algorithm and requires significantly fewer functions and decision-making time. They used prosodic features such as pitch frequency, short-term energy, and MFCC features to represent the static characteristics of a speech signal. Deceptive speech mentions non-laminar flow, flow separation in different regions, the creation and propagation of vortices, and the formation of jets. They extracted a set of 18 NLD features that contain fractal properties, kolmogorov, and lyapunov exponents entropy by the Open TSTTOOL toolbox available for MATLAB. Xue et al. used MFCC, energy features, and pitch contours feature set for deception detection [85]. They generated the above features from a balanced dataset of deceptive and non-deceptive speech recordings collected from a two-person lying game. They created a model using a majority-voting ensemble learning classifier constructed from a Gradient Boosting Classifier (GBC), SVM, and Stochastic Gradient Descent (SGD) trained on MFCC and energy features. They achieved a maximum accuracy of 55.8% for lie detection using this model. Jaiswal et al. have used a feature vector comprised of 28 dimensions, including prosody, energy, voicing probabilities, spectrum, and cepstral features [86]. Intensity, loudness, and pitch are prosodic characteristics that describe the amplitude and frequency of the speech stream. Humans perceive loudness according to energy properties. The voice probabilities represent the estimate of the percentage of vocal and unvoiced energy in the speech. The spectral features are based on spoken content and represent speaker characteristics. Cepstral characteristics highlight variations or recurrences in the features of the spectrum, quantified by frequencies. They used 12 MFCC features and got 34.23% accuracy on the Real Life Trial Deception Detection(RLTDD) dataset [87]. They used the OpenSmile software to extract voice attributes [88]. They use a 28-dimensional feature vector including prosody, energy, voicing probability, spectrum, and cepstral properties. Chebbi and Jebara have focused on investigating different pitch-based features [89]. They noted that liars tend to have a higher range in their pitch voice than truth-tellers. They calculated the pitch values frame by frame using the 'fxrapt' function provided in the 'voicebox' tool. They generated a set of 72 pith-based attributes for each sequence once from the pitch values to investigate how to discriminate between deception and truth. This 72-feature set consists of four feature families. 12 features are the various statistical measures of the pitch. 14 features related to speech voicing. 28 features in total, including their statistical measures for the first and second derivatives of the pitch, and the remaining 18 features related to the pitch. They used RLTDD dataset [87] and got 58% accuracy using audio features. Zhang et al. used Interspeech 2009 (IS09) ComParE Challenge Open-SMILE baseline feature set [88]. This feature set is considered a benchmark for many computational paralinguistic tasks. Before extracting attributes using the OpenSmile software, they used various noise reduction techniques like calculating spectral centroids, MFCC, and Median filtering to remove audience laughter, applause, and background music. The performance after applying all noise reduction techniques is improved. Spectral centroids work best among all three applied noise reduction techniques with 63% accuracy on a Box of Lies corpus [90]. Antoln et al. used Frame-level speech characteristics (log-mel spectrograms) [91]. Additionally, they demonstrated a multi-modal Automatic Deception Detection (ADD) system that integrates the gaze and voice modalities using two fusion algorithms into an attention LSTM architecture. On the Bag-of-Lies dataset [10], they found that the Attention LSTM-based systems perform noticeably better than the conventional SVM method. Using attention LSTM architecture, they achieved an accuracy of 63.8% at the segment level and 61.41% at the turn level, respectively. Garcia et al. have extracted audio features like glottal flow, voice, MCEP, Harmonic model, pitch, rhythm, and phase distortion mean and deviations [92]. They used RLTDD dataset [87]. They got 0.73 and 0.63 AUC using LSTM deep neural network and SVM. Wu et al. have used MFCC as their audio features [93]. They used GMM (Gaussian Mixture Model) to build an audio feature set for training data. They treat all acoustic features equally and then use feature encoding techniques to find hidden clues in audio for deception detection. They got 0.81 AUC on the RLTDD dataset using the MFCC feature and Support Vector Machine (SVM) classifier, [87]. Krishnamurthy et al. have used the OpenSmile software to pull features from the input audio [15]. They performed voice normalization using Z-standardization to eliminate background noise. For each audio input, they extract features with a dimension of 6373 using the IS13-ComparE OpenSmile configuration. Then the obtained feature vector of dimension 6373 is converted into a feature vector of 300 using a fully connected trained neural network. They used RLTDD dataset [87] and got 0.76 AUC using the above audio features using a Linear SVM Classifier. Gupta et al. used various frequency-based properties of audio signals [10]. They used SC, SB, SR, ZCR, CF, and MFCC. They combined these features into a 26-dimensional feature vector. They performed two-class audio classification using RF and KNN classifiers. They used the Bag-of-Lies dataset [10]. They used various classifiers such as KNN, SVM, and RF. They got the highest accuracy of 56.22% based on acoustic features. Venkatesh et al. have used the audio system based on Cepstral Coefficients (CC) and Spectral Regression Kernel Discriminant Analysis (SRKDA) of fixed-length speech sequences to detect deception [94]. SRKDA employs discriminant analysis, prompted by non-linear mapping to process data, stimulated by non-linear mapping. They used RLTDD dataset [87]. They got the Correct Classification Rate(CCR) of 76% using the CR-SRKDA method based on speech modality. Sen et al. used pitch, silence, and speech histogram as feature set for the voice modality [95]. They used the STRAIGHT toolkit to determine the fundamental frequency (F0) of the defendants' speech to estimate pitch. Since only spoken speech segments predict F0, they don't include unvoiced speech frames in their calculations. Then, they generate two features, namely mean F0 and stdev F0. from the raw F0 values. They utilize a voice activity detection (VAD) algorithm to separate the speech and silence segments from the subject's speech to retrieve silence and speech histograms. Histogram plots indicate that deceitful people pause for shorter periods more frequently than truthful people. They used RLTDD dataset [87]. They used three classifiers, SVM, Random Forest (RF), and ANN. They Got the best accuracy of 71.19% using the pitch feature of voice with RF Classifier. Table I summarizes the attributes of acoustic modality, classification accuracy, and databases used in previous literature studies for deception detection.

F. Deception detection Using Facial Clues and Body Language

Various facial clues, such as spontaneous facial expressions, emotions, and visual body language, such as hand gestures, prove to be very effective clues to detect deception [96]. Facial micro-expressions are involuntary expressions of emotion on the face with a duration of not more than 0.5

TABLE I. Acoustic Modality Related Features

Ref	Features	Dataset	Classifier	Maximum Classification Accuracy(%)
[77]	ZCR, Energy, The entropy of energy, SF, SR, CV, CD, MFCC	Columbia SRI-Colorado State University (CSC) dataset [78]	RNN	62.59
[79]	F0, ZCR, and Energy features	Swiss Research Institute IDIAP WOLF dataset [80]	SVM	80
[81]	ZCR, RMS of the frame energy, F0, Center of gravity and Second central moment of the spectrum, MFCC	Columbia SRI-Colorado State University (CSC) dataset [78]	Convolutional Bi-directional LSTM	70.30
[82]	Short Time Energy, Pitch, Formant, and Durations	Chinese Deception Detection corpus [82]	LR, J48, MLP, GBDT, SVM	84.56
[83]	Delta and Time-difference cepstrum, Delta and Time-difference energy	Criminal Interrogation Dataset [83]	Levenberg-Marquardt with PCA, LSTM with PCA	91.70
[84]	Prosodic and NLD features	Deception corpus of Soochow University [84]	RVM	70.37
[86]	Prosody, Energy, Voice probabilities, Spectrum, and Cepstral features	RLTDD [87]	SVM	34.23
[89]	72 pitch based features	RLTDD [87]	KNN	58
[10]	ZCR, SC, SB, SR, CF and MFCC	Bag-of-Lies dataset [10]	KNN, SVM and RF	56.22
[94]	Cepstral Coefficients	RLTDD [87]	SRKDA	76
[95]	Pitch, Silence, and Speech Histogram	RLTDD [87]	SVM, ANN, RF	71.19

seconds [97]. Ekman and Rosenberg originated the Facial Action Coding System (FACS) to systematize the technique to identify facial expressions [98]. FACS delivered a catalog of facial attributes using facial muscle activities. The inner eyebrow raiser, lips opener, cheek raiser, nose wrinkling, chin booster, and eye gaze are various examples of Facial Action Units (FAU). Avola et al. have used (FAU to classify deceptive and truthful behavior from the videos [20]. FAU refers to the tightening or relaxing of one or more facial muscles. It is related to the emotions felt by a person [98]. In this paper, the authors used the Convolutional Experts Constrained Local Model (CE-CLM) model from the OpenFace tool to detect FAU in each frame of each video [99], [100]. The author utilized occurrences and intensities of various FAU along with eye gaze as features for classification. They also tried to find correlation patterns between different AUs in deceitful and truthful behavior. They analyzed the importance of each Facial Action Unit using a p-value test. They noticed interesting differences in deceptive and truthful behavior in the frequencies of several Facial Action Units, such as Lowering the eyebrow, raising the upper lip, pulling the corner of the lips, and Dimpler. They used various classifiers such as Logistic Regression (LR), RF, and SVM. In their research, SVM-RBF gives the highest accuracy of 76.84% on the RLTDD dataset [87]. Su and Levine have used various facial clues to detect deception [101]. They divided the face into nine regions and used the Gabor filter to identify the activity of Various Facial Action Units. They created a feature vector of size 21 (12 primary features + 9 Secondary Features). The primary feature consists of the activity of an individual facial action unit, and the secondary consists of the activity of the combination of multiple facial action units. They mainly focused on facial muscle movements, including eye blink, mouth direction, wrinkle detection, and eyebrow motion. They created a dataset from real-life

YouTube videos involving forensic cases. They collected 324 videos and had 51.23 percent of videos with convicted suspects and 48.77 percent with innocent people. The mean runtime of a video clip is 20 seconds. They used the RF classifier and achieved an accuracy of 76.92%. Zhang et al. analyzed real-time video analysis of visual sequences to detect deception by evaluating previously identified falsehood evidence based on unintentional, seemingly trustworthy facial expressions [102]. The reliable emotions are ones that psychologists claim a sizable portion of the population cannot accurately imitate without having a real internal felt emotion. A collection of FAUs connected to trustworthy expressions is identified based on distance and texture-based attributes. They performed trials for falsehood signs for four emotions, including joy, anger, anxiety, and sadness. In Sadness emotion, they got the highest accuracy. They created a dataset of 344 facial images of twelve participants comprising six men and six women. They extracted these images from the authenticated video acquired by the Center for Unified Biometrics and Sensors (CUBS) for training and testing the procedure. Jaiswal et al. used Facial Action Units (FAU) to classify deceptive and truthful behavior from the videos [86]. They used the OpenFace toolkit [100] to extract FAU from the facial areas. They mainly focused on the Eyebrow, Eye, and Mouth regions. They used a feature vector of 18 dimensions containing different FAU. They got an accuracy of 67.20% employing features from visual modality and SVM classifier on the RLTDD dataset in their experiment [87]. Ahmed et al. have divided each video of the dataset into a chunk of 30 frames [103]. They used CE-CLM deep learning model [99] provided in OpenFace open-source software [100] to detect Facial Action Units from each chunk. FAU derived from the above step is then given to the LSTM deep neural network to achieve a Spatiotemporal relationship among the frames of videos. They used three different datasets, namely, RLTDD dataset [87], Silesian Dataset [96], and Bag-of-Lies Dataset [10]. They achieved good accuracy on an individual dataset, but accuracy decreased when they combined datasets. They got an accuracy of 89.49%, 75.82%, and 67.11% on the RLTDD dataset, Silesian Dataset, and Bag-of-Lies Dataset respectively. Abd et al. used facial expression data using Facial Action Unit [104]. They used a total of 18 FAUs. These are AU1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, 20, 23, 25, 26, 28, and 45. They found the usefulness of each AU by taking its proportion of participation in deceptive and truthful videos. They accumulated data from 102 participants (25 female and 77 male) as video clips. There are 504 clips for deceptive responses and 384 for truthful responses (a total of 888 clips). They concluded that six FAU are the most effective, and they had an immediate effect on the differentiation between liar and truth-teller. Facial emotions are considered a pivoting factor in the deception detection of humans. Various literature studies focused on facial emotions to detect deceptive behavior. Shen et al. identified that the length of fear emotion is shorter in deception compared to truth-telling [105]. In deceptive conversation, the time duration of fear expression from peak to offset is small compared to the truthful scenario. The authors measured the distance between the right eyebrow and the right eye and compared it with the distance between the left eyebrow and the left eye. They analyzed the asymmetry differences between the deceptive and honest situations using the above-calculated values. They noticed facial movements around the eyes were more asymmetric in deception compared to an honest scenario. They used a database of which 32 videos of 16 participants telling lies in 16 videos and the truth in 16 videos. They took all video clips from a high-stake game show. The video clips range in length from 3 seconds to 280 seconds, lasting an average of 56.6 seconds. They used three classifiers RF, KNN, and Bagging using a 10-fold cross-validation procedure and got 86.9033%, 85.1068%, and 86.1482% accuracies. Yang et al. have developed the Emotional Transformation Feature, which contains a combination of emotion transformation count for the basic seven emotions [106]. The emotional Transformation Feature describes the percentage of emotion switched from one emotion to another. They applied various classifiers, such as SVM, DT, RF, and KNN. They considered the Emotional Transformation Feature and got good accuracy only from Visual Modality. They used the RLTDD dataset [87]. They got an accuracy of 65% using ETF. They combined facial, and Hand gestures with ETF to get an accuracy of 87.59%. Mathur et al. have used Facial Affect Valance and Arousal in emotions as features to find deceptive behavior [107]. They hypothesized that deceivers have low valance and high arousal facial affect compared to truthful people. They used other visual features like FAU, Head Pose, and EyeGaze with facial valance and arousal. They used the Boruta Feature selection algorithm [108] to create a feature vector for an early fusion of features collected from multiple modalities. They employed the AffWildNet model trained on the Aff-Wild database [109] to capture facial affect representations of valence and arousal from the videos. They got an accuracy of 86% with a combination of Facial Affect and Visual features on RLTDD dataset [87]. Researchers employed facial muscle movement and hand gestures based on MUMIN coding scheme to detect deception in various research studies. Chebbi and Jebara used a total of 39 binary features vector based on the MUMIN coding scheme [110] from hand gestures and facial cues [87], [111]. Four of the Thirty-nine features are associated with facial emotions. Eight features are associated with hand movement. Nine features are associated with head movement. Three features are associated with the eyebrow motion, five to eyes, four to eye gaze, two to the opening of the mouth, and four to mouth lips. Each feature has a boolean value that indicates whether its activity is present or absent. The relevance of each feature is evaluated based on the Chi-square test. They identified that 3 out of the 5 features are associated with eyes. 3 features out of 4 features associated with mouth lips, 3 features out of 3 associated with the motion of eyebrows, and 2 features out of 2 associated with mouth openness effectively differentiate between deception and truth.Four out of eight and three out of nine features that are related to head and hand motions are revealed as deception signs. None of the features related to general facial emotions and eye gaze is considered a valuable feature. They used the KNN classifier on the RLTDD dataset [87]. They got 94% accuracy when they consider the entire feature set. They considered degradation in accuracy when they used only relevant features. Sen et al. have used MUMIN coding scheme [110] to annotate various facial display and hand gesture features. They used basic facial expressions, eyebrows, eyes contact direction, mouth aperture and lips, and Head and Hand trajectory [95]. They used a RLTDD dataset [87]. They used three different classifiers RF, SVM, and ANN. They got the highest accuracy of 80.7% using ANN. In another approach, They used Facial Action Unit. They extracted 18 different FAU using OpenFace Library [100]. They got an accuracy of 61.58% using the RF Algorithm. They observed that classifiers built with automatic visual features (FAU) perform worse than classifiers built with manual annotations. In another paper, they used facial features, and hand gesture features with two different classifiers RF and DT [87]. They got an accuracy of 70.24% and 76.03% from RF and DT classifiers, respectively on RLTDD dataset [87]. Venkatesh et al. used the same 39 binary features vector discussed above from hand gestures, and facial cues [94]. They utilized AdaBoost classifier by considering its excellent performance on the binary classification. They used a RLTDD dataset [87]. They used 10-fold cross-validation in their experiment and got 88% accuracy. Wu et al. used IDT (Improved Dense Trajectory) feature to identify facial action movement [93]. IDT computes local feature correspondences in consecutive frames. They used MBH (Motion Boundry Histogram) to find changes in motion rather than a constant motion to identify facial microexpressions. They got an AUC of 0.83 with an SVM classifier using a combination of IDT, and MicroExpression features on RLTDD dataset [87]. Owayjan et al. tried to find involuntary facial expressions from videos [17]. Facial Micro-expressions are the emotions that are involuntary facial expressions. The duration of microexpression is not more than 0.5 seconds [98]. In their experiment, they try to find the emotions in each video frame using distance measurements of various facial parts. If any emotion occurs for a very small number of frames between other emotions then that is considered a facial micro expression. They hypothesize that if more facial micro-expressions are in the video, then there is more probability of deception in videos. Khan et al. used nonverbal behavior from video interviews to detect deception [18]. They used 36 different fine-grained facial and eye-related micro-features (15 faces, 16 eyes, and 5 face angles). They used the Silent Talker system [112] to extract micro features. They evaluated the uniqueness of facial micromovements concerning deceitful behavior. They used different classifiers such as ANN, SVM, and RF. They got a maximum of 80% accuracy using the RF algorithm on their deception database. Researchers used hand movement and activity as feature sets to detect deception in several research studies. Avola et al. have used Hand movement to detect deception [113]. They employed hand skeleton features retrieved from RGB videos using OpenPose software [114]. There is a total of 21 different points of hand that can be located using OpenPose software. They used the velocity and acceleration of each point for the feature vector. They also find Hand Openness and Hand Elasticity as a feature. After that, they created a feature vector of 29 features. They converted this feature vector into a Fisher vector for better encoding. Finally, They employed LSTM to exploit the Feature Vector representation of deception cues. They got 90.96% accuracy on RLTDD dataset [87]. Randhavane et al. have tried to identify deceptive behavior from features like gait and body gestures [115]. Human posture and movement describe an individual's affective state. So, They used velocity and acceleration of hand and feet joints and other points as features. They used their dataset DeceptiveWalk for the training of the classifier. They employed a variety of motions as gesture features, including looking about, touching faces, touching shirts or blazers, touching hair, folding their hands, and looking at phones. They used LSTM deep learning network as a classifier to learn both temporal and movement patterns of each walk and based on they trained the classifier and got a good result of 88.41% accuracy on their dataset. Avola et al. used an RGB video sequence [116]. In their approach, the author used the OpenFace toolkit [100] to detect, align, and mask the subject's face. They employed histograms of oriented gradients (HOG) and local binary pattern (LBP) methods to obtain a dense face representation. They used improved dense trajectories (IDT) to identify trajectories of facial expressions. They employed local binary patterns from three orthogonal planes (LBP-TOP) to obtain a spatiotemporal characterization of facial muscles. They used Oriented fast and Rotated Brief (ORB) to get interest points in the relevant face area. They generated a signature of the entire video using the Fisher vector (FV) encoding utilizing the above-extracted features. Then, they input the video signature to several base-level algorithms, such as LR, SVM, XGBoost, DT, and MLP. They created a reliable meta-level classifier by concatenating the various base-level predictions and feeding them as input to a second MLP, which effectively uses stacked generalization. They accomplished experiments on two different datasets. (1) RLTDD dataset [87] and (2) Bag-of-Lies Dataset [10]. Using the Stack Generalization method, they got the best accuracy when they used all feature sets with the SVM classifier. In many research studies, people have used CNN to extract features for deception detection. Ding et al. emphasized that Both face and body contain valuable insights to check whether a subject is lying or not [117]. They tried to combine facial hints and body motions through a novel face-focused cross-stream network (FFCSN) deep learning-based architecture. They used association learning across the spatial and temporal channels for combined-deep features from facial expressions and body movements. They got the highest accuracy of 93.16% on the RLTDD datasett. Krishnamurthy et al. have used a combination of multiple modalities to detect deception [15]. They used two approaches (1) 3D-CNN and (2) Facial micro expressions such as a smile, an eyebrow raise, a frown, etc to extract facial features. The facial microexpressions are binary features, and the authors considered them as a feature vector of dimension 39. They have considered manually annotated facial microexpressions. 3D-CNN extracts static features from each image. It identifies Spatio-temporal features from the entire video that aid in locating the subtle face cues. They used the RLTDD dataset for their experiment [87]. They used multilayer perceptron(MLP) and its variant as a classifier. They got an accuracy of 93.08 using visual features and 76.19% using microexpression features. Venkatesh et al. introduced a unique Deep Recurrent CNN for automatic deception detection [118]. The proposed approach relies on the sequential frame-by-frame video input to the deep neural network pre-trained by transfer learning. The suggested method automatically extracts non-verbal facial features using the deep learning approach from a series of video frames using pre-trained CNN from GoogleNet [119]. GoogleNet performs well in object classification and action detection tasks. They connected bi-directional LSTM before the last dropout layer of GoolgeNet. LSTM removes the problem of vanishing gradients by learning and remembering long sequences to identify deception using non-verbal facial features. They used end-to-end learning procedures that train CNN and LSTM components simultaneously. They used a leave-one-out cross-validation method with 25 attempts on the RLTDD dataset [87]. They got 100% classification accuracy using visual features. They got better performance than different unimodal and multimodal approaches. They have shown that deep learning-based non-verbal features perform better than other features. In their setup, deep learning-based features gained 100% classification accuracy on the RLTDD dataset [87]. To consider their solution as a global solution, we must verify its performance on various real-life scenarios or other publicly available deception datasets. Table II summarizes the attributes of video modality, classification accuracy, and databases used in previous literature studies for deception detection.

G. *Deception detection Using Multiple Modalities*

Researchers investigated multimodal techniques, which incorporate features from multiple modalities. It is involved in the task of complex deception detection systems. These strategies aim to reduce the risk that arises in a single modality usage and the time-consuming analysis and decision-making processes required by prior techniques. Furthermore, by combining features from different modalities, the dataset is supplemented with the knowledge that is not attainable when using these modalities alone. The classifier's overall performance and confidence level show this scenario as well. García et al. have used a combination of visual, acoustic, and linguistic modalities [92]. They employed a fusion of features derived from every modality before applying the LSTM deep learning approach for classification. Sen et al. have also used a combination of visual, acoustic, and linguistic modalities [95]. They combined attributes from various modalities using both early and late fusion techniques. They employed an RF, ANN, and an SVM as three different classifiers. Zhang et al. integrated visual, verbal, and audio modalities [120]. In their research, they used a RF classifier with different combinations of feature sets. They outperformed the use of individual modalities by achieving maximum accuracy while combining all modalities. Chebbi and Jebara used features from linguistic, visual, and acoustic modalities [111]. In linguistic modality, they have taken a vector containing 21 different attributes, such as passive words usage, percentage of positive emotions, percentage of self-reference, percentage of denying words, etc. In Visual modality, they used a vector containing 39 binary features from hand gestures and facial cues. In Acoustic modality, they used 72 pitch base features set. They used the KNN classifier in their experiment. Researchers examined multiple different fusions of modalities to identify deception in their research studies. Gupta et al. used the features from visual, linguistic, acoustic, and Physiological modalities [10]. They used Electroencephalogram(EEG) features from physiological modality to improve performance achieved by visual, linguistic, and acoustic modalities. Abouelenien et al. used features of the linguistic, thermal, and physiological modalities [14]. Although multimodal deception detection systems aim to reduce the risk associated with depending on a single modality, they also make it challenging to combine the best features of each modality. The multiple modality approach is the democratic approach. The democratic technique works well when every voter is good on their own because even if a tiny portion of them are wrong in a specific case, the other experts comprise this shortcoming. However, even if one voter is perfect, the ensemble will perform poorly if all the other experts perform poorly. Therefore, when the overall result from multiple modalities is poor, it is challenging to determine the perfect features that perform better from individual modalities. Table III summarizes the attributes of various modalities, classification accuracy, and databases used in previous literature studies for deception detection.

3. DATASETS AVAILABLE FOR DECEPTION DETECTION

This section will discuss the relevant publicly available datasets for Deception Detection.

A. *Bag-of-Lies Deception Detection dataset*

Gupta et al. created this dataset based on a realistic scenario [10]. This collection contains 325 separate data samples. In a sample size of 325 data samples, 162 are deceptive, and 163 are truthful. They collect data in video, audio, eye gaze, and electroencephalography (EEG) modalities. The participants in this study were 35 student volunteers from various origins who were all fluent in English. There were ten female and twenty-five male participants, each of whom was shown six to ten images from the chosen set. The recordings span a duration from 3.5 seconds to 42 seconds. Subjective interviews are the most common method of generating deception datasets. These interviews

TABLE II. Visual Modality Related Features

Ref	Features	Dataset	Classifier	Maximum Classification Accuracy(%)
[20]	Facial Action Unit	RLTDD	LR, RF, and SVM	76.84 (SVM-RBN)
[101]	Facial Action Unit	Custom dataset of real life youtube videos created by author	RF	76.92
[86]	Facial Action Unit	RLTDD	SVM	67.20
[105]	Facial Action Unit and Facial Emotion	RLTDD	RF, KNN, and Bagging	86.90 (RF)
[87]	Facial and Hand gesture based manually annotated microexpression	RLTDD	RF, DT	76.03 (DT)
[95]	Facial and Hand gesture based manually annotated microexpression	RLTDD	RF, SVM, and NN	61.58(RF)
[111]	Facial and Hand gesture based manually annotated microexpressions	RLTDD	KNN	94
[94]	Facial and Hand gesture based manually annotated microexpressions	RLTDD	AdaBoost	88
[15]	3D CNN	RLTDD	3D CNN	93.08
[118]	pre-trained CNN from GoogleNet	RLTDD	Bi-directional LSTM	100
[117]	Facial and Body Language features based on FFCSN	RLTDD	FFCSN	93.16
[106]	Facial Emotion, Facial and Hand gesture based manually annotated microexpressions	RLTDD	DT, RF, and KNN	87.59
[107]	Facial Affect and Emotions, Facial Action Unit	RLTDD	AdaBoost	86
[113]	Body Language, Hand Gestures	RLTDD	LSTM	90.96
[115]	Body Language, Hand Gestures	Custom Dataset created by author	LSTM	88.41

require participants to tell lies or to tell the truth. But in this dataset, Some hypothetical scenario is given to participants where they must express truthful and deceitful opinions. This dataset includes multiple modalities while providing a real objective purpose to detect deception. It is an actual data collection that enables a real-world deception situation in which participants can choose to be truthful or deceitful.

B. *Real Life Trial Deception Detection (RLTDD) dataset*

Verónica Pérez-Rosas et al. created this corpus in high-stakes, realistic scenarios [87]. This dataset includes 61 misleading and 60 truthful data instances. The population of speakers consists of 21 ladies and 35 gents, varying in age from 16 to 60. A typical video clip in the data sample lasts 28 seconds. A deceptive video lasts 27.7 seconds, while a truthful video lasts 28.3 seconds on average. The author collected data in video, audio, and linguistic (English-language transcripts) modalities. The authors gathered the data by looking for public multimedia sources with trial hearing records where they can accurately observe and verify deception and truthfulness.

C. *Miami University Deception Detection Database*

Lloyd, E. Paige et al. created Miami University Deception Detection Database (MU3D) [121]. There are 320 video clips in the dataset. The participants in the dataset are Black and White, with both genders stating the truth and lying, and were university students. The dataset contains 80 volunteers (20 Black women, 20 Black men, 20 White women, and 20 White men) who discussed their social interactions truthfully and deceitfully. The average duration of the videos in this collection is 35.72 seconds. Each participant in the dataset produced four videos with categories of positive truth, negative truth, positive lie, and negative lie. There is a total of 320 videos addressing the participants' ethnicity, participants' gender, facial valence, and truthfulness of the statement. The dataset provides information in three modalities: audio, video, and English-language transcripts.

TABLE III. Multiple Modality Related Features

Ref	Modalities and Features	Dataset	Optimal Combination of Modality and Classifier	Maximum Classification Accuracy(%)
[92]	- Visual: Facial Micro expressions - Acoustic: Glottan Flow, hmpdd, hmpdm, mcep - Linguistic: LIWC, POS, and Bag of words	RLTDD	Acoustic with LSTM classifier	0.73(AUC)
[14]	- Thermal: Forehead and Periorbital region of face - Physiological: Pulse rate, Blood Pressure, Breathing rate, and Skin conductance - Linguistic: Unigrams, LIWC	Multimodal Dataset for Deception Detection	(Physiological + Thermal Periorbital region) featues with SVM classifier	79.31
[95]	- Visual: Facial micro expressions, Hand gestures, and Facial Action Unit - Acoustic: Pitch related features and Silent and Speech Histogram - Linguistic: Unigram and LIWC	RLTDD	Linguistic, Visual, Acoustic with Neural Network classifier.	72.88
[120]	- Visual: Facial micro expressions, Hand gestures - Acoustic: Interspeech 2009 (IS09) Challenge low level and functional acoustic features - Linguistic: Unigrams, LIWC, and POS	Box of Lies	Linguistic, Visual, Acoustic with RF classifier.	73
[89]	- Visual: Facial micro expressions, Hand gestures - Acoustic: 72 pitch base features - Linguistic: 21 features such as passive words usage, percentage of positive emotions, percentage of self reference, percentage of deny words, etc.	RLTDD	Linguistic, Visual, Acoustic with KNN	85
[10]	- Visual: Local Binary Pattern features of facial Image, Eye Gaze, Eye blink rate, Eye fixations pattern, and Pupil size deviation - Acoustic: Spectral centroid, Spectral bandwidth, Spectral rolloff, ZCR, Chroma frequencies, and MFCC - Linguistic: Bag of Words, LIWC - Physiological: Electroencephalogram(EEG)	Bag-of-Lies Dataset	Linguistic, Visual, Acoustic, Physiological with score level late fusion with various classifiers such as KNN, SVM	66.17
[94]	- Visual: Facial micro expressions, Hand gestures - Acoustic: Cepstral Coefficients (CC) - Linguistic: Bag-of-N-Grams (BoNG)	RLTDD	Linguistic with SVM, Visual with AdaBoost, Acoustic with CC-SRKDA then used Majority voting for Final Decision.	97

D. A Multimodal Dataset for Deception Detection

Pérez-Rosas et al. created this dataset [96]. The authors have done a series of experiments to obtain this dataset and asked participants to generate deceptive or truthful responses in three situations: best friend, abortion, and fake crime. The authors prepared a dataset with 30 (5 female and 25 male) student volunteers, aged between 22 to 38 years, who spoke English, and came from different ethnic origins. They provided physiological, thermal, and visual modality features in the dataset. They utilized four separate sensors to collect physiological data.

In addition to the above datasets, other datasets are available in various modalities. The Columbia-SRI-Colorado dataset [78] is based on a hypothetical scenario on audio modality. It contains an audio interview of 32 hours with 32 (16 men, 16 women) American English participants. The ReLiDDB dataset [122] is based on a hypothetical scenario on audio modality. In this dataset, each participant told true or false stories. This dataset involved 40 participants. The average length of audio data is approximately 80 seconds. The Open domain dataset [123] is based on linguistic modality and created using crowd-sourcing. The dataset contains 7168 sentences (3584 truth, 3584 lies) from participants aged between 18 to 72. The authors used Amazon Mechanical Turk and suggested that every participant worker provide seven misleading and seven honest sentences on topics of their interest. EEG-P300 dataset [124] is based on EEG-based physiological modality. There are 11 participants with an average of 20 years of age in this dataset. The author created this dataset in lab controlled environment for a hypothetical scenario.

4. DISCUSSION

The findings of this study offer insight into different techniques for spotting deceit. The researcher used diverse machine learning models with various modalities separately or in combination to identify deceptive activities. Conventional methods require human intervention, so they are inapplicable in recent world scenarios because of their time-consuming traits. Machine learning models reduce the need for human intervention and speed up the production of results. We have discussed features from physiological, thermal, linguistic, visual, and acoustic modalities in combination with different machine-learning models. In thermal modality, various research studies observed temperature changes in the periorbital region of the face. In linguistic modality, researchers identified various semantic and syntactic linguistic features. In the visual modality, they employed various facial clues, emotions, and eye-related characteristics as input to machine learning models. We include analyses of diverse machine learning models that learn and make predictions or judgments based on features extracted from various modalities. We review a range of machine learning models such as SVM, ANN, RF, KNN, AdaBoost, and deep learning approaches with LSTM in our study.

SVM (Support Vector Machine) is a well-accepted classifier for the classification task. It performs exceptionally well in high-dimensional space with large sets of input features. It uses little storage, is less prone to an overfitting issue, and performs well on linear and non-linear data. Although it operates worse as the number of classes increases, it is ideally adapted to deception detection since there are only two classes deceit and truth. We located that SVM was used in numerous research studies on visual features and obtained minimum and maximum accuracy of 61.58% and 76.84%. We identified minimum and maximum accuracies ranging from 34.23% to 80% in the acoustic modality using SVM. On linearly separable data, SVM performs well. Hence, the accuracy of literature research increases when SVM is employed with RBF kernel to change feature space to a higher dimensional to handle non-separable data. SVM is applied to linguistic features in one study [94], whereas they employed other classifiers on other modalities. They achieved an accuracy of 97% using a majority voting process. Random Forest (RF) is an ensemble learning algorithm that constructs multiple decision trees and combines their results to make a final prediction. It can be computationally and memory intensive when dealing with large datasets, but it is often highly accurate. It reveals the significance of each feature in the dataset. Researchers used RF classifiers on acoustic, linguistic, and visual modalities in many research studies. RF performs better than other classifiers, such as SVM, ANN, and DT, combining audio and visual features. According to the paper [120], RF gets 73% accuracy on the RLTTD dataset [87] when applied to linguistic, audio, and visual modalities. In a study [105], the RF classifier beats the bagging and KNN classifiers and achieves 86.90% accuracy on facial clues such as facial action units and emotions.

AdaBoost (Adaptive Boosting) is a machine-learning approach that combines several weak classifiers to produce a robust classifier. It emphasizes data instances that are challenging to classify while downplaying the significance of easily classified data instances. This tendency enables the algorithm to concentrate on the data instances that are the most difficult and enhances the overall performance. It requires less training time and provides high accuracy in various real-world classification problems. It works well on binary and multi-class classifiers. It requires a large amount of data to train effectively, or else it leads to overfitting. It is sensitive to noisy and imbalanced data. According to the paper [107], AdaBoost gets 86% accuracy on the RLTTD dataset [87] when used with facial clues. We have seen less number of research studies using the AdaBoost algorithm, but it is promising for future enhancements. K-nearest neighbor (KNN) is a simple, fast, and versatile classifier. It is a non-parametric machine learning model. It works for both classification and regression problems. It is computationally intensive when dealing with high-dimensional and large datasets. It is sensitive to outliers and requires feature scaling to operate. Compared to acoustic modality, KNN gives exceptional performance for visual modality. KNN also provides adequate performance when multiple modalities are combined to detect deception.

Artificial Neural Networks (ANN) classifier is a popular machine learning model that imitates the procedure of the human brain. It has excelled at many challenging pattern recognition tasks. It performs well on linear and non-linear data and is noise resistant. It is computationally intensive and prone to overfitting when dealing with limited-size datasets. It requires a large amount of data and proper hyperparameter tuning to train effectively. According to the paper [95], ANN gets 72.88% accuracy on the RLTTD dataset [87] when applied to linguistic, audio, and visual modalities. ANN suggests moderate performance in many research studies with various modalities compared to other machine learning models. However, we observed better performance in CNN and LSTM networks. Convolutional Neural Networks (CNN) and Long Short Term Memory (LSTM) Networks are deep learning architectures that are very promising in diverse computer vision and natural language processing tasks. CNN is translation-invariant and very useful in extracting features from videos and images. One can use a pre-trained CNN model of one task in another task that reduces the need for large datasets and computing resources. CNNs are susceptible to overfitting with small datasets. LSTM works well in long-term sequential data such as video, audio, and language sequences. A combination of CNN with LSTM has proven to be very effective in facial feature extraction and facial muscle movement recognition. In a deceptive environment, facial muscle movement provides a significant clue of the subjects' veracity behavior. Spatiotemporal event-related data are available in the audio, visual, and language modalities. When dealing with spatiotemporal data, CNN and LSTM work efficiently. Combining CNN and LSTM allows efficient capture of facial expressions, hand and body motion, and eye gaze movement. The performance of CNN and LSTM in linguistic, auditory, visual, and multimodal modalities is quite acceptable in numerous research studies.

So, Performance is not solely improved by the machine learning model or employed modality. It also depends on the dataset considered in the research study and the optimal set of features selected in every modality. We observed that the machine learning model that captures spatiotemporal relationships performs better than others, and results are acceptable when we combine multiple modalities in deception detection.

5. CHALLENGES AND FUTURE DIRECTION IN DECEPTION DETECTION

In early research, people used a physiological device such as a polygraph, electrogram, and Brain imaging sensors to detect deception. These tools are expensive, time-consuming, and non-invasive. This physiological modality demands qualified human operators with human intervention. We, therefore, need less expensive, non-invasive, and machine-learning-based modalities to detect deception. Linguistic modality-based tools, including LIWC, GloVe, and WordtoVec are available to detect deceit. The majority of these tools are only available in the English language, and one must supply a corpus that is pertinent to the field to use them. As a result, a machine learning model trained to detect deception in one context may not perform well when used in another context. Video and audio modalities have a temporal correlation that renders it difficult to identify deception-related indications, and therefore various research studies provided very low accuracy in previous research. The availability of large-scale datasets is another major obstacle in deception detection research. There are only a few globally accessible datasets available for deception detection that comprise multiple modalities. Researchers created many of these datasets in a simulated laboratory environment. These datasets contain a limited number of subjects and a static set of domain-related questions. The subjects participating in these datasets may not experience the same level of cognitive load as in real-life scenarios. Therefore, the machine learning models trained on these datasets may not perform well in a real-life scenario. Data from the real world have a significant amount of noise. Researchers performed deception detection with good results without combining datasets from several domains. However, the performance suffers when they employ training and testing data from datasets of different areas. According to psychological research, verbal clues are preferable to non-verbal ones, although professional investigators have observed the opposite as correct. Deep learning methods suggest that verbal clues are effective in deception detection. There is a significant risk of overfitting scenarios in deep learning models because of the diminutive training data of the globally accessible corpus. Sometimes, Cultural habits and stressful situations also increase the false positive rate in deception detection.

Researchers have become interested in the multimodal fusion technique in recent studies. In numerous research investigations, researchers used Facial Action Units (FAU) as a feature set. In the future, feature selection approaches will become efficient techniques to determine the crucial FAU. Numerous academic studies have employed manually annotated facial gestures from the RLTDD dataset. In the future, it will be possible to recognize micro facial gestures automatically using optical flow-based motion tracking algorithms. One can apply a combination of transfer learning with the deep learning model to get over the issue of limited data and mitigate the problem of overfitting. Researchers have already demonstrated LBP-TOP, 3D-CNN, and LSTM as efficient methods to detect temporal correlation in audio, video, and linguistic modalities. One can use a combination of Facial microexpression, facial emotion, facial affect related to facial emotion, and facial symmetry data to detect deception in the future.

6. CONCLUSION

This paper provides a detailed study of the diverse modalities employed to tackle the immensely challenging problem of deception detection using machine learning techniques. We have discussed the comparative effectiveness of several features from different modalities used to identify deception. Physiological modalities, including Polygraph testing, Electroencephalography (EEG), and Brain imaging, require sensors and physical devices. These modalities are non-invasive, costly, and time-consuming. These techniques require human intervention, and the subject is aware of the examination, which makes these methods unreliable. Non-contact deception detection techniques avoid placing sensors on the body of a participant. They conduct a test without the awareness of the participant. So, we have discussed deception detection based on thermal, linguistic, acoustic, and visual modality features. Several research studies have indicated that cognitive load increases in the case of deception. The periorbital region of the face experiences an increase in temperature when cognitive load increases. The thermal infrared camera monitors the changes in temperature in the different areas of the face to detect deception. A linguistic feature provides accurate results if the training and validation data are from the same knowledge base. The linguistic features analyze syntactic and semantic relationships of the data to identify deception. Acoustic features related to audio signals use pitch and other frequency-based features to perform automatic deception detection. Visual modality uses a variety of facial clues to detect deception, such as micro facial expressions, eye gaze, eye blink rate, etc. The human face is considered the most reliable representation of internal cognitive load. Various neural networks and deep learning-based techniques, such as CNN, RNN, and LSTM, are effective at detecting deception since they also examine the spatiotemporal relationships of data. Numerous research studies have encountered that when the subject's cognitive load increases, the capacity of deception detection also increases. We have also discussed various well-known datasets used in different literature studies. Researchers have recently looked into multimodal techniques, which integrate features from multiple modalities. These methods strive to eliminate the risk of relying on a single modality and the time-consuming analysis and decision-making procedures required by earlier techniques. In addition, combining the features of different modalities enriches the dataset with information. Individual modalities alone make it inaccessible. The classifier's overall performance and confidence level reflect this.

References

[1] A. Vrij, *Detecting lies and deceit: The psychology of lying and implications for professional practice*. Wiley series in the psychology of crime, policing and law,Wiley, 2000.

[2] L. Su and M. D. Levine, "High-stakes deception detection based on facial expressions," in *2014 22nd International Conference on Pattern Recognition*, 2014, pp. 2519–2524.

[3] J. Charles F. Bond and B. M. DePaulo, "Accuracy of deception judgments," *Personality and Social Psychology Review*, vol. 10, no. 3, pp. 214–234, 2006, pMID: 16859438. [Online]. Available: https://doi.org/10.1207/s15327957pspr1003_2

[4] M. Burzo, M. Abouelenien, V. Perez-Rosas, and R. Mihalcea, "Multimodal deception detection," in *The Handbook of Multimodal-Multisensor Interfaces: Signal Processing, Architectures, and Detection of Emotion and Cognition-Volume 2*, 2018, pp. 419–453.

[5] K. A. Crockett, S. Zoltán, J. D. O'Shea, T. Szklarski, A. Malamou, and G. Boultadakis, "Do europe's borders need multi-faceted biometric protection," *Biometric Technology Today*, vol. 2017, pp. 5–8, 2017.

[6] D. Grubin and L. Madsen, "Lie detection and the polygraph: A historical review," *The Journal of Forensic Psychiatry & Psychology*, vol. 16, no. 2, pp. 357–369, 2005. [Online]. Available: https://doi.org/10.1080/14789940412331337353

[7] P. Ekman, M. O'Sullivan, W. V. Friesen, and K. R. Scherer, "Invited article: Face, voice, and body in detecting deceit," *Journal of nonverbal behavior*, vol. 15, no. 2, pp. 125–135, 1991.

[8] M. Derksen, "Control and resistance in the psychology of lying," *Theory & Psychology*, vol. 22, no. 2, pp. 196–212, 2012. [Online]. Available: https://doi.org/10.1177/0959354311427487

[9] E. Lendaro, *Investigating Phantom Motor Execution as Treatment of Phantom Limb Pain*. Chalmers Tekniska Hogskola (Sweden), 2021.

[10] V. Gupta, M. Agarwal, M. Arora, T. Chakraborty, R. Singh, and M. Vatsa, "Bag-of-lies: A multimodal dataset for deception detection," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2019, pp. 83–90.

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

[12] B. A. Rajoub and R. Zwiggelaar, "Thermal facial analysis for deception detection," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 6, pp. 1015–1023, 2014.

[13] I. Pavlidis, N. L. Eberhardt, and J. A. Levine, "Seeing through the face of deception," *Nature*, vol. 415, no. 6867, pp. 35–35, 2002.

[14] M. Abouelenien, V. Pérez-Rosas, R. Mihalcea, and M. Burzo, "Detecting deceptive behavior via integration of discriminative features from multiple modalities," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 5, pp. 1042–1055, 2017.

[15] G. Krishnamurthy, N. Majumder, S. Poria, and E. Cambria, "A deep learning approach for multimodal deception detection," *CoRR*, vol. abs/1803.00344, 2018. [Online]. Available: http://arxiv.org/abs/1803.00344

[16] W. Merghani, A. K. Davison, and M. H. Yap, "A review on facial micro-expressions analysis: datasets, features and metrics," *arXiv preprint arXiv:1805.02397*, 2018.

[17] M. Owayjan, A. Kashour, N. Al Haddad, M. Fadel, and G. Al Souki, "The design and development of a lie detection system using facial micro-expressions," in *2012 2nd International Conference on Advances in Computational Tools for Engineering Applications (ACTEA)*, 2012, pp. 33–38.

[18] W. Khan, K. Crockett, J. O'Shea, A. Hussain, and B. M. Khan, "Deception in the eyes of deceiver: A computer vision and machine

learning based automated deception detection," *Expert Systems with Applications*, p. 114341, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417420310289

[19] P. Ekman and W. V. Friesen, "Facial action coding system," *Environmental Psychology & Nonverbal Behavior*, 1978.

[20] D. Avola, L. Cinque, G. L. Foresti, and D. Pannone, "Automatic deception detection in rgb videos using facial action units," in *Proceedings of the 13th International Conference on Distributed Smart Cameras*, ser. ICDSC 2019. New York, NY, USA: Association for Computing Machinery, 2019. [Online]. Available: https://doi.org/10.1145/3349801.3349806

[21] J. G. Proudfoot, J. L. Jenkins, J. K. Burgoon, and J. F. Nunamaker, "Deception is in the eye of the communicator: Investigating pupil diameter variations in automated deception detection interviews," in *2015 IEEE International Conference on Intelligence and Security Informatics (ISI)*, 2015, pp. 97–102.

[22] D. Borza, R. Itu, and R. Danescu, "In the eye of the deceiver: Analyzing eye movements as a cue to deception," *Journal of Imaging*, vol. 4, no. 10, p. 120, 2018.

[23] N. Carissimi, C. Beyan, and V. Murino, "A multi-view learning approach to deception detection," in *2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018)*, 2018, pp. 599–606.

[24] J. A. Larson, "Lying and its detection." 1932.

[25] F. A. Kozel, L. J. Revell, J. P. Lorberbaum, A. Shastri, J. D. Elhai, M. D. Horner, A. Smith, Z. Nahas, D. E. Bohning, and M. S. George, "A pilot study of functional magnetic resonance imaging brain correlates of deception in healthy young men," *The Journal of Neuropsychiatry and Clinical Neurosciences*, vol. 16, no. 3, pp. 295–305, 2004.

[26] G. Ganis, J. P. Rosenfeld, J. Meixner, R. A. Kievit, and H. E. Schendan, "Lying in the scanner: covert countermeasures disrupt deception detection by functional magnetic resonance imaging," *Neuroimage*, vol. 55, no. 1, pp. 312–319, 2011.

[27] P. Bhuvaneswari and J. S. Kumar, "Influence of linear features in nonlinear electroencephalography (eeg) signals," *Procedia computer science*, vol. 47, pp. 229–236, 2015.

[28] T. A. Gannon, A. R. Beech, and T. Ward, "Risk assessment and the polygraph," *The use of the polygraph in assessing, treating and supervising sex offenders: A practitioner's guide*, pp. 129–154, 2009.

[29] M. Derksen, "Control and resistance in the psychology of lying," *Theory & Psychology*, vol. 22, no. 2, pp. 196–212, 2012.

[30] M. K. Taylor, D. S. Horning, J. F. Chandler, J. B. Phillips, J. Y. Khosravi, J. E. Bennett, H. Halbert, B. J. Fern, and H. Gao, "A comparison of approaches to detect deception," NAVAL AEROSPACE MEDICAL RESEARCH LAB PENSACOLA FL, Tech. Rep., 2010.

[31] S. Leal and A. Vrij, "Blinking during and after lying," *Journal of Nonverbal Behavior*, vol. 32, no. 4, pp. 187–194, 2008.

[32] K. K. Lim, M. Friedrich, J. Radun, and K. Jokinen, "Lying through the eyes: Detecting lies through eye movements," in *Proceedings of the 6th Workshop on Eye Gaze in Intelligent Human Machine Interaction: Gaze in Multimodal Interaction*, ser. GazeIn '13. New York, NY, USA: Association for Computing Machinery, 2013, p. 51–56. [Online]. Available: https://doi.org/10.1145/2535948.2535954

[33] X. Fang, Y. Sun, X. Zheng, X. Wang, X. Deng, and M. Wang, "Assessing deception in questionnaire surveys with eye-tracking," *Frontiers in Psychology*, vol. 12, 2021. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.774961

[34] A. Vrij, J. Oliveira, A. Hammond, and H. Ehrlichman, "Saccadic eye movement rate as a cue to deceit," *Journal of Applied Research in Memory and Cognition*, vol. 4, no. 1, pp. 15–19, 2015.

[35] J. G. Proudfoot, J. L. Jenkins, J. K. Burgoon, and J. F. Nunamaker Jr, "More than meets the eye: How oculometric behaviors evolve over the course of automated deception detection interactions," *Journal of Management Information Systems*, vol. 33, no. 2, pp. 332–360, 2016.

[36] P. K. Lakkapragada, "Using eye-tracking to understand user behavior in deception detection system interaction," 2016.

[37] S. George, M. Manohara Pai, R. M. Pai, and S. K. Praharaj, "Eye blink count and eye blink duration analysis for deception detection," in *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2017, pp. 223–229.

[38] F. Marchak, "Detecting false intent using eye blink measures," *Frontiers in Psychology*, vol. 4, 2013. [Online]. Available: https://www.frontiersin.org/article/10.3389/fpsyg.2013.00736

[39] B. S. Perelman, "Detecting deception via eyeblink frequency modulation," *PeerJ*, vol. 2, p. e260, 2014.

[40] S. Mann, A. Vrij, S. Leal, P. A. Granhag, L. Warmelink, and D. Forrester, "Windows to the soul? deliberate eye contact as a cue to deceit," *Journal of Nonverbal Behavior*, vol. 36, no. 3, pp. 205–215, 2012.

[41] M. Monaro, P. Capuozzo, F. Ragucci, A. Maffei, A. Curci, C. Scarpazza, A. Angrilli, and G. Sartori, "Using blink rate to detect deception: A study to validate an automatic blink detector and a new dataset of videos from liars and truth-tellers," in *International Conference on Human-Computer Interaction*. Springer, 2020, pp. 494–509.

[42] T. Seymour, C. Baker, and J. Gaunt, "Combining blink, pupil, and response time measures in a concealed knowledge test," *Frontiers in Psychology*, vol. 3, 2013. [Online]. Available: https://www.frontiersin.org/article/10.3389/fpsyg.2012.00614

[43] E. Trifiletti, S. D'Ascenzo, L. Lugli, V. M. Cocco, G. A. Di Bernardo, C. Iani, S. Rubichi, R. Nicoletti, and L. Vezzali, "Truth and lies in your eyes: Pupil dilation of white participants in truthful and deceptive responses to white and black partners," *Plos one*, vol. 15, no. 10, p. e0239512, 2020.

[44] M. T. B. Noman and M. A. R. Ahad, "Mobile-based eye-blink detection performance analysis on android platform," *Frontiers in ICT*, vol. 5, 2018. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fict.2018.00004

[45] R. M. Schuetzler, "Countermeasures and eye tracking deception detection," 2012.

[46] I. Pavlidis and J. Levine, "Monitoring of periorbital blood flow rate through thermal image analysis and its application to polygraph testing," in *2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 3, 2001, pp. 2826–2829 vol.3.

[47] P. Tsiamyrtzis, J. Dowdall, D. Shastri, I. Pavlidis, M. Frank, and P. Ekman, "Lie detection-recovery of the periorbital signal through tandem tracking and noise suppression in thermal facial video," in *Proceedings of SPIE Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense IV*, vol. 5778, 2005, pp. 29–31.

[48] D. A. Pollina, A. B. Dollins, S. M. Senter, T. E. Brown, I. Pavlidis, J. A. Levine, and A. H. Ryan, "Facial skin surface temperature changes during a "concealed information" test," *Annals of biomedical Engineering*, vol. 34, no. 7, pp. 1182–1189, 2006.

[49] U. Jain, B. Tan, and Q. Li, "Concealed knowledge identification using facial thermal imaging," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2012, pp. 1677–1680.

[50] L. Warmelink, A. Vrij, S. Mann, S. Leal, D. Forrester, and R. P. Fisher, "Thermal imaging as a lie detection tool at airports," *Law and human behavior*, vol. 35, no. 1, pp. 40–48, 2011.

[51] M. Abouelenien, V. Pérez-Rosas, R. Mihalcea, and M. Burzo, "Deception detection using a multimodal approach," in *Proceedings of the 16th International Conference on Multimodal Interaction, ICMI 2014, Istanbul, Turkey, November 12-16, 2014*, A. A. Salah, J. F. Cohn, B. W. Schuller, O. Aran, L. Morency, and P. R. Cohen, Eds. ACM, 2014, pp. 58–65. [Online]. Available: https://doi.org/10.1145/2663204.2663229

[52] K. Park, H. W. Suk, H. Hwang, and J.-H. Lee, "A functional analysis of deception detection of a mock crime using infrared thermal imaging and the concealed information test," *Frontiers in Human Neuroscience*, vol. 7, p. 70, 2013. [Online]. Available: https://www.frontiersin.org/article/10.3389/fnhum.2013.00070

[53] A. Derakhshan, M. Mikaeili, T. Gedeon, and A. M. Nasrabadi, "Identifying the optimal features in multimodal deception detection," *Multimodal Technologies and Interaction*, vol. 4, no. 2, 2020. [Online]. Available: https://www.mdpi.com/2414-4088/4/2/25

[54] M. L. Newman, J. W. Pennebaker, D. S. Berry, and J. M. Richards, "Lying words: Predicting deception from linguistic styles," *Personality and social psychology bulletin*, vol. 29, no. 5, pp. 665–675, 2003.

[55] B. M. DePaulo, J. J. Lindsay, B. E. Malone, L. Muhlenbruck, K. Charlton, and H. Cooper, "Cues to deception." *Psychological bulletin*, vol. 129, no. 1, p. 74, 2003.

[56] J. Bachenko, E. Fitzpatrick, and M. Schonwetter, "Verification and implementation of language-based deception indicators in civil and criminal narratives," in *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*. Manchester, UK: Coling 2008 Organizing Committee, Aug. 2008, pp. 41–48. [Online]. Available: https://aclanthology.org/C08-1006

[57] V. Hauch, I. Blandón-Gitlin, J. Masip, and S. L. Sporer, "Are computers effective lie detectors? a meta-analysis of linguistic cues to deception," *Personality and social psychology Review*, vol. 19, no. 4, pp. 307–342, 2015.

[58] L. Zhou, J. K. Burgoon, J. F. Nunamaker, and D. Twitchell, "Automating linguistics-based cues for detecting deception in text-based asynchronous computer-mediated communications," *Group decision and negotiation*, vol. 13, no. 1, pp. 81–106, 2004.

[59] C. L. Toma and J. T. Hancock, "Reading between the lines: Linguistic cues to deception in online dating profiles," in *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*, ser. CSCW '10. New York, NY, USA: Association for Computing Machinery, 2010, p. 5–8. [Online]. Available: https://doi.org/10.1145/1718918.1718921

[60] M. Yancheva and F. Rudzicz, "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)*. Sofia, Bulgaria: Association for Computational Linguistics, Aug. 2013, pp. 944–953. [Online]. Available: https://aclanthology.org/P13-1093

[61] R. Mihalcea and C. Strapparava, "The lie detector: Explorations in the automatic recognition of deceptive language," in *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, ser. ACLShort '09. USA: Association for Computational Linguistics, 2009, p. 309–312.

[62] C. Chen, H. Zhao, and Y. Yang, "Deceptive opinion spam detection using deep level linguistic features," in *Natural Language Processing and Chinese Computing*, J. Li, H. Ji, D. Zhao, and Y. Feng, Eds. Cham: Springer International Publishing, 2015, pp. 465–474.

[63] J. W. Pennebaker, M. E. Francis, and R. J. Booth, "Linguistic inquiry and word count: Liwc 2001," *Mahway: Lawrence Erlbaum Associates*, vol. 71, no. 2001, p. 2001, 2001.

[64] Á. Almela, R. Valencia-García, and P. Cantos, "Seeing through deception: A computational approach to deceit detection in written communication," in *Proceedings of the Workshop on Computational Approaches to Deception Detection*. Avignon, France: Association for Computational Linguistics, Apr. 2012, pp. 15–22. [Online]. Available: https://aclanthology.org/W12-0403

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

[66] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Advances in Neural Information Processing Systems*, C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger, Eds., vol. 26. Curran Associates, Inc., 2013. [Online]. Available: https://proceedings.neurips.cc/paper/2013/file/9aa42b31882ec039965f3c4923ce901b-Paper.pdf

[67] T. Fornaciari and M. Poesio, "Automatic deception detection in italian court cases," *Artificial intelligence and law*, vol. 21, no. 3, pp. 303–340, 2013.

[68] V. Pérez-Rosas and R. Mihalcea, "Cross-cultural deception detection," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Baltimore, Maryland: Association for Computational Linguistics, Jun. 2014, pp. 440–445. [Online]. Available: https://aclanthology.org/P14-2072

[69] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, "Finding deceptive

opinion spam by any stretch of the imagination," *arXiv preprint arXiv:1107.4557*, 2011.

[70] J. Li, M. Ott, C. Cardie, and E. Hovy, "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)*, 2014, pp. 1566–1576.

[71] B. Verhoeven and W. Daelemans, "CLiPS stylometry investigation (CSI) corpus: A Dutch corpus for the detection of age, gender, personality, sentiment and deception in text," in *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. Reykjavik, Iceland: European Language Resources Association (ELRA), May 2014, pp. 3081–3085. [Online]. Available: http://www.lrec-conf.org/proceedings/lrec2014/pdf/1_Paper.pdf

[72] S. V. Fernandes and M. S. Ullah, "A comprehensive review on features extraction and features matching techniques for deception detection," *IEEE Access*, vol. 10, pp. 28 233–28 246, 2022.

[73] T. Painter and A. Spanias, "Perceptual coding of digital audio," *Proceedings of the IEEE*, vol. 88, no. 4, pp. 451–515, 2000.

[74] S. V. Fernandes and M. S. Ullah, "Phychoacoustic masking of delta and time-difference cepstrum features for deception detection," in *2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*. IEEE, 2020, pp. 0213–0217.

[75] N. Srivastava and S. Dubey, "Deception detection using artificial neural network and support vector machine," in *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, 2018, pp. 1205–1208.

[76] M. Sanaullah and K. Gopalan, "Deception detection in speech using bark band and perceptually significant energy features," in *2013 IEEE 56th International Midwest Symposium on Circuits and Systems (MWSCAS)*, 2013, pp. 1212–1215.

[77] S. Desai, M. Siegelman, and Z. Maurer, "Neural lie detection with the csc deceptive speech dataset," *Stanford University*, 2017.

[78] F. Enos, *Detecting deception in speech*. Columbia University, 2009.

[79] H. Tao, P. Lei, M. Wang, J. Wang, and H. Fu, "Speech deception detection algorithm based on svm and acoustic features," in *2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT)*, 2019, pp. 31–33.

[80] H. Hung and G. Chittaranjan, "The idiap wolf corpus: exploring group behaviour in a competitive role-playing game," in *Proceedings of the 18th ACM international conference on Multimedia*, 2010, pp. 879–882.

[81] Y. Xie, R. Liang, H. Tao, Y. Zhu, and L. Zhao, "Convolutional bidirectional long short-term memory for deception detection with acoustic features," *IEEE Access*, vol. 6, pp. 76 527–76 534, 2018.

[82] C. Fan, H. Zhao, X. Chen, X. Fan, and S. Chen, "Distinguishing deception from non-deception in chinese speech," in *2015 Sixth International Conference on Intelligent Control and Information Processing (ICICIP)*, 2015, pp. 268–273.

[83] S. V. Fernandes and M. S. Ullah, "Use of machine learning for deception detection from spectral and cepstral features of speech signals," *IEEE Access*, vol. 9, pp. 78 925–78 935, 2021.

[84] Y. Zhou, H. Zhao, X. Pan, and L. Shang, "Deception detecting from speech signal using relevance vector machine and non-linear dynamics features," *Neurocomputing*, vol. 151, pp. 1042–1052, 2015.

[85] A. Xue, H. Rohde, and P. Finkelstein, "An acoustic automated lie detector," *Princeton University*, 2019.

[86] M. Jaiswal, S. Tabibu, and R. Bajpai, "The truth and nothing but the truth: Multimodal analysis for deception detection," *CoRR*, vol. abs/1903.04484, 2019. [Online]. Available: http://arxiv.org/abs/1903.04484

[87] V. Pérez-Rosas, M. Abouelenien, R. Mihalcea, and M. Burzo, "Deception detection using real-life trial data," in *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, ser. ICMI '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 59–66. [Online]. Available: https://doi.org/10.1145/2818346.2820758

[88] F. Eyben, M. Wöllmer, and B. W. Schuller, "Opensmile: the munich versatile and fast open-source audio feature extractor." in *ACM Multimedia*, A. D. Bimbo, S.-F. Chang, and A. W. M. Smeulders, Eds. ACM, 2010, pp. 1459–1462. [Online]. Available: http://dblp.uni-trier.de/db/conf/mm/mm2010.html#EybenWS10

[89] S. Chebbi and S. B. Jebara, "An audio-visual based feature level fusion approach applied to deception detection." in *VISIGRAPP (4: VISAPP)*, 2020, pp. 197–205.

[90] F. Soldner, V. Pérez-Rosas, and R. Mihalcea, "Box of lies: Multimodal deception detection in dialogues," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 1768–1777. [Online]. Available: https://aclanthology.org/N19-1175

[91] A. Gallardo-Antolín and J. M. Montero, "Detecting deception from gaze and speech using a multimodal attention lstm-based framework," *Applied Sciences*, vol. 11, no. 14, 2021. [Online]. Available: https://www.mdpi.com/2076-3417/11/14/6393

[92] R. Rill-García, H. J. Escalante, L. Villaseñor-Pineda, and V. Reyes-Meza, "High-level features for multimodal deception detection in videos," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2019, pp. 1565–1573.

[93] Z. Wu, B. Singh, L. S. Davis, and V. S. Subrahmanian, "Deception detection in videos," *CoRR*, vol. abs/1712.04415, 2017. [Online]. Available: http://arxiv.org/abs/1712.04415

[94] S. Venkatesh, R. Ramachandra, and P. Bours, "Robust algorithm for multimodal deception detection," in *2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, 2019, pp. 534–537.

[95] U. M. Sen, V. Perez-Rosas, B. Yanikoglu, M. Abouelenien, M. Burzo, and R. Mihalcea, "Multimodal deception detection using real-life trial data," *IEEE Transactions on Affective Computing*, pp. 1–1, 2020.

[96] V. Pérez-Rosas, R. Mihalcea, A. Narvaez, and M. Burzo, "A

multimodal dataset for deception detection," in *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. Reykjavik, Iceland: European Language Resources Association (ELRA), May 2014, pp. 3118–3122. [Online]. Available: http://www.lrec-conf.org/proceedings/lrec2014/pdf/869_Paper.pdf

[97] P. Ekman, "Lie catching and microexpressions," *The philosophy of deception*, vol. 1, no. 2, p. 5, 2009.

[98] E. L. Rosenberg and P. Ekman, *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, 2020.

[99] A. Zadeh, T. Baltrusaitis, and L. Morency, "Deep constrained local models for facial landmark detection," *CoRR*, vol. abs/1611.08657, 2016. [Online]. Available: http://arxiv.org/abs/1611.08657

[100] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L.-P. Morency, "Openface 2.0: Facial behavior analysis toolkit," in *2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018)*, 2018, pp. 59–66.

[101] L. Su and M. Levine, "Does "lie to me" lie to you? an evaluation of facial clues to high-stakes deception," *Computer Vision and Image Understanding*, vol. 147, pp. 52–68, 2016, spontaneous Facial Behaviour Analysis. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1077314216000345

[102] Z. Zhang, V. Singh, T. E. Slowe, S. Tulyakov, and V. Govindaraju, "Real-time automatic deceit detection from involuntary facial expressions," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–6.

[103] H. U. D. Ahmed, U. I. Bajwa, F. Zhang, and M. W. Anwar, "Deception detection in videos using the facial action coding system," *CoRR*, vol. abs/2105.13659, 2021. [Online]. Available: https://arxiv.org/abs/2105.13659

[104] S. H. Abd, I. A. Hashim, and A. S. Jalal, "Optimized action units features for efficient design of deception detection system," *Iraqi Journal of Information and Communication Technology*, vol. 1, no. 1, p. 104–111, Dec. 2021. [Online]. Available: https://www.ijict.edu.iq/index.php/ijict/article/view/160

[105] X. Shen, G. Fan, C. Niu, and Z. Chen, "Catching a liar through facial expression of fear," *Frontiers in Psychology*, vol. 12, 2021. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.675097

[106] J.-T. Yang, G.-M. Liu, and S. C.-H. Huang, "Emotion transformation feature: Novel feature for deception detection in videos," in *2020 IEEE International Conference on Image Processing (ICIP)*, 2020, pp. 1726–1730.

[107] L. Mathur and M. J. Matarić, *Introducing Representations of Facial Affect in Automated Multimodal Deception Detection*. New York, NY, USA: Association for Computing Machinery, 2020, p. 305–314. [Online]. Available: https://doi.org/10.1145/3382507.3418864

[108] M. B. Kursa and W. R. Rudnicki, "Feature selection with the boruta package," *Journal of Statistical Software*, vol. 36, no. 11, p. 1–13, 2010. [Online]. Available: https://www.jstatsoft.org/index.php/jss/article/view/v036i11

[109] D. Kollias, P. Tzirakis, M. A. Nicolaou, A. Papaioannou, G. Zhao, B. Schuller, I. Kotsia, and S. Zafeiriou, "Deep affect prediction in-the-wild: Aff-wild database and challenge, deep architectures, and beyond," *International Journal of Computer Vision*, vol. 127, no. 6, pp. 907–929, 2019.

[110] J. Allwood, L. Cerrato, K. Jokinen, C. Navarretta, and P. Paggio, "The mumin coding scheme for the annotation of feedback, turn management and sequencing phenomena," *Language Resources and Evaluation*, vol. 41, no. 3, pp. 273–287, 2007.

[111] S. Chebbi and S. B. Jebara, "Deception detection using multimodal fusion approaches," *Multimedia Tools and Applications*, pp. 1–30, 2021.

[112] J. Rothwell, Z. Bandar, J. O'Shea, and D. McLean, "Silent talker: a new computer-based system for the analysis of facial cues to deception," *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, vol. 20, no. 6, pp. 757–777, 2006.

[113] D. Avola, L. Cinque, M. De Marsico, A. Fagioli, and G. L. Foresti, "Lietome: Preliminary study on hand gestures for deception detection via fisher-lstm," *Pattern Recognition Letters*, vol. 138, pp. 455–461, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167865520303123

[114] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7291–7299.

[115] T. Randhavane, U. Bhattacharya, K. Kapsaskis, K. Gray, A. Bera, and D. Manocha, "The liar's walk: Detecting deception with gait and gesture," *CoRR*, vol. abs/1912.06874, 2019. [Online]. Available: http://arxiv.org/abs/1912.06874

[116] D. Avola, M. Cascio, L. Cinque, A. Fagioli, and G. L. Foresti, "Lietome: An ensemble approach for deception detection from facial cues," *International Journal of Neural Systems*, vol. 31, no. 02, p. 2050068, 2021, pMID: 33200620. [Online]. Available: https://doi.org/10.1142/S0129065720500689

[117] M. Ding, A. Zhao, Z. Lu, T. Xiang, and J. Wen, "Face-focused cross-stream network for deception detection in videos," *CoRR*, vol. abs/1812.04429, 2018. [Online]. Available: http://arxiv.org/abs/1812.04429

[118] S. Venkatesh, R. Ramachandra, and P. Bours, "Video based deception detection usingÂ deep recurrent convolutional neuralÂ network," in *Computer Vision and Image Processing*, N. Nain, S. K. Vipparthi, and B. Raman, Eds. Singapore: Springer Singapore, 2020, pp. 163–169.

[119] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1–9.

[120] J. Zhang, S. I. Levitan, and J. Hirschberg, "Multimodal Deception Detection Using Automatically Extracted Acoustic, Visual, and Lexical Features," in *Proc. Interspeech 2020*, 2020, pp. 359–363.

[121] E. P. Lloyd, J. C. Deska, K. Hugenberg, A. R. McConnell, B. T. Humphrey, and J. W. Kunstman, "Miami university deception detection database," *Behavior research methods*, vol. 51, no. 1, pp. 429–439, 2019.

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

[123] V. Pérez-Rosas and R. Mihalcea, "Experiments in open domain deception detection," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Lisbon, Portugal: Association for Computational Linguistics, Sep. 2015, pp. 1120–1125. [Online]. Available: https://aclanthology.org/D15-1133

[124] A. Turnip, M. F. Amri, H. Fakrurroja, A. I. Simbolon, M. A. Suhendra, and D. E. Kusumandari, "Deception detection of eeg-p300 component classified by svm method," in *Proceedings of the 6th International Conference on Software and Computer Applications*, ser. ICSCA '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 299–303. [Online]. Available: https://doi.org/10.1145/3056662.3056709

Maheta Hardikkumar Harishbhai
Maheta Hardikkumar Harishbhai is a Ph.D. research scholar from Gujarat Technological University, Ahmedabad, Gujarat, India. He completed his M.Tech. in Information Technology from Dharmsinh Desai University, Gujarat, India, in 2014. He completed his B.Tech. in Information Technology from Nirma University, Gujarat, India, in 2009. He is currently working as an Assistant Professor at the Department of Information Technology, Shantilal Shah Engineering College, Bhavnagar, Gujarat, India. He has 13 years of teaching experience. His research interest includes artificial intelligence, computer vision, image processing, machine learning, and data science. He has published research papers in prestigious conferences and journals concerning artificial intelligence and machine learning.

Dr. Dineshkumar B. Vaghela
Dr. Dineshkumar B. Vaghela completed his Ph.D. in Computer Engineering from Gujarat Technological University, Chandkheda, Gujarat, India, in 2017. He completed his M.E. in Computer Engineering from Sardar Patel University, V.V. Nagar, Gujarat, India, in 2010. He completed his B.E. in Information Technology from Gujarat University, Ahmedabad, Gujarat, India, in 2004. He is currently working as an Assistant Professor at the Department of Information Technology, Shantilal Shah Engineering College, Bhavnagar, Gujarat, India. He has 18 years of teaching experience. His research interest spans computer vision, image processing, machine learning, cloud computing, big data analytics, and IoT. He has published research papers in prestigious conferences and journals concerning computer vision, image processing, machine learning, and cloud computing.