

Received February 11, 2022, accepted February 28, 2022, date of publication March 8, 2022, date of current version March 17, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3157821

A Comprehensive Review on Features Extraction and Features Matching Techniques for Deception Detection

SINEAD V. FERNANDES¹, (Member, IEEE), AND
MUHAMMAD SANA ULLAH¹, (Senior Member, IEEE)

Department of Electrical and Computer Engineering, Florida Polytechnic University, Lakeland, FL 33805, USA

Corresponding author: Muhammad Sana Ullah (mullah@floridapoly.edu)

ABSTRACT Over a few decades, a remarkable amount of research has been conducted in the field of speech signal processing particularly on deception detection for security applications. In this study, a comprehensive review on recent machine learning approaches using verbal and non-verbal features is presented for deception detection. A brief overview on different feature extraction techniques, the results of recognition rate, and computational time based on machine learning methods are summarized in a tabular format. In addition, numerous datasets used as primary sources of deception detection in the review articles are also presented in this work. Key findings from the reviewed articles are summarized and a few major issues related to deception detection approaches are examined. A statistical analysis which conducted by extracting the significant information from the eighty-eight (88) scientific papers over the last thirty (30) years are provided in this review paper. The results emphasize on the trends of research in deception detection as well as further research opportunities for researchers as a part of continuous progress.

INDEX TERMS Deception detection, machine learning, non-verbal features, principal component analysis, verbal features.

I. INTRODUCTION

Detecting human emotion has peaked researcher's interest for generations. However, how well humans or machines ultimately perform the task of deceptive speech remains a challenging question for a criminal investigation. In the case of a speaker under stress, increased activation of the sympathetic or the parasympathetic nervous system is observed to occur when a speaker is angry, fearful, or sad. This increased activation leads to changes in heart rate, blood pressure, and muscle activity [3]. Consequently, the articulatory and respiratory movements are affected by speech production [3], [4]. Therefore, it is important to address these issues in a collective way for the experts who works in the fields of law enforcement, education, health care, government agencies, border crossings, military screenings, regular job screenings, telecommunications, informants at embassies and consulates around the world [3]–[6]. It well known that human speech has emotion and nonlinguistic information that encoded in

it where deception is one of them. In fact, deception is incorporated in everyday interactions, yet it is challenging for untrained and trained professionals to accurately detect it without the use of intrusive measures [1], [2].

Nowadays security is a requirement for all systems and is incorporated in everyday interactions. An increased need for more efficient artificial intelligence security systems has arisen to execute larger and more powerful tasks at a higher productivity rate. So, what is deception? Deception can be described as intentionally causing an individual to accept false statements as one(s) that are true. From a psychological perspective, an individual is being deceptive when subconscious or conscious movements present themselves including shortened length of speech, a flushed face, changes in the individual's voice frequency, avoidant eye contact, changes in the diameter of the eye pupil, as well as presenting a more rigid body [4], [5]. Traditional methods have attempted to take advantage of these deceptive indicators to detect deception with a relatively high accuracy rate through the combinatory use of various devices such as the polygraph, cardiovascular activity monitor governed by the sympathetic

The associate editor coordinating the review of this manuscript and approving it for publication was Donato Impedovo¹.

and parasympathetic nervous systems (i.e., blood pressure), heartbeat sensors, strain gauges to measure respiratory rate, and electrodermal activity (i.e., sweatiness of fingertips) [6]. The polygraph measures some human responses like respiratory rate, electro-dermal activity, heart rate, and others through direct contact, which can lead to numerous challenges and complexities in terms of implementation. Monitoring systems of this nature are intrusive, require the subject to be a cooperative, and experienced interviewer with years of training to operate and accurately perform the polygraph examinations.

Over the years, methods to analyze and recognize non-verbal and verbal feature characteristics have been developed to detect deception. It is well-known that there is a natural intuition to pick up on a speaker's emotions and purposeful language to detect some form of deception when listening. When deciding upon whether a speaker is being deceptive or truthful, most listeners have access to facial expressions as well as involuntary muscle movements besides the voice itself. This paper has significant implications in the telecommunication field and can be integrated to route 911 and other emergency calls. It can be used to address the emotional state of the caller in high-stress environments for telephone response surveys to improve upon the robustness of the system.

This paper provides an overview on deception detection techniques and classification methods specifically Linear Regression [1], J48 Decision Tree [1], [72], Multi-Layer Perceptron [1], Gradient Boosting Decision Tree [1], Levenberg-Marquardt [6], [44], BFGS Quasi-Newton [6], Long Short Term Memory [6], [39], [41], [58], Support Vector Machine [1], [13], [15], [39], [41], [45], [46] [58], [79], GMM [13], RVM [15], ensemble [39], Random Forest Decision Tree classifier [73], improved dense trajectories and OpenFace [79] and Ripper rule induction classifier [74]. In addition, non-verbal features like thermal facial analysis [7], [23], [32], [36], [38] eye movement analysis [8], [22], [26]–[28], visual cues [9], [10], text analysis with BERT [16], brain activity analysis [21], [33], [35], multimodal features [24], [29], infrared imaging [31], and facial expressions [18], [25], [27], [30], [37] have been used for deception detection research and reported encouraging results. Similarly, verbal speech features such as acoustic [11], STE [1], [46], time-difference energy [6], delta energy [6], energy [39], [40], MFCC and energy [39], zero crossing rate [40], [45], [46], entropy of energy [40], spectral centroid [40], spectral spread [40], spectral entropy [40], spectral flux [40], [75], spectral roll-off [40], chroma vector [40], chroma deviation [40], significant energy [42], [44], linear prediction cepstrum coefficients [43], bark energy [44], fundamental frequency [45], [46], time-difference cepstrum [6], delta cepstrum [6], acoustic spectral based Mel cepstral features with energy [13], MFCC [39], [40], [45], [46], [58], log-energy of MFCC [58], thirteen cepstral coefficients [58] are reported to reassure results for deception detection.

Furthermore, a breath and in-depth analysis of above mention features, generally known as spectral and cepstral features, are investigated thoroughly in this paper by addressing their progress, challenges and opportunities for improvement.

The remainder of this paper is organized as follows. Section II presents the contemporary feature approaches in deception detection. Within Section II, Subsections 1 and 2 discuss non-verbal and verbal features (i.e., spectral energy and cepstral features) used for deception detection and how they are generally developed. Section III discusses the Principal Component Analysis (PCA) along with previous research results after it was applied to various data sets to improve recognition rates. Section IV presents various feature classification methods, their respective recognition rate results, and time duration for both non-verbal and verbal features used for deception detection. Section V concludes the paper and discusses potential future research directions.

II. CONTEMPORARY FEATURE APPROACHES IN DECEPTION DETECTION

In recent years, numerous avenues have been studied to test the possibility of more accurately detecting deception in humans using technological aid through the use of various non-verbal and verbal features. Researchers have been studying more innovative ways of detecting deception such as computational methods like artificial intelligence (AI), more specifically machine learning algorithms using non-verbal and verbal cues. Figure 1 that shows a systematic approach of key steps to conduct this deception detection research.

A. EXPERIMENTAL DATABASE

For conduct this comprehensive research, the authors used the database which is a collection of utterances from the audio recording of a male suspect under criminal investigation. The suspect was determined to have given deceptive statements under questioning during polygraph testing [3], [4], [6], [12], [14], [86]. Audio recordings of three sessions of polygraph testing with the same questions by the investigator and the same responses by the suspect will be used for analysis and synthesis. For reference, available two pairs of truthful utterances or ground truth (label as Q7 and Q9) and deceptive utterances (label as Q4 and Q5) of the word 'No' from each recording will be selected for preliminary investigation. These utterances are sampled at the rate of 16,000 samples per second. Figure 2 shows the graphical view of the signals for all three sessions in the time domain.

B. NON-VERBAL FEATURE

Researchers have been learning about the multitude of ways in which humans present their deceptive ruses and how to detect them without the use of intrusive measures. Pavlidis and Levine studied thermal facial analysis [7]. Nugroho, Nasrun, and Setianingsih studied detecting deception using pupil dilation and eye blink analysis with a database consisting of 30 subjects [8]. Yap, Rajoub, Ugail, and Zwiggelaar

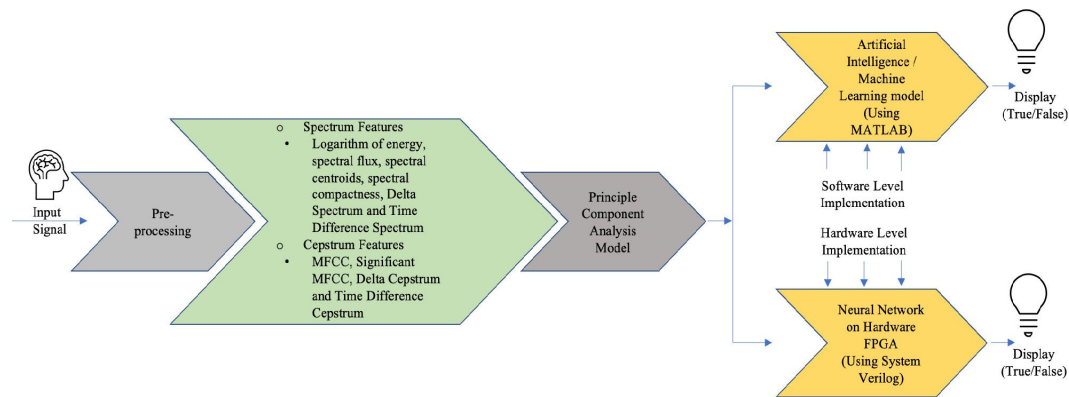


FIGURE 1. A systematic approach and a general process of deception detection research.

TABLE 1. Non-Verbal features used for detecting deception.

Ref	Feature Type	Database
[7]	Thermal Facial Analysis	-
[8]	Pupil Dilation and Eye Blink Analysis	30 Subjects
[9]	Visual Cues of Facial Behavior	-
[10]	Visual Cues (HMM-Based)	-
[15]	The Prosodic features, Nonlinear Linear Dynamic features	-
[16]	Text Analysis with BERT	-
[21]	Brain Waves	Interrogation Data
[22]	Eyeblink Pattern	5 Subjects
[23]	Facial Thermal Imaging	16 Subjects
[24]	Multimodal Deception	-
[25]	Facial Expressions	43 Subjects
[26]	Eye Blink Count and Eye Blink Duration (M)	35 Subjects
	Eye Blink Count and Eye Blink Duration (F)	15 Subjects
[27]	Facial Expressions – Eye-Blink, Eyebrow Motion, Wrinkle Occurrence, and Mouth Motion Combo	324 Video Clips of various guilty suspects
[28]	Eye Tracking	-
[29]	Multimodal Features	-
[30]	Facial Expressions	-
[31]	Infrared Imaging using Time Domain Analysis	11 Subjects
	Infrared Imaging using Frequency Domain Analysis	11 Subjects
[18]	Facial Micro-Expressions	-
[32]	Thermal Imaging	-
[33]	Brain Activities	11 Male Subjects
[34]	Head Movement Analysis	10 Subjects
[35]	Brain Activities	5 Subjects
[36]	Thermal Imaging	-
[37]	Facial Expressions	-
[38]	Stress-Induced Facial Perspiration via Thermal Imaging	40 Subjects
[76]	Keyboard Dynamics	60 Participants
[77]	Unexpected Questions and Mouse Dynamics	40 Participants
[78]	Keystroke Dynamics	190 Subjects
[79]	Facial Expression	Videotaped Interviews
[80]	Facial Displays and Hand Gesture	61 Trial Videos
[81]	Multimodal Feature (Lexical, Acoustic and Visual)	121 Trial Videos
[29]	Improved Debse Trajectories and Microexpression	104 Trial Videos
[82]	Involuntary Facial Expression	12 People (344 Facial Images)
[83]	Audio, Visual, Textual (Static and Nonstatic) and Microexpression	121 Trial Video
[84]	Macro and Micro Facial Expression	High Stakes YouTube Videos
[85]	Facial Action Units	121 Trial Videos

studied various visual cues of facial behavior to detect deception [9]. Tschepnakis *et al.* studied HMM-based visual cue

analysis to detect deception [10]. Prosodic and nonlinear linear dynamic features were studied by Zhou, Zhao, Pan,

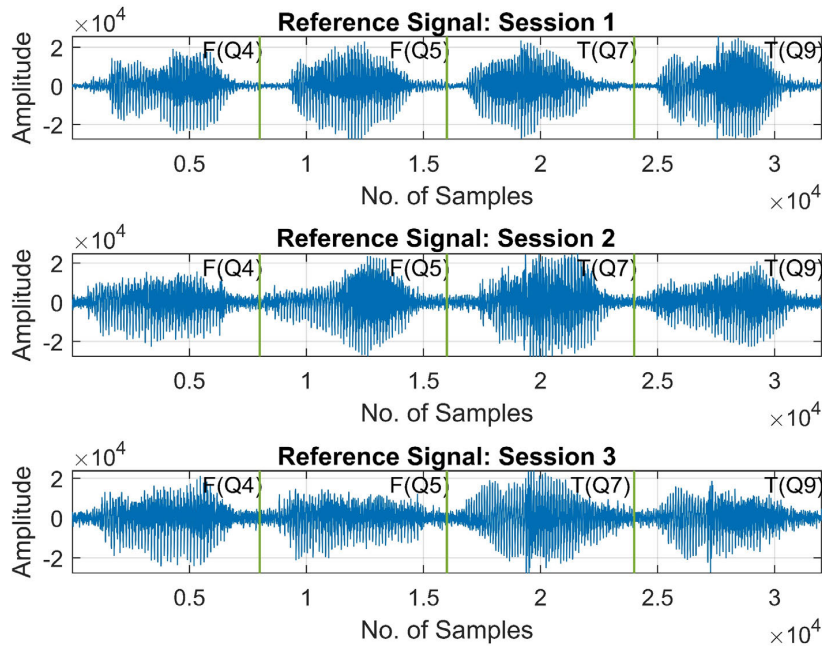


FIGURE 2. Reference speech signals.

and Shang to distinguish deception [15]. Barsever *et al.* studied detecting deception using text analysis with BERT [16]. Amir, Ahmed, and Chowdhry used interrogation data to study the brain waves and see how it performs when tasked with detecting deception [21]. Singh, Rajiv, and Chandra collected data from five (5) subjects and studied how various eyeblink patterns can be utilized to detect deception [22]. Using facial thermal analysis, Jain *et al.* used a database with data from 16 subjects to study detecting deception [23]. Thannoon, Ali, and Hashim studied the facial expressions in [25]. George, Pai, Pai, and Praharaj studied eye blink count and eye blink duration using an intra - gender database consisting of 15 male subjects and 15 female subjects to study how it can be utilized to detect deception [26].

Azhan, Zaman, and Bhuiyan studied 324 video clips of guilty suspects and studied various facial expressions like eye blink count, eyebrow motion, wrinkle occurrence, and mouth motion [27]. Eye-tracking for deception detection was studied by Lakkapragada [28]. Su, Levine, Owayjan, Kashour, Al Haddad, Fadel, Al Souki, Bailey, Demyanov, Ramamohanarao, and Leckie all studied facial analysis techniques for deception detection [18], [30], [37]. Thermal imaging was studied by both Bedoya-Echeverry *et al.* and Dcosta *et al.* [36], [38]. [34] used head movement analysis to study how it could be used to detect deception. As well as psychological and physiological evaluations on human perception [2], [13], [17]–[20] and studies on other facial, linguistic, gestural, thermal, and biometric indications. Table 1 shows the non-verbal deception detection research work previously published along with the databases used in some studies.

C. VERBAL FEATURES

As listeners, it is well-known that there is a natural intuition to pick up on a speaker's emotions and purposeful

language to detect some form of deception. When deciding upon whether a speaker is being deceptive or truthful, most listeners have access to facial expressions as well as involuntary muscle movements to consider besides the voice itself. Using speech to analyze deception provides a non-intrusive experience, especially without needing to attach sensors on a body to chart and read blood pressure, respiration, and pulse. Additionally, speech-based analysis systems can be used to analyze pre-recorded speech signals at any point in time, are inexpensive to produce, can be operated effortlessly, and can be designed to be portable devices. Over the years, researchers have been studying methods to analyze and recognize speech signal features and characteristics to detect deception. Spectral energy features and cepstral features being two types of speech signal features used to detect deception.

1) SPECTRAL ENERGY FEATURES

Spectral energy features were developed using the psychoacoustic masking property of human speech perception [47]. The psychoacoustic masking property of human speech perception is utilized to extract the spectral energy speech features, where the "irrelevant" speech signal information that typically goes undetected by the human ear is identified [47]. By modeling the non-linear perceptions of the human hearing sensation, the process of extracting spectral information can improve greatly [49]. Psychoacoustic principles including the absolute threshold of hearing, critical band frequency analysis, simultaneous masking, temporal masking, and the spread of masking along the basilar membrane were incorporated during the speech signal analysis process [47].

The absolute threshold of hearing can be described as the smallest level of a pure tone that can be detected by a listener

TABLE 2. Spectral energy verbal features used for detecting deception.

Ref	Feature	Type	Database
[1]	STE	Speech	Chinese Deception Detection corpus
[6]	Time-Difference Energy Delta Energy	Speech	Criminal Interrogation
[39]	Energy	Speech	CSC corpus
	MFCC and Energy Combination	Speech	CSC corpus
[40]	Zero Crossing Rate	Speech	CSC corpus
	Energy	Speech	CSC corpus
	Entropy of Energy	Speech	CSC corpus
	Spectral Centroid	Speech	CSC corpus
	Spectral Spread	Speech	CSC corpus
	Spectral Entropy	Speech	CSC corpus
	Spectral Flux	Speech	CSC corpus
	Spectral Rolloff	Speech	CSC corpus
	Chroma Vector	Speech	CSC corpus
	Chroma Deviation	Speech	CSC corpus
[42]	Significant Energy	Speech	Criminal Interrogation
[44]	Bark Energy	Speech	Criminal Interrogation
	Significant Energy	Speech	Criminal Interrogation
[45]	Fundamental Frequency	Speech	Interview in isolated environment
	Zero Crossing Rate	Speech	Interview in isolated environment
	Energy	Speech	Interview in isolated environment
[46]	Fundamental Frequency	Speech	Swiss Research Institute IDIAP WOLF data set
	Short-term energy	Speech	Swiss Research Institute IDIAP WOLF data set
	Zero-crossing rate	Speech	Swiss Research Institute IDIAP WOLF data set

in a noiseless environment [47]. An example being a listener hearing the ticking sound of a clock in an empty room.

Typically, the absolute threshold of hearing is calculated in terms of sound pressure level (SPL) in decibels (dB) [47]. The absolute (quiet) threshold, $T_q(f)$ is approximated [47], [50] by the non-linear function that is shown in (1).

$$T_q(f) = 3.64 \left(\frac{f}{1000} \right)^{-0.8} - 6.5e^{-0.6 \left(\frac{f}{1000} - 3.3 \right)^2} + 10^{-3} \left(\frac{f}{1000} \right)^4 \quad (\text{dB SPL}) \quad (1)$$

This equation is the representation of a listener with acute hearing where f is the frequency in Hz. The absolute threshold of hearing is also related to another acoustical metric known as the dB sensation level (dB SL) [47]. Relative to a listener's individual unmasked detection threshold for the stimulus, the intensity level difference is denoted by the SL [47], [51]. The SL is used because it quantifies listener-specific audibility instead of an absolute level [47].

Simultaneous masking is a type of auditory masking that occurs in the frequency domain when two sounds occur at the same duration, one of which is unwanted and one of which is wanted but inaudible [47], [52]. Auditory masking is when a louder sound becomes undetectable to the human ear due to a weaker but audible sound [52]–[54]. Masking is a process where one sound is virtually inaudible due to the presence of another sound [47]. From the perspective of the frequency domain, phase relationships between stimuli as well as the relative shapes of masker and the maskee

magnitude spectra are used to establish to what extent the presence of particular spectral energy will mask the presence of other spectral energy [47]. An alternative way to explain this concept is to understand how the stronger noise (tone masker) effectively blocks the detection of the weaker signal at the critical band location by creating an excitation with an adequate strength on the basilar membrane [47].

Non-simultaneous masking or temporal masking is another type of auditory masking that occurs when a sound is made inaudible by another sound, that either immediately precedes or immediately follows the original sound [52], [55]. Pre-masking is one type of temporal masking that immediately precedes the presence of a masker [52]. Post-masking is another type of temporal masking that immediately follows the masker and obscures sound [52]. With the purpose of perceptual coding, a listener will neglect to perceive signals below the elevated audibility thresholds produced by a masker when unexpected audio signal transients generate pre-masking and post-masking regions in time [47].

Using the absolute threshold of hearing, a modified version known as the detection threshold is used for spectrally complex quantization noise [47]. Its shape fluctuating at any given time based on the stimuli present, thus the detection threshold being a time-varying function of the input signal [47]. Based on how the human ear naturally performs spectral analysis, the threshold estimation is calculated [47].

First, along the basilar membrane in the cochlea or inner ear, a frequency-to-place transformation occurs [47], [56]. When a frequency-to-place transformation transpires, an acoustic stimulus produces a sound wave that moves the

ossicular bones and the eardrum attached to it [47]. The mechanical vibrations at the oval window, are then transferred to the cochlea which is a fluid-filled, spiral-shaped structure that happens to hold the coiled basilar membrane [47]. Along the length of the basilar membrane, the cochlear structure produces travel waves that were previously excited by the mechanical vibrations [47]. At frequency-specific membrane positions, these travel waves generate peak responses that “tune” various neural receptors, connected along the length of the basilar membrane, to various frequency bands depending on their locations [47]. From the oval window, the traveling wave on the basilar membrane then propagates until the traveling waves approach the area with a resonant frequency near that of the stimulus frequency for sinusoidal stimuli [47]. After the traveling wave slows down and the magnitude increases to a peak, the traveling wave then decays rapidly beyond the peak [47]. For the stimulus frequency, the location of the peak is known as the “best place” and the frequency that best excites a particular place is known as the “best frequency” [47]. With that, the frequency-to-place transformation occurs. From a signal-processing perspective, the cochlea can be viewed as a band of highly overlapping bandpass filters due to the frequency-to-place transformation [47]. The magnitude responses are nonlinear and asymmetric [47]. The bandwidth of the cochlear filter passbands is nonuniform and increases along with the increasing frequency [47]. Additionally, the cochlear filter passbands are quantified by the “critical bandwidth” that is a function of frequency [47]. Critical bandwidth is when the perceived intensity of sound remains constant at a constant SPL for a narrow-band noise source [47]. It remains constant when the noise bandwidth is increased up to the critical bandwidth, although when the loudness begins to increase for any SPL beyond the critical bandwidth [47]. Essentially, the loudness level remains constant as long as the noise energy stays within a singular cochlear critical bandwidth and it increases when the noise energy is forced into the adjacent cochlear critical bandwidth [47].

To enumerate the cochlear filter passbands, the critical bandwidth is calculated as a function of frequency [47]. The critical bandwidth, $BW_c(f)$ is calculated approximately using the non-linear function shown in (2) where frequency, f is in Hz.

$$BW_c(f) = 25 + 75 \left[1 + 1.4 \left(\frac{f}{1000} \right)^2 \right]^{0.69} \quad (\text{Hz}) \quad (2)$$

Frequency in Hz is converted to the Bark scale, $z(f)$ for analysis purposes using (3) where one critical band distance is referred to as “one Bark” [47].

$$z(f) = 13 \tan^{-1}(0.00076f) + 3.5 \tan^{-1} \left[\left(\frac{f}{7500} \right)^2 \right] \quad (\text{Bark}) \quad (3)$$

The first critical band started with the resolution frequency (DF) to exclude DC. Critical bandwidth tends to

remain constant from about 100 Hz up to 500 Hz and increases to approximately 20 percent of the center frequency above 500 Hz [44], [48]. Additionally, the non-simultaneous and simultaneous masking phenomena are induced by the auditory time-frequency analysis in the critical band filter bank to shape the coding distortion spectrum [47]. The perceptual models allocate bits for signal components so that the quantization noise is shaped to exploit the detection thresholds that are determined by the energy within a critical band, for a complex sound [47].

Fan *et al.* studied detecting deception using speech signals by extracting the short-time energy (STE) feature among others, from the Chinese Deception Detection corpus [1]. Fernandes *et al.* used a criminal interrogation data set to extract the time-difference energy feature and the delta energy feature to detect deception [6]. Xue *et al.* using the CSC corpus extracted MFCC and energy features to study how using the features singularly and combinatory effect the deception detection recognition rate [39]. Desai *et al.* used the Columbia- SRI-Colorado (CSC) corpus as their database to research detecting deception using combinations of several different spectral energy features including zero-crossing rate, energy, the entropy of energy, spectral flux, spectral roll-off, chroma vector, and chroma deviation [40]. Cosetl and Lopez used a criminal interrogation database to extract significant energy feature and use it to distinguish between deceptive and non- deceptive speech [42]. Ullah and Gopalan extracted the Bark energy and significant energy features from stressed speech signals to detect deception from a criminal interrogation database [44]. Srivastava and Dubey collected speech signal data from an interview they conducted in an isolated environment to study detecting deception using the fundamental frequency, zero-crossing rate, and energy features [45]. Tao *et al.* used the fundamental frequency, zero-crossing rate, and energy features to study how it affects deception detection while using the Swiss Research Institute IDIAP WOLF data set [46]. Table 2 shows the spectral energy features and the databases used in previous research work for deception detection.

2) CEPSTRAL FEATURES

In general, cepstral features are developed based on the MFCC feature. MFCCs are computed by summing up the weighted log energy magnitudes in a band around a center frequency as shown in (4), where $n = 1, 2, \dots, K$ is the number of cepstral coefficients, K is equal to the number of band index and S_k represents the Hamming window function.

$$MFCC_n = \sum_{k=1}^K \log_{10}(S_k) \cos \left[n \left(k - \frac{1}{2} \right) \frac{\pi}{K} \right] \quad (4)$$

Deception detection based on extracted cepstrum features was studied to understand how speech features can be used to detect human emotion and deception. Cepstral representation of an utterance provides a depiction of the local spectral properties of the signal [2], [7], [22]. When

TABLE 3. Cepstral verbal features used for detecting deception.

Ref	Feature	Type	Database
[6]	Time-Difference Cepstrum Delta Cepstrum	Speech	Criminal Interrogation
[13]	Acoustic Spectral-based Mel cepstral features with energy	Speech	Columbia-SRI-Colorado (CSC) corpus
[39]	MFCC	Speech	Columbia-SRI-Colorado (CSC) corpus
	MFCC, and Energy Combination	Speech	Columbia-SRI-Colorado (CSC) corpus
[40]	MFCC	Speech	Columbia-SRI-Colorado (CSC) corpus
[45]	MFCC	Speech	Interview in isolated environment
[46]	MFCC	Speech	Swiss Research Institute IDIAP WOLF data set
[58]	MFCC	Audio	Real-Life Trial Data including 121 videos
	Log-Energy of MFCC	Audio	Real-Life Trial Data including 121 videos
	13 Cepstral Coefficients	Audio	Real-Life Trial Data including 121 videos

analyzing deceptive speech using cepstral speech characteristics, it exhibits an increased amplitude, decreased speech duration, and increased fundamental frequency [4].

Detecting deception using cepstral features has limited research. The delta cepstrum features added to the static MFCC features strongly improves speech recognition [59]. Wang *et al.* studied speech features including MFCC and energy features using the CSC corpus [41]. Ullah *et al.* presented the results of detecting deception through the process of analyzing human speech signals and their extracted cepstrum features [4], [6]. Chowdhury *et al.* studied the MFCCs and other speech features in their research work [14]. Chowdhury *et al.* studied the effects of using characteristics of speech to detect deception and noted an increased duration of speech, fundamental frequency, and amplitude when a person was being deceptive [14]. Graciarena *et al.* studied the results of detecting deceptive and non-deceptive speech based on prosodic, lexical, and acoustic features using the CSC corpus [13]. P. Benson in [57] obtained and analyzed an audiotape of a pilot's speech during a serious aircraft malfunction, engine failure of the single-engine F-16. This investigation revealed that speech under stress is shorter and simpler than that of normal speech. Gopalan and Wenndt in [12] studied the initial results of analysis of speech features for speech under stress and for detecting deception from speech utterances of a criminal suspect using Benson's research as a steppingstone. Using the CSC corpus, Desai *et al.* extracted the MFCC speech feature among other features and used it for detecting deception [40]. In [45], Srivastava *et al.* extracted the MFCC speech features from their deception detection database that they created using the data from interviews they conducted in an isolated environment. Tao *et al.* studied speech deception detection using MFCC features and the Swiss Research Institute data set. In [58], Venkatesh *et al.* studied one hundred and twenty-one real-life trial videos to extract and examine the MFCC, Log-Energy of MFCC, and thirteen cepstral coefficient speech features to detect deception. Table 3 highlights previous research work for deception detection using various cepstral features.

III. PRINCIPAL COMPONENT ANALYSIS

While working on data deception, it was observed that large datasets often make it challenging to interpret the results more accurately [6]. The incorporation of the principal component analysis (PCA) provided a solution by reducing the data dimensions while increasing the interpretability of the data and minimizing the loss of information [6]. The PCA is a type of reduction method that takes into consideration the original dataset as rows representing elements in high dimensional space [60]. The rows are arranged to directions that characterize the optimal set of features [60]. By constructing a group of new latent variables, the PCA is then able to reduce the original data dimensions [6]. From the new mapping space, the main variation information is then extracted along with the statistical features [6]. The original data can then construct the new solution of the spatial features [6]. To reduce the dimensions of the projection space, the variables in the new mapping space are composed of linear combinations of the original dataset [6]. The correlation between variables is then eliminated and the complexity of the principal characteristic analysis is simplified due to the statistical eigenvectors in the projection space being orthogonal to each other [61].

Fernandes and Ullah proposed using the PCA to improve their speech-based deception detection recognition rate results using various speech features [6]. Using the time - difference energy speech feature, they achieved an 8.34% increase in recognition rate, and while using the time- difference cepstrum feature, they achieved an 8.33% increase in recognition rate [6]. While using the delta energy speech feature, they achieved a 12.5% increase in recognition rate and a 29.17% increase using the delta cepstrum speech feature [6]. Roopa and Asha proposed using the principal component analysis (PCA) to improve their diabetes disease prediction approach and achieved a 6.03% increase [60]. For their underwater image recognition study, Bi and Du proposed using the PCA to improve their image recognition rate and achieved a 20.3% increase after applying the PCA to their data [62]. To solve the irregular packing problem, Gua *et al.* proposed a packing algorithm based on the PCA methodology which resulted in an increased filling rate, decreased packing time,

TABLE 4. Principal component analysis based previous research.

Ref	Type of Study	Recognition Rate Before PCA	Recognition Rate After PCA
[6]	Deception Detection – Time-Difference Energy Speech Feature	91.66%	100%
[6]	Deception Detection – Delta Energy Speech Feature	75%	87.50%
[6]	Deception Detection – Time-Difference Cepstrum Speech Feature	83.33%	91.66%
[6]	Deception Detection – Delta Cepstrum Speech Feature	70.83%	100%
[60]	Diabetes Disease Prediction	76.07%	82.10%
[62]	Underwater Image Recognition PCA	63.37%	83.67%
[63]	Irregular Packing - Experiment 2 Example 1	74.00%	79.09%
	Irregular Packing - Experiment 2 Example 2	75.27%	81.37%
[64]	Genetic Study - Heart Data	84.01%	85.37%
	Genetic Study Adult Data	79.07%	91.21%

and increased packing number as compared to the MGA method [63]. Zheng *et al.* proposed a PCA-based support vector classifier and noted an increased identification rate in their heart and adult data sets, as compared to the conventional support vector classifier [64]. Table 4 shows the previous research work conducted using the PCA and how it affected the recognition rate results. Overall, applying the PCA does show an increased rate in recognition results.

IV. FEATURE CLASSIFICATION METHODS USED IN DECEPTION DETECTION

In many speech processing tasks, deep neural networks have been successfully used in speaker verification [65], [66], speech enhancement [67], [68], and speech recognition [69]–[71], deception detection [4], [44], and emotion recognition [14]. Fan *et al.* constructed and used a Chinese corpus consisting of 15 male and 15 female recordings [1]. They extracted four kinds of speech features from the database including STE, pitch, format, and duration for male and female subjects [1]. Using logic regression (LR), J48 decision tree, multi-layer perceptron (MLP), SVM, and gradient boosting decision tree (GBDT) to test the effectiveness of using a combination of the features for deception detection [1]. The highest rate for both genders was achieved using the GBDT classification method with recognition rates ranging between 82% and 85% [1]. The time-difference spectral energy feature achieved a 79.16% recognition rate before applying the PCA, and 100% after applying the PCA using the Levenberg-Marquardt feature matching technique.

Using the LSTM feature matching technique, the feature achieved a 91.66% recognition rate before applying the PCA and 100% after applying the PCA. Using the BFGS Quasi-Newton feature matching technique, the feature achieved a 75% recognition rate before applying the PCA and 100% after applying the PCA. The delta spectral energy feature achieved a 75% recognition rate before applying the PCA, and 87.5% after applying the PCA using the Levenberg-Marquardt feature matching technique. Using the LSTM feature matching technique, the feature achieved a 58.3% recognition rate before applying the PCA and 83.33% after applying the PCA. Using the BFGS Quasi-Newton feature matching technique, the feature achieved a 50% recognition

rate before applying the PCA and 87.5% after applying the PCA. The time-difference cepstrum feature achieved an 83.33% recognition rate before applying the PCA, and 79.16% after applying the PCA using the Levenberg-Marquardt feature matching technique. Using the LSTM feature matching technique, the feature achieved a 91.66% recognition rate before applying the PCA and 100% after applying the PCA. Using the BFGS Quasi-Newton feature matching technique, the feature achieved an 87.50% recognition rate before applying the PCA and 100% after applying the PCA. The delta cepstrum feature achieved a 79.16% recognition rate before applying the PCA, and 91.66% after applying the PCA using the Levenberg-Marquardt feature matching technique. Using the LSTM feature matching technique, the feature achieved a 50% recognition rate before applying the PCA and 75% after applying the PCA. Using the BFGS Quasi-Newton feature matching technique, the feature achieved a 70.83% recognition rate before applying the PCA and 100% after applying the PCA.

Graciarena *et al.* reported on distinguishing deceptive speech from non-deceptive speech using the CSC corpus and various classification models and features [13]. They computed 215 prosodic features including pitch, energy, and duration, 20 lexical features including filled pause counts, dialog act labels, and syntax-based features, as well as acoustic features including spectral-based Mel cepstral features with energy, simple delta features, double delta features, and triple delta features [13]. Graciarena *et al.* achieved the highest recognition rate using a combination of the acoustic and prosodic features as input into the GMM/SVM models with a recognition rate of 64.4% [13]. Using relevance vector machine (RVM) and non-linear dynamic features, Zhou *et al.* proposed an intra-gender deception detection approach [15]. With a combination of the prosodic and non-linear linear dynamic features of a male and female subject, they were able to achieve the top recognition rate of 70.3% and 70.15%, respectively, using the RVM as compared to SVM or the radial basis function neural network (RBFNN) models [15]. Venkatesh *et al.* proposed the extraction of MFCC, cepstral coefficients, and log-energy of MFCC to detect deception using various classification methods including SVM and LSTM [58]. Using SVM, they achieved the

TABLE 5. Features and classification methods based comparison results.

Ref	Feature	Classification Method	Computational Time	Recognition Rate
[1]	STE, Pitch, Format, and Duration features (M)	LR	~	76.37%
	STE, Pitch, Format, and Duration features (M)	J48 Decision Tree	~	76.75%
	STE, Pitch, Format, and Duration features (M)	MLP	~	75.22%
	STE, Pitch, Format, and Duration features (M)	SVM	~	74.41%
	STE, Pitch, Format, and Duration features (M)	GBDT	~	82.36%
	STE, Pitch, Format, and Duration features (F)	LR	~	77.82%
	STE, Pitch, Format, and Duration features (F)	J48 Decision Tree	~	81.93%
	STE, Pitch, Format, and Duration features (F)	MLP	~	77.05%
	STE, Pitch, Format, and Duration features (F)	SVM	~	76.04%
	STE, Pitch, Format, and Duration features (F)	GBDT	~	84.56%
[6]	Time-Difference Energy (Pre-PCA)	Levenberg-Marquardt	7-20 epochs	79.16%
	Time-Difference Energy (Post-PCA)	Levenberg-Marquardt	5-18 epochs	100%
	Time-Difference Energy (Pre-PCA)	LSTM	2000 epochs	91.67%
	Time-Difference Energy (Post-PCA)	LSTM	2000 epochs	100%
	Time-Difference Energy (Pre-PCA)	BFGS Quasi-Newton	15-20 epochs	75%
	Time-Difference Energy (Post-PCA)	BFGS Quasi-Newton	12-20 epochs	100%
	Delta Energy (Pre-PCA)	Levenberg-Marquardt	7-20 epochs	75%
	Delta Energy (Post-PCA)	Levenberg-Marquardt	5-18 epochs	87.5%
	Delta Energy (Pre-PCA)	LSTM	2000 epochs	58.33%
	Delta Energy (Post-PCA)	LSTM	2000 epochs	83.33%
	Delta Energy (Pre-PCA)	BFGS Quasi-Newton	15-20 epochs	50%
	Delta Energy (Post-PCA)	BFGS Quasi-Newton	13-20 epochs	87.5%
	Time-Difference Cepstrum (Pre-PCA)	Levenberg-Marquardt	6-20 epochs	83.33%
	Time-Difference Cepstrum (Post-PCA)	Levenberg-Marquardt	5-18 epochs	83.33%
	Time-Difference Cepstrum (Pre-PCA)	LSTM	2000 epochs	83.33%
	Time-Difference Cepstrum (Post-PCA)	LSTM	2000 epochs	91.66%
	Time-Difference Cepstrum (Pre-PCA)	BFGS Quasi-Newton	15-20 epochs	87.5%
	Time-Difference Cepstrum (Post-PCA)	BFGS Quasi-Newton	14-20 epochs	70.83%
	Delta Cepstrum (Pre-PCA)	Levenberg-Marquardt	5-20 epochs	79.16%
	Delta Cepstrum (Post-PCA)	Levenberg-Marquardt	5-18 epochs	91.66%
	Delta Cepstrum (Pre-PCA)	LSTM	2000 epochs	50%
Delta Cepstrum (Post-PCA)	LSTM	2000 epochs	75%	
Delta Cepstrum (Pre-PCA)	BFGS Quasi-Newton	17-20 epochs	70.83%	
Delta Cepstrum (Post-PCA)	BFGS Quasi-Newton	11-20 epochs	100%	
[8]	Pupil Dilation and Eye Blink Analysis	~	~	84%
[13]	Acoustic Features: Spectral-based Mel cepstral features with energy, plus simple, double, and triple delta features	GMM	~	62.1%
	Prosodic Features: 215 Features	SVM	~	62.7%
	Prosodic/Lexical Features: 215 Prosodic and 20 Lexical features	SVM	~	62.9%
	Acoustic and Prosodic Features	GMM and SVM	~	64.4%
	Acoustic, Prosodic, and Lexical Features	GMM and SVM	~	64.0%
[15]	The Prosodic features, NLD features (M)	RBFNN	~	42.13%
	The Prosodic features, NLD features (M)	SVM	0.2351s	68.14%
	The Prosodic features, NLD features (M)	RVM	0.0031s	70.37%
	The Prosodic features, NLD features (F)	RBFNN	~	29.13%
	The Prosodic features, NLD features (F)	SVM	0.3164s	66.87%
	The Prosodic features, NLD features (F)	RVM	0.0042s	70.15%
[16]	Text Analysis with BERT	BERT in a GAN	~	93.6%
[23]	Facial Thermal Imaging	BAC	~	83.5%
[27]	Facial Expressions – Eye-Blink, Eyebrow Motion, Wrinkle Occurrence, and Mouth Motion Combo	~	~	76.2%
[30]	Facial Expression	~	~	76.92%
[31]	Infrared Imaging using Time Domain Analysis	~	~	69%
	Infrared Imaging using Frequency Domain Analysis	~	~	84%
[18]	Facial Micro-Expressions	~	~	85%
[32]	Thermal Imaging	~	~	64%
[33]	Brain Activities	~	~	70.83%
[34]	Head Movement Analysis	~	~	58.25%
[35]	Brain Activities	~	~	-
[36]	Thermal Imaging	~	~	79.2%
[37]	Facial Expressions	~	~	70.26%

TABLE 5. (Continued.) Features and classification methods based comparison results.

[38]	Stress-Induced Facial Perspiration via Thermal Imaging	~	~	80%
[39]	MFCC	SVM	~	51.8%
	Pitch	SVM	~	50.7%
	Energy	SVM	~	50.2%
	MFCC	LSTM	~	54.6%
	Pitch	LSTM	~	51.8%
	Energy	LSTM	~	47.0%
	MFCC	Ensemble	~	54.6%
	Pitch	Ensemble	~	51.1%
	Energy	Ensemble	~	53.0%
	MFCC and Energy	Ensemble	~	55.8%
	MFCC, Pitch, and Energy	Ensemble	~	53.4%
[40]	LexRNN: Lexical features	LSTM	~	81.03%
	HybridRNN: Speaker-dependent and lexical features	LSTM	~	84.11%
	AudioRNN: Speaker-dependent acoustic features	LSTM	~	62.59%
[41]	MFCC	SVM	~	51.8%
	Energy	SVM	~	50.2%
	MFCC	LSTM	~	54.6%
	Energy	LSTM	~	47.0%
[44]	Bark Energy	Levenberg-Marquardt	3-20 epochs	83.33%
	Significant Energy	Levenberg-Marquardt	3-20 epochs	66.67%
[45]	Fundamental Frequency, Zero-Crossing Rate, MFCC, Frames Function, and Energy	SVM	~	100%
	Fundamental Frequency, Zero-Crossing Rate, MFCC, Frames Function, and Energy	ANN	~	93.33%
[46]	Fundamental Frequency, Short-Term Energy, Zero-Crossing Rate, and MFCC	SVM	~	82.47%
[58]	MFCC	SVM	~	64%
	Cepstral Coefficients	SVM	~	66%
	Log-Energy of MFCC	SVM	~	72%
	MFCC	LSTM	~	46%
	Cepstral Coefficients	LSTM	~	58%
	Log Energy of MFCC	LSTM	~	50%
[72]	Critical "hot spot" segments identified to tell if a speaker is truthful or lying	Bagging, AdaBoost, and J48 combination	~	68.6%
[73]	Lexical and Speech	RFDT Classifier	~	50.45%
[74]	Acoustic/Prosodic, Lexical, and Speaker-Dependent Combination	Ripper Rule Induction Classifier	~	66.4%
[76]	Keyboard Dynamic	Logistic	~	100%
	Keyboard Dynamic	SVM	~	90%
	Keyboard Dynamic	LMT	~	90%
	Keyboard Dynamic	Random Forest	~	100%
[77]	Unexpected Questions and Mouse Dynamics	Random Forest	~	92.5%
	Unexpected Questions and Mouse Dynamics	Logistics	~	95%
	Unexpected Questions and Mouse Dynamics	SVM	~	90%
	Unexpected Questions and Mouse Dynamics	LMT	~	92.5%
[79]	Facial Micro Expressions	SVM & Deep Neural Networks	~	57%
[80]	Unigrams	Decision Tree	~	60.33%
	Unigrams	Random Forest	~	56.19%
	Bigrams	Decision Tree	~	53.71%
	Bigrams	Random Forest	~	51.20%
	Facial Displays	Decision Tree	~	70.24%
	Facial Displays	Random Forest	~	76.03%
	Hand Gesture	Decision Tree	~	61.98%
	Hand Gesture	Random Forest	~	62.80%

TABLE 5. (Continued.) Features and classification methods based comparison results.

[81]	Lexical Feature	SVM	~	56.12%
	Acoustic Features (Prosody, Energy, Voice Probability, Spectral and Cepstral)	SVM	~	34.23%
	Visual (OpenFace)	SVM	~	67.20%
	Lexical + Acoustic + Visual (Feature – Level Fusion)	SVM	~	78.95%
	Lexical + Acoustic + Visual (Decision – Level Fusion)	SVM	~	76.12%
	Lexical + Acoustic + Visual (Utterance based Feature Fusion)	SVM	~	74.02%
[29]	Improved Debse Trajectories (IDT)	Linear SVM	~	77.31%
	Micro expression	Random Forest	~	80.64%
	Transcripts	Naïve Bayes	~	66.25%
	MFCC	Kernel SVM	~	81.71%
	IDT+Microexpression+Transcript+MFCC	Linear SVM	~	87.73%
[82]	Facial Expression (Angry)	LIBSVM	~	86.02%
	Facial Expression (Enjoyment)	LIBSVM	~	73.16%
	Facial Expression (Fear)	LIBSVM	~	80.46%
	Facial Expression (Sadness)	LIBSVM	~	90.15%
[83]	Audio	MLP	~	52.38%
	Visual	MLP	~	93.08%
	Textual (Static)	MLP	~	80.16%
	Textual (Non-Static)	MLP	~	90.24%
	Micro Expression	MLP	~	76.19%
[84]	Macro Facial Expression	Random Forest	~	73.85%
	Micro Facial Expression	Random Forest	~	56.92%
	Macro & Micro Facial Expression	Random Forest	~	76.92%
[85]	Facial Action Units	RBF - SVM	~	76.84%

highest recognition rate of 72% with the log-energy of MFCC feature whereas while using the LSTM, they achieve the highest recognition rate of 58% with the cepstral coefficients feature [58]. Ullah and Gopalan proposed extracting the Bark energy and significant energy features to detect deception [44]. Using the significant energy feature, they achieved a 66.67% recognition rate with the Levenberg-Marquardt classification method [44]. Using the Bark energy feature, they achieved a higher recognition rate at 83.33% with the Levenberg- Marquardt classification method [44]. Using the LSTM classification method, Desai *et al.* achieved recognition rates of 81.03% using the LexRNN features, 84.11% using the HybridRNN features, and 62.59% using the AudioRNN features [40]. Using the SVM classification method and MFCC speech feature, Wang *et al.* were able to achieve a 51.8% recognition rate whereas when using the LSTM classification method and the same feature, they were able to achieve a 54.6% recognition rate [41].

When using the energy speech feature and the SVM classification method, Wang *et al.* were able to achieve a 50.2% recognition rate whereas when using the LSTM classification method and the same speech feature, they were able to achieve a 47% rate of recognition [41]. Enos *et al.* proposed detecting deception through the use of critical “hot spot” segments in the speech where they achieved a recognition rate of 68.6% using a combination of bagging, AdaBoost, and J48 [72]. Using various lexical and speech features, Warnita and Lestari were able to achieve a 50.45% recognition rate using the random forest decision tree (RFDT) classifier [73]. Tao *et al.* achieved an 82.47% recognition rate using an SVM classifier and a combination of various speech features

including fundamental frequency, STE, zero-crossing rate, and MFCC as inputs [46]. Kumar, Kim, and Stern achieved a recognition rate of 100% using SVM and using a combination of the fundamental frequency, zero-crossing rate, MFCC, frames function, and energy [45]. They were also able to achieve a recognition rate of 93.33% using an artificial neural network (ANN) for the same set of features [45]. Xue *et al.* proposed using MFCC, pitch, and energy features to detect deception using various classification methods [39]. Using SVM, Xue *et al.* were able to achieve a 51.8% recognition rate using the MFCC feature. They were also able to achieve a 54.6% recognition rate using LSTM and the MFCC feature [39]. Although while using the ensemble classification method, they were able to achieve the highest recognition rate of 55.8% using the MFCC and energy features [39]. Using the *Ripper* rule induction classifier, Hirschberg *et al.* were able to achieve a recognition rate of 66.4% using a combination of acoustic/prosodic, lexical, and speaker-dependent features [74]. Table 5 presents the previous deception detection studies along with what database and classification methods were used in the studies as well as their recognition rate results.

V. CONCLUSION

This study presented a comprehensive review of various verbal and non-verbal features extracted for deception detection as well as the recognition rate results of the various feature matching techniques used. Overall, the time-difference energy feature extracted and developed by Fernandes and Ullah in [6] showed the highest recognition rate of 100% after applying the PCA using three unique feature matching

techniques. Thus, the time-difference energy feature could be a potential feature for speech-based deception detection. The authors suggested further research using more speech utterances from a multitude of speakers, which was hard to obtain, can confirm the results of the proposed feature classification methods in detecting deception [6].

The limitations recognized in the previous research work include the size of the database and the type of the database. A large intragender deception database was one recommended solution. Another limitation highlighted was limiting the analysis of the data to single-word utterances as compared to analyzing full sentences in certain studies.

Studying the use of a field-programmable gate array (FPGA) with the post PCA data is an implementation method where the results in this comprehensive review could be used to expand as further research direction to create a product based on a software-hardware device with improved accuracy for real-life applications.

ACKNOWLEDGMENT

The authors would like to give thanks to Mohammad Safayet Hossain for providing the LSTM MATLAB code that helped them with this comprehensive review paper.

REFERENCES

- [1] C. Fan, H. Zhao, X. Chen, X. Fan, and S. Chen, "Distinguishing deception from non-deception in Chinese speech," in *Proc. 6th Int. Conf. Intell. Control Inf. Process. (ICICIP)*, Wuhan, China, Nov. 2015, pp. 268–273.
- [2] B. A. Rajoub and R. Zwiggelar, "Thermal facial analysis for deception detection," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 6, pp. 1015–1023, Jun. 2014.
- [3] M. Sanaullah and K. Gopalan, "Distinguishing deceptive speech from truthful speech using MFCC," in *Proc. 7th Int. Conf. Circuits, Syst. Signals (WSEAS)*, Cambridge, MA, USA, Feb. 2013, pp. 167–171.
- [4] S. V. Fernandes and M. S. Ullah, "Psychoacoustic masking of delta and time-difference cepstrum features for deception detection," in *Proc. 11th IEEE Annu. Ubiquitous Comput., Electron. Mobile Commun. Conf. (UEMCON)*, New York, NY, USA, Oct. 2020, pp. 213–217.
- [5] Z. Labibah, M. Nasrun, and C. Setianingsih, "Lie detector with the analysis of the change of diameter pupil and the eye movement use method Gabor wavelet transform and decision tree," in *Proc. IEEE Int. Conf. Internet Things Intell. Syst. (IOTAIS)*, Bali, Indonesia, Nov. 2018, pp. 214–220.
- [6] 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. 78925–78935, 2021.
- [7] I. Pavlidis and J. Levine, "Thermal facial screening for deception detection," in *Proc. 2nd Joint 24th Annu. Conf. Annu. Fall Meeting Biomed. Eng. Soc. Eng. Med. Biol.*, Houston, TX, USA, Oct. 2002, pp. 1143–1144.
- [8] R. H. Nugroho, M. Nasrun, and C. Setianingsih, "Lie detector with pupil dilation and eye blinks using Hough transform and frame difference method with fuzzy logic," in *Proc. Int. Conf. Control, Electron., Renew. Energy Commun. (ICCREC)*, Yogyakarta, Indonesia, Sep. 2017, pp. 40–45.
- [9] M. H. Yap, B. Rajoub, H. Ugail, and R. Zwiggelar, "Visual cues of facial behaviour in deception detection," in *Proc. IEEE Int. Conf. Comput. Appl. Ind. Electron. (ICCAIE)*, Penang, Malaysia, Dec. 2011, pp. 294–299.
- [10] G. Tsechpenakis, D. Metaxas, M. Adkins, J. Kruse, J. K. Burgoon, M. L. Jensen, T. Meservy, D. P. Twitchell, A. Deokar, and J. F. Nunamaker, "HMM-based deception recognition from visual cues," in *Proc. IEEE Int. Conf. Multimedia Expo*, Amsterdam, The Netherlands, 2005, pp. 824–827.
- [11] 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. 76527–76534, 2018.
- [12] K. Gopalan and S. Wendt, "Speech analysis using modulation-based features for detecting deception," in *Proc. 15th Int. Conf. Digit. Signal Process.*, Cardiff, U.K., Jul. 2007, pp. 619–622.
- [13] M. Graciarena, E. Shriberg, A. Stolcke, F. Enos, J. Hirschberg, and S. Kajarekar, "Combining prosodic lexical and cepstral systems for deceptive speech detection," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, May 2006, pp. 1033–1036.
- [14] M. Sanaullah and M. H. Chowdhury, "Neural network based classification of stressed speech using nonlinear spectral and cepstral features," in *Proc. IEEE 12th Int. New Circuits Syst. Conf. (NEWCAS)*, Trois-Rivieres, QC, Canada, Jun. 2014, pp. 33–36.
- [15] 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, Mar. 2015.
- [16] D. Barsever, S. Singh, and E. Neftci, "Building a better lie detector with BERT: The difference between truth and lies," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Glasgow, U.K., Jul. 2020, pp. 1–7.
- [17] J. Hillman, A. Vrij, and S. Mann, "Um they were wearing: The effect of deception on specific hand gestures," *Legal Criminol. Psychol.*, vol. 17, no. 2, pp. 336–345, Sep. 2012.
- [18] 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 *Proc. 2nd Int. Conf. Adv. Comput. Tools Eng. Appl. (ACTEA)*, Beirut, Lebanon, Dec. 2012, pp. 33–38.
- [19] Y. Zhou, P. Tsiamyrtzis, P. Lindner, I. Timofeyev, and I. Pavlidis, "Spatiotemporal smoothing as a basis for facial tissue tracking in thermal imaging," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 5, pp. 1280–1289, May 2013.
- [20] S. Feng, R. Banerjee, and Y. Choi, "Syntactic stylometry for deception detection," in *Proc. 50th Annu. Meeting Assoc. Comput. Linguistics (ACL)*, Jan. 2012, pp. 171–175.
- [21] S. Amir, N. Ahmed, and B. S. Chowdhry, "Lie detection in interrogations using digital signal processing of brain waves," in *Proc. 3rd Int. Conf. Instrum., Commun., Inf. Technol. Biomed. Eng. (ICICI-BME)*, Bandung, Indonesia, Nov. 2013, pp. 209–214.
- [22] B. Singh, P. Rajiv, and M. Chandra, "Lie detection using image processing," in *Proc. Int. Conf. Adv. Comput. Commun. Syst.*, Coimbatore, India, Jan. 2015, pp. 1–5.
- [23] U. Jain, B. Tan, and Q. Li, "Concealed knowledge identification using facial thermal imaging," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Kyoto, Japan, Mar. 2012, pp. 1677–1680.
- [24] V. Pérez-Rosas, R. Mihalcea, A. Narvaez, and M. Burzo, "A multimodal dataset for deception detection," in *Proc. 9th Int. Conf. Lang. Resour. Eval. (LREC)*, Reykjavik, Iceland, May 2014, pp. 3118–3122.
- [25] H. H. Thannoon, W. H. Ali, and I. A. Hashim, "Detection of deception using facial expressions based on different classification algorithms," in *Proc. 3rd Sci. Conf. Electr. Eng. (SCEE)*, Baghdad, Iraq, Dec. 2018, pp. 51–56.
- [26] S. George, M. M. Manohara Pai, R. M. Pai, and S. K. Praharaj, "Eye blink count and eye blink duration analysis for deception detection," in *Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI)*, Udipi, India, Sep. 2017, pp. 223–229.
- [27] S. Azhan, A. Zaman, and M. R. Bhuiyan, "Using machine learning for lie detection: Classification of human visual morphology," M.S. thesis, Dept. Comput. Sci. Eng., Brac Univ., Dhaka, Bangladesh, 2018.
- [28] P. K. Lakkapragada, "Using eye-tracking to understand user behavior in deception detection system interaction," M.S. thesis, Dept. Comput. Sci., Missouri Univ. Sci. Technol., Columbia, MO, USA, 2016.
- [29] Z. Wu, B. Singh, L. S. Davis, and V. S. Subrahmanian, "Deception detection in videos," in *Proc. 32nd AAAI Conf. Artif. Intell. (AAAI)*, New Orleans, LA, USA, Dec. 2017, pp. 1695–1702.
- [30] L. Su and M. D. Levine, "High-stakes deception detection based on facial expressions," in *Proc. 22nd Int. Conf. Pattern Recognit.*, Stockholm, Sweden, Aug. 2014, pp. 2519–2524.
- [31] Y. Azar and M. Campisi, "Detection of falsification using infrared imaging: Time and frequency domain analysis," in *Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI)*, New Delhi, India, Sep. 2014, pp. 1021–1026.
- [32] L. Warmelink, A. Vrij, S. Mann, S. Leal, D. Forrester, and R. P. Fisher, "Thermal imaging as a lie detection tool at airports," *Law Hum. Behav.*, vol. 35, no. 1, pp. 40–48, 2011.
- [33] A. I. Simbolon, A. Turnip, J. Hutahaean, Y. Siagian, and N. Irawati, "An experiment of lie detection based EEG-P300 classified by SVM algorithm," in *Proc. Int. Conf. Autom., Cognit. Sci., Opt., Micro Electro-Mech. Syst., Inf. Technol. (ICACOMIT)*, Bandung, Indonesia, Oct. 2015, pp. 68–71.

- [34] D.-I. Noje and R. Malutan, "Head movement analysis in lie detection," in *Proc. Conf. Grid, Cloud High Perform. Comput. Sci. (ROLCG)*, Cluj-Napoca, Romania, Oct. 2015, pp. 1–4.
- [35] M. A. Rahman and M. Ahmad, "Lie detection from fNIR signal and NeuroImage," in *Proc. Int. Conf. Med. Eng., Health Informat. Technol. (MediTec)*, Dhaka, Bangladesh, Dec. 2016, pp. 1–6.
- [36] S. Bedoya-Echeverry, H. Belalcázar-Ramírez, H. Loaiza-Correa, S. E. Nope-Rodríguez, C. R. Pinedo-Jaramillo, and A. D. Restrepo-Girón, "Detection of lies by facial thermal imagery analysis," *Revista Facultad de Ingeniería*, vol. 26, no. 44, pp. 47–59, Jan. 2017.
- [37] S. Demyanov, J. Bailey, K. Ramamohanarao, and C. Leckie, "Detection of deception in the mafia party game," in *Proc. ACM Int. Conf. Multimodal Interact.*, Seattle, WA, USA, Nov. 2015, pp. 335–342.
- [38] M. Dcosta, D. Shastri, R. Vilalta, J. K. Burgoon, and I. Pavlidis, "Perinatal indicators of deceptive behavior," in *Proc. 11th IEEE Int. Conf. Workshops Autom. Face Gesture Recognit. (FG)*, Ljubljana, Slovenia, May 2015, pp. 1–8.
- [39] A. Xue, H. Rohde, and P. A. Finkelstein, "An acoustic automated lie detector," Princeton Univ., Princeton, NJ, USA, Tech. Rep. 204877090, 2019.
- [40] S. Desai, M. Siegelman, and Z. Maurer, "Neural lie detection with the CSC deceptive speech dataset," Stanford Univ., Stanford, CA, USA, Tech. Rep. 49234152, 2017.
- [41] W. Wang, X. Yu, Y. H. Wang, and R. Swaminathan, "Audio fingerprint based on spectral flux for audio retrieval," in *Proc. Int. Conf. Audio, Lang. Image Process.*, Shanghai, China, Jul. 2012, pp. 1104–1107.
- [42] R. Cabrera Cosetl and J. M. D. Baez Lopez, "Voice stress detection: A method for stress analysis detecting fluctuations on lippold microtremor spectrum using FFT," in *Proc. 21st Int. Conf. Electr. Commun. Comput. (CONIELECOMP)*, San Andres Cholula, Mexico, Feb. 2011, pp. 184–189.
- [43] E. Wong and S. Sridharan, "Comparison of linear prediction cepstrum coefficients and mel-frequency cepstrum coefficients for language identification," in *Proc. Int. Symp. Intell. Multimedia, Video Speech Process. (ISIMP)*, Hong Kong, May 2001, pp. 95–98.
- [44] M. Sanaullah and K. Gopalan, "Deception detection in speech using bark band and perceptually significant energy features," in *Proc. IEEE 56th Int. Midwest Symp. Circuits Syst. (MWSCAS)*, Columbus, OH, USA, Aug. 2013, pp. 1212–1215.
- [45] N. Srivastava and S. Dubey, "Deception detection using artificial neural network and support vector machine," in *Proc. 2nd Int. Conf. Electron., Commun. Aerosp. Technol. (ICECA)*, Coimbatore, India, Mar. 2018, pp. 1205–1208.
- [46] H. Tao, P. Lei, M. Wang, J. Wang, and H. Fu, "Speech deception detection algorithm based on SVM and acoustic features," in *Proc. IEEE 7th Int. Conf. Comput. Sci. Netw. Technol. (ICCSNT)*, Dalian, China, Oct. 2019, pp. 31–33.
- [47] T. Painter and A. Spanias, "Perceptual coding of digital audio," *Proc. IEEE*, vol. 88, no. 4, pp. 451–513, Apr. 2000.
- [48] J. S. Park, J. S. Yoon, Y. H. Seo, and G. J. Jang, "Spectral energy based voice activity detection for real-time voice interface," *J. Theor. Appl. Inf. Technol.*, vol. 95, no. 17, pp. 4305–4312, Sep. 2017.
- [49] B. Winduratna, "FM analysis/synthesis-based audio coding," in *Proc. Audio Eng. Soc. Conv. (AES)*, vol. 104, May 1998, pp. 9798–9805.
- [50] E. Terhardt, "Calculating virtual pitch," *Hearing Res.*, vol. 1, no. 2, pp. 155–182, Mar. 1979.
- [51] B. C. J. Moore, *Introduction to the Psychology of Hearing*. New York, NY, USA: Academic, 1977.
- [52] H. M. S. Naing, R. Hidayat, B. Winduratna, and Y. Miyayaga, "Psychoacoustical masking effect-based feature extraction for robust speech recognition," *Int. J. Innov. Comput., Inf. Control*, vol. 15, no. 5, pp. 1641–1654, 2019.
- [53] D. Naveen and A. J. Rani, "Implementation of psychoacoustic model in audio compression using Munich and Gammachirp wavelets," *Int. J. Eng. Sci. Technol.*, vol. 2, no. 5, pp. 1066–1072, May 2010.
- [54] H. K. Maganti and M. Matassoni, "A perceptual masking approach for noise robust speech recognition," *EURASIP J. Audio, Speech, Music Process.*, vol. 2012, no. 1, pp. 1–9, Dec. 2012.
- [55] P. Dai and I. Y. Soon, "A temporal frequency warped (TFW) 2D psychoacoustic filter for robust speech recognition system," *Speech Commun.*, vol. 54, no. 3, pp. 402–413, Mar. 2012.
- [56] E. Zwicker and H. Fastl, *Psychoacoustics Facts and Models*. Berlin, Germany: Springer-Verlag, 1990.
- [57] P. Benson, "Analysis of the acoustic correlates of stress from an operational aviation emergency," in *Proc. ESCA-NATO Tutorial Res. Workshop Speech Under Stress Condition*, Lisbon, Portugal, Sep. 1995, pp. 25-1–25-4.
- [58] S. Venkatesh, R. Ramachandra, and P. Bours, "Robust algorithm for multimodal deception detection," in *Proc. IEEE Conf. Multimedia Inf. Process. Retr. (MIPR)*, San Jose, CA, USA, Mar. 2019, pp. 534–537.
- [59] K. Kumar, C. Kim, and R. M. Stern, "Delta-spectral cepstral coefficients for robust speech recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Prague, Czech Republic, May 2011, pp. 4784–4787.
- [60] H. Roopa and T. Asha, "A linear model based on principal component analysis for disease prediction," *IEEE Access*, vol. 7, pp. 105314–105318, 2019.
- [61] Y. Zhang, B. Zhang, and Z. Wu, "Multi-model modeling of CFB boiler bed temperature system based on principal component analysis," *IEEE Access*, vol. 8, pp. 389–399, 2020.
- [62] P. Bi and X. Du, "Application of locally invariant robust PCA for underwater image recognition," *IEEE Access*, vol. 9, pp. 29470–29481, 2021.
- [63] B. Guo, Z. Liang, Q. Peng, Y. Li, and F. Wu, "Irregular packing based on principal component analysis methodology," *IEEE Access*, vol. 6, pp. 62675–62686, 2018.
- [64] Z. Chunhong, J. Licheng, and L. Yongzhao, "Support vector classifier based on principal component analysis," *J. Syst. Eng. Electron.*, vol. 19, no. 1, pp. 184–190, Feb. 2008.
- [65] E. Variani, X. Lei, E. McDermott, I. L. Moreno, and J. Gonzalez-Dominguez, "Deep neural networks for small footprint text-dependent speaker verification," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Florence, Italy, May 2014, pp. 4052–4056.
- [66] H. Lee, P. Pham, Y. Largman, and A. Y. Ng, "Unsupervised feature learning for audio classification using convolutional deep belief networks," in *Proc. 22nd Int. Conf. Neural Inf. Process. Syst. (NIPS)*, Vancouver, BC, Canada, Dec. 2009, pp. 1096–1104.
- [67] Z. Yang, J. Lei, K. Fan, and Y. Lai, "Keyword extraction by entropy difference between the intrinsic and extrinsic mode," *Phys. A, Stat. Mech. Appl.*, vol. 392, no. 19, pp. 4523–4531, 2013.
- [68] S. Gholami-Boroujeny, A. Fallatah, B. P. Heffernan, and H. R. Dajani, "Neural network-based adaptive noise cancellation for enhancement of speech auditory brainstem responses," *Signal, Image Video Process.*, vol. 10, no. 2, pp. 389–395, Feb. 2015.
- [69] Y. Xu, J. Du, L.-R. Dai, and C.-H. Lee, "An experimental study on speech enhancement based on deep neural networks," *IEEE Signal Process. Lett.*, vol. 21, no. 1, pp. 65–68, Jan. 2014.
- [70] G. E. Dahl, D. Yu, L. Deng, and A. Acero, "Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition," *IEEE Trans. Audio, Speech, Language Process.*, vol. 20, no. 1, pp. 30–42, Jan. 2012.
- [71] J. Lei, C. Zhang, Y. Fang, Z. Gu, N. Ling, and C. Hou, "Depth sensation enhancement for multiple virtual view rendering," *IEEE Trans. Multimedia*, vol. 17, no. 4, pp. 457–469, Apr. 2015.
- [72] F. Enos, E. Shriberg, M. Graciarena, J. Hirschberg, and A. Stolcke, "Detecting deception using critical segments," in *Proc. Interspeech*, Antwerp, Belgium, Aug. 2007, pp. 2281–2284.
- [73] T. Warnita and D. P. Lestari, "Construction and analysis of Indonesian-interviews deception corpus," in *Proc. 20th Conf. Oriental Chapter Int. Coordinating Committee Speech Databases Speech I/O Syst. Assessment (O-COCOSDA)*, Seoul, South Korea, Nov. 2017, pp. 1–6.
- [74] J. Hirschberg, S. Benus, J. M. Brenier, F. Enos, S. Friedman, S. Gilman, C. Girand, M. Graciarena, A. Kathol, L. Michaelis, B. Pellom, E. Shriberg, and A. Stolcke, "Distinguishing deceptive from non-deceptive speech," in *Proc. 9th Eur. Conf. Speech Commun. Technol. (INTERSPEECH)*, Lisbon, Portugal, Sep. 2005, pp. 1833–1836.
- [75] A. Ghosal and S. Dutta, "Automatic male-female voice discrimination," in *Proc. Int. Conf. Issues Challenges Intell. Comput. Techn. (ICICT)*, Ghaziabad, India, Feb. 2014, pp. 731–735.
- [76] M. Monaro, C. Galante, R. Spolaor, Q. Q. Li, L. Gamberini, M. Conti, and G. Sartori, "Covert lie detection using keyboard dynamics," *Sci. Rep.*, vol. 8, no. 1, p. 1976, Jan. 2018.

- [77] M. Monaro, L. Gamberini, and G. Sartori, "The detection of faked identity using unexpected questions and mouse dynamics," *PLoS ONE*, vol. 12, no. 5, May 2017, Art. no. e0177851.
- [78] M. Monaro, B. Singh, L. S. Davis, and V. S. Subrahmanian, "Type me the truth!: Detecting deceitful users via keystroke dynamics," in *Proc. 12th Int. Conf. Availability, Rel. Secur. (ARES)*, Reggio Calabria, Italy, Sep. 2017, pp. 1–6.
- [79] M. Monaro, S. Maldera, C. Scarpazza, G. Sartori, and N. Navarin, "Detecting deception through facial expressions in a dataset of videotaped interviews: A comparison between human judges and machine learning models," *Comput. Hum. Behav.*, vol. 127, Feb. 2022, Art. no. 107063.
- [80] V. Perez-Rosas, M. Abouelenien, R. Mihalcea, and M. Burzo, "Deception detection using real-life trial data," in *Proc. ACM Int. Conf. Multimodal Interact. (ICMI)*, Seattle, WA, USA, Nov. 2015, pp. 59–66.
- [81] M. Jaiswal, S. Tabibu, and R. Bajpai, "The truth and nothing but the truth: Multimodal analysis for deception detection," in *Proc. IEEE 16th Int. Conf. Data Mining Workshops (ICDMW)*, Barcelona, Spain, Dec. 2016, pp. 938–943.
- [82] Z. Zhang, V. Singh, T. E. Slowe, S. Tulyakov, and V. Govindaraju, "Real-time automatic deceit detection from involuntary facial expressions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Minneapolis, MN, USA, Jun. 2007, pp. 1–6.
- [83] G. Krishnamurthy, N. Majumder, S. Poria, and E. Cambria, "A deep learning approach for multimodal deception detection," in *Proc. 19th Int. Conf. Comput. Linguistics Intell. Text Process. (CICLing)*, Hanoi, Vietnam, Mar. 2018, pp. 1–10.
- [84] L. Su and M. Levine, "Does 'lie to me' lie to you? An evaluation of facial clues to high-stakes deception," *Comput. Vis. Image Understanding*, vol. 147, pp. 52–68, Jun. 2016.
- [85] D. Avola, G. L. Foresti, L. Cinque, and D. Pannone, "Automatic deception detection in RGB videos using facial action units," in *Proc. 13th Int. Conf. Distrib. Smart Cameras (ICDSC)*, Trento, Italy, Sep. 2019, pp. 1–6.
- [86] S. V. Fernandes and M. S. Ullah, "Development of spectral speech features for deception detection using neural networks," in *Proc. IEEE 12th Annu. Inf. Technol., Electron. Mobile Commun. Conf. (IEMCON)*, Vancouver, BC, Canada, Oct. 2021, pp. 198–203.



SINEAD V. FERNANDES (Member, IEEE) received the B.S. degree in computer engineering and the M.S. degree in electrical engineering under the supervision of Dr. Muhammad Sana Ullah from Florida Polytechnic University (FPU), Lakeland, FL, USA, in 2016 and 2021, respectively. She is currently working as a Test Development Engineer at Jabil Healthcare, St. Petersburg, FL, USA. Her research interests include speech signal proceeding and machine learning, particularly deception detection. Besides, she was also one of the Founding Member of the IEEE HKN-Mu Omega chapter with the Florida Polytechnic University, which was established in 2020. She has been an IEEE Student Member at Florida Poly Chapter, since 2018.



MUHAMMAD SANA ULLAH (Senior Member, IEEE) received the B.S. degree in electrical and electronic engineering from the Chittagong University of Engineering and Technology, Chittagong, Bangladesh, in 2008, the M.S. degree in electrical and computer engineering from Purdue University Northwest, Hammond, IN, USA, in 2013, and the Ph.D. degree in electrical and computer engineering from the University of Missouri-Kansas City, Kansas City, MO, USA, under the supervision of Dr. M. Chowdhury.

He joined the College of Engineering, Florida Polytechnic University, Lakeland, FL, USA, as an Assistant Professor of computer engineering. His current research interests include a relatively new methodology and nanotechnology for the next generation of computing and other micro- and nano-electronic applications, modeling of RLC interconnects and RF interconnect in high-density integrated circuits (ICs), investigation of a tunneling device based on transition metal dichalcogenide material, graphene, carbon nanotube, and other emerging 2-D nanomaterials, very large scale integration signal processing, and higher-order statistics and spectra in signal processing.

• • •