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

SINEAD V. FERNANDES, (Graduate Student Member, IEEE), 
AND MUHAMMAD S. ULLAH, (Senior Member, IEEE) 
Department of Electrical and Computer Engineering, Florida Polytechnic University, Lakeland, FL 33805 USA 
Corresponding author: Muhammad S. 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 are 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 this 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, 4, 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. Therefore, it is important to address this 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, 4, 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
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], [73], Multi Layer Perceptron [1], Gradient Boosting Decision Tree [1], Levenberg-Marquardt [6], [45], BFGS Quasi-Newton [6], Long Short Term Memory [6], [40], [42], [59], Support Vector Machine [1], [13], [15], [40], [42], [46], [47] [59], [80], GMM [13], RVM [15], ensemble [40], Random Forest Decision Tree classifier [74], improved dense trajectories and OpenFace [80] and Ripper rule induction classifier [75]. In addition, non-verbal features like thermal facial analysis [7], [23], [33], [37], [39] eye movement analysis [8], [22], [26]-[28], visual cues [9]-[10], text analysis with BERT [16], brain activity analysis [21], [34], [36], multimodal features [24], [29], infrared imaging [31], and facial expressions [25], [27], [30], [32], [38] have been used for deception detection research and reported encouraging results. Similarly, verbal speech features such as acoustic [11], STE [1], [47], time-difference energy [6], delta energy [6], energy [40], [41], MFCC and energy [40], zero crossing rate [41], [46]-[47], entropy of energy [41], spectral centroid [41], spectral spread [41], spectral entropy [41], spectral flux [41], [76], spectral rolloff [41], chroma vector [41], chroma deviation [41], significant energy [43], [45], linear prediction cepstrum coefficients [44], bark energy [45], fundamental frequency [46]-[47], time-difference cepstrum [6], delta cepstrum [6], acoustic spectral based Mel cepstral features with energy [13], MFCC [40]-[41], [46]-[47], [59], log-energy of MFCC [59], thirteen cepstral coefficients [59] 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, 88]. 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 studied various visual cues of facial behavior to detect deception [9]. Tsechpenakis et al. studied HMM-based visual cue analysis to detect deception [10]. Prosodic and nonlinear linear dynamic features were studied by Zhou, Zhao, Pan, 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 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]. Then noon, Ali, and Hashim studied the facial expressions of 43 subjects to study how deception alters facial expressions [25]. George, Pai, Pai, and Praraj, studied eye blink count and eye blink duration using an intragender 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 [30], [32], [38]. Thermal imaging was studied by both Bedoya -Echeverry et al. and Dcesta et al. [37], [39]. [35] 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 [48]. 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 [48]. By modeling the non-linear perceptions of the human hearing sensation, the process of extracting spectral information can improve greatly [50]. 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 [48].

### 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 |
| [32] | Facial Micro-Expressions | - |
| [33] | Thermal Imaging | - |
| [34] | Brain Activities | 11 Male Subjects |
| [35] | Head Movement Analysis | 10 Subjects |
| [36] | Brain Activities | 5 Subjects |
| [37] | Thermal Imaging | - |
| [38] | Facial Expressions | - |
| [39] | Stress-Induced Facial Perspiration via Thermal Imaging | 40 Subjects |
| [77] | Keyboard Dynamics | 60 Participants |
| [78] | Unexpected Questions and Mouse Dynamics | 40 Participants |
| [79] | Keystroke Dynamics | 190 Subjects |
| [80] | Facial Expression | Videotaped Interviews |
| [81] | Facial Displays and Hand Gesture | 61 Trial Videos |
| [82] | Multimodal Feature (Lexical, Acoustic and Visual) | 121 Trial Videos |
| [83] | Improved Debe Trajectories and Microexpression | 104 Trial Videos |
| [84] | Involuntary Facial Expression | 12 People (344 Facial Images) |
| [85] | Audio, Visual, Textual (Static and Nonstatic) and Microexpression | 121 Trial Video |
| [86] | Macro and Micro Facial Expression | High Stakes YouTube Videos |
| [87] | Facial Action Units | 121 Trial Videos |
The absolute threshold of hearing can be described as the smallest level of a pure tone that can be detected by a listener in a noiseless environment [48]. 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) [48]. The absolute (quiet) threshold, $T_q(f)$ is approximated [48, 51] by the non-linear function that is shown in (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) [48]. Relative to a listener’s individual unmasked detection threshold for the stimulus, the intensity level difference is denoted by the SL [48], [52]. The SL is used because it quantifies listener-specific audibility instead of an absolute level [48].

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 [48], [53]. Auditory masking is when a louder sound becomes undetectable to the human ear due to a weaker but audible sound [53] - [55]. Masking is a process where one sound is virtually inaudible due to the presence of another sound [48]. 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 [48]. 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 [48].

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 [53], [56]. Pre-masking is one type of temporal masking that immediately precedes the presence of a masker [53]. Post-masking is another type of temporal masking that immediately follows the masker and obscures sound [53]. 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 [48].

Using the absolute threshold of hearing, a modified version known as the detection threshold is used for spectrally complex quantization noise [48]. 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 [48]. Based on how the human ear naturally performs spectral analysis, the threshold estimation is calculated [48].

First, along the basilar membrane in the cochlea or inner ear, a frequency-to-place transformation occurs [48], [57]. 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 [48]. 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 [48]. Along the length of the basilar membrane, the cochlear structure produces travel waves that were previously excited by the mechanical vibrations [48]. 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 [48]. 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 [48]. After the traveling wave slows down and the magnitude increases to a peak, the traveling wave then decays rapidly beyond the peak [48]. 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” [48]. 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 [48]. The magnitude responses are nonlinear and asymmetric [48]. The bandwidth of the cochlear filter passbands is nonuniform and increases along with the increasing frequency [48]. Additionally, the cochlear filter passbands are quantified by the “critical bandwidth” that is a function of frequency [48]. Critical bandwidth is when the perceived intensity of sound remains constant at a constant SPL for a narrow-band noise source [48]. 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 [48]. 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 [48].

To enumerate the cochlear filter passbands, the critical bandwidth is calculated as a function of frequency [48]. The critical bandwidth, $BW_c(f)$ is calculated approximately using the non-linear function shown in (2) where frequency, $f$ is in Hz.

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

$$BW_c(f)$$
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” [48].

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

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 [45, 49]. 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 [48]. 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 [48].

Fan et al. studied detecting deception using speech signals by extracting the short-time energy (STE) feature among
I. INTRODUCTION

Detecting deception using cepstral features has limited success when compared to other methods. The effectiveness of cepstral analysis is due to the unique properties of speech signals. Cepstral analysis is based on the cepstrum, which is a representation of the signal in the frequency domain. The cepstrum is obtained by taking the logarithm of the magnitude spectrum of the signal and then applying a Fourier transform. This process allows for the extraction of information about the signal's structure.

Previous studies have shown that the cepstrum contains information about the signal's pitch and amplitude. These features are useful in distinguishing between honest and deceptive speech. For example, a study by Ullah et al. [42] found that deceptive speech has a lower fundamental frequency than honest speech. Additionally, the cepstrum can be used to detect changes in speech characteristics that are indicative of deception.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3157821, IEEE Access

VOLUME XX, 2017

II. RELATED WORK

A. Speech Features

Speech features are used to describe the characteristics of speech signals. They can be classified into two main categories: static and dynamic features. Static features, such as fundamental frequency, zero-crossing rate, and energy, are extracted from the entire speech signal. Dynamic features, such as the delta cepstrum, are calculated from the changes in the signal over time.

B. Deception Detection

Deception detection is a subfield of speech recognition that focuses on identifying deception in speech signals. This can be achieved through the use of cepstral features, which have been shown to be effective in distinguishing between truthful and deceptive speech.

C. Cepstral Analysis

Cepstral analysis is a technique used to analyze speech signals. It involves taking the logarithm of the magnitude spectrum of the signal and then applying a Fourier transform. This process allows for the extraction of information about the signal's structure.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/
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 [61]. 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 [63]. 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, and increased packing number as compared to the MGA method [64]. 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 [65]. 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 [66] – [67], speech enhancement [68] – [69], and speech recognition [70, 71, 72], deception detection [4, 45], 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.

| Ref | Type of Study | Recognition Rate Before PCA | Recognition Rate After PCA |
|-----|---------------|-----------------------------|----------------------------|
| [6] | Deception Detection – Time-Difference Energy Speech Feature | 79.07% | 84.01% |
| [6] | Deception Detection – Delta Energy Speech Feature | 70.83% | 84.01% |
| [6] | Deception Detection – Time-Difference Cepstrum Speech Feature | 83.33% | 84.01% |
| [6] | Deception Detection – Delta Cepstrum Speech Feature | 75% | 84.01% |
| [61] | Diabetes Disease Prediction | 75% | 91.66% |
| [63] | Underwater Image Recognition PCA | 63.37% | 91.66% |
| [64] | Irregular Packing - Experiment 2 Example 1 | 74.00% | 87.50% |
| | Irregular Packing - Experiment 2 Example 2 | 75.27% | 87.50% |
| [65] | Genetic Study - Heart Data | 84.01% | 87.50% |
| | Genetic Study Adult Data | 79.07% | 91.66% |

VOLUME XX, 2017
| 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 Expressions | ~ | ~ | 76.92% |
| [31] | Infrared Imaging using Time Domain Analysis | ~ | ~ | 69% |
|     | Infrared Imaging using Frequency Domain Analysis | ~ | ~ | 84% |
TABLE 5. (Continued.) Feature and Classification Method Recognition Rate Comparison

| Ref | Feature                              | Classification Method | Computational Time | Recognition Rate |
|-----|--------------------------------------|-----------------------|--------------------|------------------|
| [32] | Facial Micro-Expressions              | ~                     | ~                  | 85%              |
| [33] | Thermal Imaging                      | ~                     | ~                  | 64%              |
| [34] | Brain Activities                      | ~                     | ~                  | 70.83%           |
| [35] | Head Movement Analysis                | ~                     | ~                  | 58.25%           |
| [36] | Brain Activities                      | ~                     | ~                  | -                |
| [37] | Thermal Imaging                      | ~                     | ~                  | 79.2%            |
| [38] | Facial Expressions                    | ~                     | ~                  | 70.26%           |
| [39] | Stress-Induced Facial Perspiration via Thermal Imaging | ~                     | ~                  | 80%              |
| [40] | 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%            |
| [41] | LexRNN: Lexical features              | LSTM                  | ~                  | 81.03%           |
|      | HybridRNN: Speaker-dependent and lexical features | LSTM                  | ~                  | 84.11%           |
|      | AudioRNN: Speaker-dependent acoustic features | LSTM                  | ~                  | 62.59%           |
| [42] | MFCC                                  | SVM                   | ~                  | 51.8%            |
|      | Energy                                | SVM                   | ~                  | 50.2%            |
|      | MFCC                                  | LSTM                  | ~                  | 54.6%            |
|      | Energy                                | LSTM                  | ~                  | 47.0%            |
| [45] | Bark Energy                           | Levenberg-Marquardt   | 3-20 epochs        | 83.33%           |
|      | Significant Energy                    | Levenberg-Marquardt   | 3-20 epochs        | 66.67%           |
| [46] | 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%           |
| [47] | Fundamental Frequency, Short-Term Energy, Zero-Crossing Rate, and MFCC | SVM                     | ~                  | 82.47%           |
| [59] | 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%              |
| [73] | Critical “hot spot” segments identified to tell if a speaker is truthful or lying | Bagging, AdaBoost, and J48 combination | ~                  | 68.6%            |
| [74] | Lexical and Speech                    | RFDT Classifier       | ~                  | 50.45%           |
| [75] | Acoustic/Prosodic, Lexical, and Speaker-Dependent Combination | Ripper Rule Induction Classifier | ~                  | 66.4%            |
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 nonlinear 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].

Using SVM, they achieved the 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 [59]. Ullah and Gopalan proposed extracting the Bark energy and significant energy features to detect deception [45]. Using the significant energy feature, they achieved a 66.67% recognition rate with the Levenberg-Marquardt classification method [45]. Using the Bark energy feature, they achieved a higher recognition rate at 83.33% with the Levenberg-Marquardt classification method [45]. 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 [41]. 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 [42]. When using the energy

### TABLE 5. (Continued.) Feature and Classification Method Recognition Rate Comparison

| Ref. | Feature | Classification Method | Computational Time | Recognition Rate |
|------|---------|-----------------------|--------------------|-----------------|
| [77] | Keyboard Dynamic | Logistic | ~ | 100% |
|      | Keyboard Dynamic | SVM | ~ | 90% |
|      | Keyboard Dynamic | LMT | ~ | 90% |
| [78] | Unexpected Questions and Mouse Dynamics | Random Forest | ~ | 70.24% |
|      | Unexpected Questions and Mouse Dynamics | Logistics | ~ | 70.15% |
|      | Unexpected Questions and Mouse Dynamics | SVM | ~ | 90% |
|      | Unexpected Questions and Mouse Dynamics | LMT | ~ | 70.15% |
| [80] | Facial Micro Expressions | SVM & Deep Neural Networks | ~ | 57% |
| [81] | 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% |
| [82] | 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% |
| [83] | Improved Deese Trajectories (IDT) | Linear SVM | ~ | 77.31% |
|      | Microexpression | Random Forest | ~ | 80.64% |
|      | Transcripts | Naive Bayes | ~ | 66.25% |
|      | MFCC | Kernel SVM | ~ | 81.71% |
|      | IDT+Microexpression+Transcript+MFCC | Linear SVM | ~ | 87.73% |
| [84] | Facial Expression (Angry) | LIBSVM | ~ | 86.02% |
|      | Facial Expression (Enjoyment) | LIBSVM | ~ | 73.16% |
|      | Facial Expression (Fear) | LIBSVM | ~ | 80.46% |
|      | Facial Expression (Sadness) | LIBSVM | ~ | 90.15% |
| [85] | Audio | MLP | ~ | 52.38% |
|      | Visual | MLP | ~ | 93.08% |
|      | Textual (Static) | MLP | ~ | 80.16% |
|      | Textual (Non Static) | MLP | ~ | 90.24% |
|      | Microexpression | MLP | ~ | 76.19% |
| [86] | Macro Facial Expression | Random Forest | ~ | 73.85% |
|      | Micro Facial Expression | Random Forest | ~ | 56.92% |
|      | Macro & Micro Facial Expression | Random Forest | ~ | 76.92% |
| [87] | Facial Action Units | RBF & SVM | ~ | 76.84% |

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/
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 [42]. 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 [73]. 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 [74]. 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 [47]. 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 [46]. They were also able to achieve a recognition rate of 93.33% using an artificial neural network (ANN) for the same set of features [46]. Xue et al. proposed using MFCC, pitch, and energy features to detect deception using various classification methods [40]. 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 [40]. 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 [40]. 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 [75]. Table 5 presents the previous 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 us 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 6th International Conference on Intelligent Control and Information Processing (ICICIP), Wuhan, China, Nov. 2015, pp. 268-273.

[2] B. Rajoub and R. Zwiggelaar, “Thermal Facial Analysis for Deception Detection,” IEEE Transactions on Information Forensics and Security, vol. 9, no. 6, pp. 1015-1023, Apr. 2014.

[3] M. Sanaullah and K. Gopalan, “‘Distinguishing deceptive speech from truthful speech using MFCC,’ in Proceedings of the 7th International Conference on Circuits, Systems and 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 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (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 changes of diameter pupil and the eye movement use method Gabor wavelet transform and decision tree,” in IEEE International Conference on Internet of Things and Intelligence System (IOTIAS), 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, May. 2021.

[7] I. Pavlidis and J. Levine, “‘Thermal facial screening for deception detection,’ in Proceedings of the Second Joint 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society, 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 International Conference on Control, Electronics, Renewable Energy and Communication (ICCERC), Yogyakarta, Indonesia, Sep. 2017, pp. 40-45.

[9] M. H. Yap, B. Rajoub, H. Uegai, and R. Zwiggelaar, “‘Visual cues of facial behaviour in deception detection,’ in IEEE International Conference on Computer Applications and Industrial Electronics (ICCAIE), Penang, Malaysia, Dec. 2011, pp. 294–299.

[10] G. Tsechpenakis, D. Metaxas, M. Adkins, J. Kruse, J. K. Burgoon, M. Jensen, J. Meservy, D. P. Twitchell, A. Deokar, and J. F. Nunemaker, “‘HMM-based deception recognition from visual cues,’ in IEEE International Conference on Image Processing (ICIP), Cambridge, MA, USA, Sep. 2002, pp. 401–404.

[11] X. Xie, R. Liang, H. Tao, Y. Zhu, and L. Zhao, “‘Convolutional bidirectional long short-term memory for deception detection with acoustic features,’ in IEEE Access, vol. 6, pp. 76527–76534, Nov. 2018.

[12] K. Gopalan and S. Wenndt, “‘Speech analysis using modulation-based features for detecting deception,’ in 15th International Conference on Digital Signal Processing, 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 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings, 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 IEEE 12th International New Circuits and Systems Conference (NEWCAS), Trois-Rivières, 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. Nefci, “Building a better lie detector with BERT: The difference between truth and lies,” in International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 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 and Criminological Psychology, vol. 17, 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 2nd International Conference on Advances in Computational Tools for Engineering Applications (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 Transactions on Biomedical Engineering, vol. 60, no. 5, pp. 1280–1289, May. 2013.

[20] S. Feng, R. Banerjee and Y. Choi, “Syntactic stylometry for deception detection” in Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL), pp. 171–175, Jan. 2012.

[21] S. Amir, N. Ahmed, and B. S. Chowdhry, “Lie detection in interrogations using digital signal processing of brain waves,” in 3rd International Conference on Instrumentation, Communications, Information Technology and Biomedical Engineering (ICICI-BME), Bandung, Indonesia, Nov. 2013, pp. 209–214.

[22] B. Singh, P. Rajiv, and M. Chandra, “Lie detection using image processing,” in International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, Jan. 2015.

[23] U. Jain, B. Tan, and Q. Li, “Concealed knowledge identification using facial thermal imaging,” in IEEE International Conference on Acoustics, Speech and Signal Processing (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 Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC’14), 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 3rd Scientific Conference of Electrical Engineering (SCEE), Baghdad, Iraq, Dec. 2018, pp. 51–56.

[26] S. George, M. M. M. Päi, R. M. Päi, and S. K. Praharaj, “Eye blink count and eye blink duration analysis for deception detection,” in International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udum, 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,” MSc thesis, BRAC University, 2018.

[28] P. K. Lakkapragada, “Using eye-tracking to understand user behavior in deception detection system interaction,” MSc Thesis, Missouri University of Science and Technology, USA, 2016.

[29] Z. Wu, B. Singh, L. S. Davis, and V. S. Subrahmanian, “Deception detection in videos,” in 32nd AAAI Conference on Artificial Intelligence (AAAI-18), New Orleans, Louisiana, USA, Dec. 2017, pp. 1697–1702.

[30] L. Su and M. D. Levine, “High-stakes deception detection based on facial micro-expressions,” in 22nd International Conference on Pattern Recognition, 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 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Delhi, India, Sep. 2014, pp. 1021–1026.

[32] 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 2nd International Conference on Advances in Computational Tools for Engineering Applications (ACTEA), Beirut, Lebanon, Dec. 2012, pp. 33–38.

[33] L. Warmelink, A. Vrij, S. Mann, S. Leaf, 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.

[34] A. I. Simbolon, A. Turnip, J. Hutahean, Y. Siagian, and N. Irawati, “An experiment of lie detection based EEG-F300 classified by SVM algorithm,” in International Conference on Automation, Cognitive Science, Optics, Micro Electro-Mechanical System, and Information Technology (ICACOMIT), Bandung, Indonesia, Oct. 2015, pp. 68–71.

[35] D. I. Noje and R. Malutan, “Head movement analysis in lie detection,” in Conference Grid, Cloud & High-Performance Computing in Science (ROLCG), Cluj-Napoca, Romania, Oct. 2015.

[36] M. A. Rahman and M. Ahmad, “Lie detection from FNIR signal and NeuroImage,” in International Conference on Medical Engineering, Health Informatics and Technology (MedTec), Dhaka, Bangladesh, Dec. 2016, pp. 1-6.

[37] S. Bedoya-Echeverry, H. Belalczar-Ramirez, H. Loaiza-Coara, S. E. Nope-Rodriguez, C. R. Pinedo-Jaramillo, and A. D. Restrepo-Giron, “Detection of lies by facial thermal imagery analysis,” in Revista Facultad De Ingeniería, vol. 26, no. 44, pp. 47–59, 2017.

[38] J. Bailey, S. Demyanov, K. Ramamohan Rao, and C. Leckie, “Detection of deception in the Mafia party game,” in Proceedings of the 2015 ACM International Conference on Multimodal Interaction (ICMI), Seattle, Washington, USA, Nov. 2015, pp. 335–342.

[39] M. Dcosta, D. Shastri, R. Vilalta, J. K. Burgon and I. Pavlidis, “Perinatal indicators of deceptive behavior,” in 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Ljubljana, Slovenia, May. 2015, pp. 1-8.

[40] A. Xue, H. Rohde, and P. A. Finkelstein, “An Acoustic Automated Lie Detector,” Princeton University, 2019.

[41] S. Desai, M. Siegelman, and Z. Maurer, “Neural lie Detection with the CSC Deceptive Speech Dataset,” Stanford University, 2017.

[42] W. Wang, X. Yu, Y. H. Wang, and R. Swaminathan, “Audio fingerprint based on spectral flux for audio retrieval,” in International Conference on Audio, Language and Image Processing, Shanghai, China, Jul. 2012, pp. 1104–1107.

[43] R. C. Cosseti and J. M. D. B. Lopez, “Voice stress detection: A method for stress analysis detecting fluctuations on Lippold microtremor spectrum using FFT,” in 21st International Conference on Electrical Communications and Computers (CONIELE-COMP), San Andres Cholula, Mexico, Feb. 2011, pp. 184–189.

[44] E. Wong and S. Sridharan, “Comparison of linear prediction cepstrum coefficients and mel-frequency cepstrum coefficients for language identification,” in Proceedings of International Symposium on Intelligent Multimedia, Video and Speech Processing (ISMIP), Hong Kong, China, Apr. 2003, pp. 95–98.

[45] M. Sanaullah and K. Gopalan, “Deception detection in speech using bark band and perceptually significant energy features,” in IEEE 56th International Midwest Symposium on Circuits Systems (MWSCAS), Columbus, OH, USA, Aug. 2013, pp. 1212–1215.

[46] N. Srivastava and S. Dubey, “Deception detection using artificial neural network and support vector machine,” in 2nd International Conference on Electronic, Communication and Aerospace Technology (ICECA), Coimbatore, India, Mar. 2018, pp. 1205-1208.

[47] H. Tao, P. Lei, M. Wang, J. Wang, and H. Fu, “Speech Deception Detection Algorithm Based on SVM and Acoustic Features,” in IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT), Dalian, China, Oct. 2019, pp. 31–33.

[48] T. Painter and A. Spanias, “Perceptual coding of digital audio,” in Proceedings of the IEEE, vol. 88, no. 4, pp. 451–513, Apr. 2000.

[49] J. S. Park, J. S. Yoon, Y. H. Seo, and G. J. Jang, “Spectral energy based voice activity detection for real-time voice interface,” in Journal of Theoretical and Applied Information Technology, vol. 95, no. 17, pp. 4305–4312, Sep. 2017.

[50] B. Winduratna, “FM Analysis/Synthesis-Based Audio Coding,” 70th Audio Engineering Society Convention (AES), vol. 104, May. 1998.

[51] E. Terhardt, “Calculating virtual pitch,” Hearing Research, vol. 1, no. 2, pp. 155–182, Mar. 1978.

[52] B. C. J. Moore, Introduction to the Psychology of Hearing, Academic, New York, USA, 1977.

[53] H. M. S. Naing, R. Hidayat, B. Winduratna, and Y. Miyangya, “Psychoacoustical Masking Effect-Based Feature Extraction For Robust-Speech Recognition,” International Journal of Innovative
SINEAD V. FERNANDEZ (Graduate Student Member, IEEE) received the B.S. degree in computer engineering and the M.S. in Electrical Engineering under the supervision of Dr. Muhammad Sana Ullah from Florida Polytechnic University (FPU), Lakeland, FL, USA in 2016 and 2021, respectively. Her research interests include speech signal processing and machine learning, particularly deception detection. She was IEEE student member at Florida Poly chapter since 2018. Besides, she was also one of the Founding Member of the IEEE HKN-Mu Omega chapter with the Florida Polytechnic University which establish in 2020. Currently, she is working as test development engineer at Jabil Healthcare, St. Petersberg, Florida.

MUHAMMAD S. 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 2D nanomaterials; very large scale integration signal processing; and higher-order statistics and spectra in signal processing.