J. Electrical Systems 20-4s (2024): 619-633

[1] Fahad Abdulridha

[2] Baraa M. Albaker

A Review on Non-Invasive Multimodal Approaches to Detect Deception Based on Machine Learning Techniques

Abstract: - Detecting deception has been investigated by the scientific community for over a century due to its importance in the justice system and homeland security. Attempts to come up with an approach, a system or a framework that serves the purpose of discerning lies from truths has therefore been a major field. This has led researchers to automate the detection process and reduce its invasiveness as much as possible. In addition, machine learning techniques are used with multiple channels of information, known as modals, to increase accuracy in what is known as a multimodal approach. As a result, several research and datasets are currently available, and it could be challenging to identify successful patterns, gaps, and future directions. In this paper, over fifty state-of-the-art publications in the field of deception detection using non-invasive approaches based on machine learning techniques are analyzed after reviewing more than one thousand publications from Scopus, IEEE Xplore, Web of Science, ScienceDirect, and Google Scholar. The work presents the classification techniques and datasets used with their detection performance and finally analyzing the data to draw conclusions. The reported detection accuracy ranges from about 50% to 95% for monomodal approaches based on facial expression, body movement, audio, or thermal imaging. In conclusion, the multimodal approach shows promising results as it reaches a detection accuracy approaching 100%. It outperforms any alternative non-invasive approach, especially when dealing with small datasets, which seems to be the biggest challenge in this field. Future research directions should focus on experimenting with multimodal systems by developing larger datasets as well as implementing classification algorithms that can work with multiple modals effectively.

Keywords: Deception Detection, Lying Detection, Machine Learning, Multimodal, Non-Invasive, Facial Expressions, Audio Features, Thermal Imaging.

I. INTRODUCTION

The concept of Detecting Deception (DD) in humans has been a great area of interest in human history, dating back to ancient Greeks and Indians more than 2000 years ago. The earliest attempts at detecting deception date back to around 900 – 600 B.C.E. in the works of the Hindu Dharmasastra of Gautama [1]. Which indicates that the importance of DD was understood for a very long time and many attempts were made to detect deceit in humans, especially in high stake situations such as court rooms and investigations. The first real implementation of DD was the Polygraph [2], a physiological sensor that detects various biological signals such as heart rate, body temperature and breathing patterns, which are analyzed by an expert during an interview with the subject to determine deception in each of the questions answered. Many other attempts have been made in order to improve upon the Polygraph due to its lack of accuracy and validity [3]. Earlier works such as using facial expressions in order to detect deception by identification of Action Units (AU) in the face (macro or micro movements in the face) that correlate most with lying [4]. Ekman attempted to use voice and body movement [5]. All these attempts and more such as using Electroencephalogram (EEG) [6] share the same shortcomings of the Polygraph, namely, the involvement of an interviewer in direct contact with the subject and a judge or an expert who analyzes the results and makes the decision based on the collected readings and information whether the subject is being deceptive or not [7]. This shifted the research of DD in the direction of Automated Deception Detection (ADD) and avatar mediated interviewing as well as non-invasive modalities for gathering and analyzing data to address these issues which will be explored further in later sections.

Another challenge and perhaps the most important in the field of DD is increasing the accuracy of deceit detection and reduction of false positives as much as possible. One key approach in recent years has been mixing multiple sources of information such as facial expressions, Brain fingerprint, thermal imaging, body movement, voice, etc. and analyzing them together. This is to increase the likelihood of correctly identifying deception with increased confidence in the results. In this paper, techniques and methods used in the field of DD are reviewed and their effectiveness are reported in a multimodal approach as well as identifying the possible future research in this regard.

[1, 2] College of Engineering, Al-Iraqia University, Baghdad, Iraq

Email: [1] fahad.m.abdulridha@aliraqia.edu.iq, [2] baraamalbaker@ymail.com

II. RELATED WORK

Although the field of deception detection is well covered from all over the world, and as briefly discussed earlier and will expand on later, in last five years the focus has shifted greatly towards DD with innovative and more practical approaches especially in the non-invasive domain. Furthermore, experimenting with combining multiple modals has also been rising in popularity. In fact, a simple keyword search in the Scopus database in published articles relating to deception detection from the last 13 years reveals there has been as much published papers with the keyword "multimodal" in the last 3 years as there was in the 10 years prior.

Despite these trends, there hasn't been an equal effort in the review front, and more importantly, it hasn't been reflective of the current landscape and great shift the field of ADD is going through, focusing on a single general modal at best. Table 1 presents a list of review articles and their scope of work. These review articles do indeed offer great value in their comprehensiveness due to their relatively narrow scope in an otherwise wide-ranging field, this can be valuable for researchers with a specific modal in mind, it is less so when it comes to the multimodal oriented future research.

Table 1: Review articles and their modal scope in the last five years

Paper	Text	Facial Expressions	Body Movement	Audio	Thermal Imaging	EEG	Multimodal
[8]				✓			
[9]	✓						
[10]	✓						
[11]						✓	
[12]	✓						
[13]							
[14]	✓						
[15]	✓			✓			
Scope of this work		✓	✓	✓	✓		✓

III. DATA COLLECTION

For the data collection process, two routes have been taken, as shown in Figure 1. First, a list of publications was exported from the Scopus, IEEE Xplore, Web of Science, and ScienceDirect databases using the search term "Deception Detection" which yielded 2001 publication. Any publication before 2018 was filtered out leaving 841 publications, then it was further narrowed down by removing irrelevant journals and papers as well as only keeping journal articles and some conference papers, leaving 157 publications in the list. The remaining publications were manually examined to filter out irrelevant or off topic articles or sources, leaving us with a total of 40 papers. The resulting list was then categorized based on its employed DD modal with an additional category for review papers and multimodal papers, many papers got listed under multiple categories. The second route was through manual search in Google Scholar for relevant papers as well as for review papers that might have been missed in the above databases list, these papers added to the final categorized list yielding a total of 50 papers.

IV. INVASIVENESS IN DECEPTION DETECTION SYSTEMS

Invasiveness of a given deception detection system being employed is a measure of the degree of contact of the various parts of a system that are aimed at gathering, identifying, analyzing and finally classifying the data from the subject as either truthful of deceitful. It measures how much does the system need to invade the space of the subject throughout the process of DD from preparing, interviewing and all the way to the final classification for it to function as intended. This includes how the various sensors being used in the system are handled, placed, operated, their number, their degree, and interaction with the subject. It can also include the design of the interview, how much of the DD system and its inner workings they can see. For example, the original polygraph can be considered a highly invasive system, since it relies on attaching various sensors directly on the subject's body to read their physiological changes. It might have also shown the gathered data to the subject and the expert trying to analyze them all during the interview. On the other hand, a DD system with only a teleprompter displaying written interview questions and a simple camera placed far away from the subject or even concealed from them during the interview, resulting in no human involvement, can be considered highly non-invasive.

While the property of invasiveness of a given system may not negatively impact the functionality of a said system. For instance, an implemented algorithm for classification does not care about how the data is obtained or the experience of the subject under investigation. So long as it is provided with the expected data in the expected format, it will perform as intended and produce results as designed. However, for a DD system that is designed for real life applications in mind, it is imperative to take into consideration how the system is affecting the experience of the subject. A system that incentivizes the subject to try harder to conceal the truth or puts the subject in a state of unrest and stress will always be suboptimal. This has been shown by Buller and Burgoon [17] in their "Interpersonal Deception Theory" where the interview context, the environment, the incentives, the relationship context and even the initial expectation of dishonesty can influence the behavior, strategy, restraint between the sender (the subject under investigation) and receiver (the judge or the interviewer) and can even have an effect on the "truth leakage" from the deceiver.

V. DECEPTION DETECTION APPROACHES

Three main categories of DD modals have emerged in recent literature, those that are based on the 'leakage hypotheses', 'reality monitoring' and 'truth default'[16]. In this paper, we focused on the modals that are based on the 'leakage hypotheses', which is the idea that lying causes involuntary physiological reactions in the body of the subject such as eye movement, changes in heart rate, facial expressions, twitching or movement in arms or legs, body shifting, increase in body and face temperature and changes in voice pitch to name a few. These are the most relevant sources of data for ADD, specifically with non-invasive approaches such as prerecorded video and audio. Figure 2 depicts the categories of DD approaches, modals, and sub-modal.

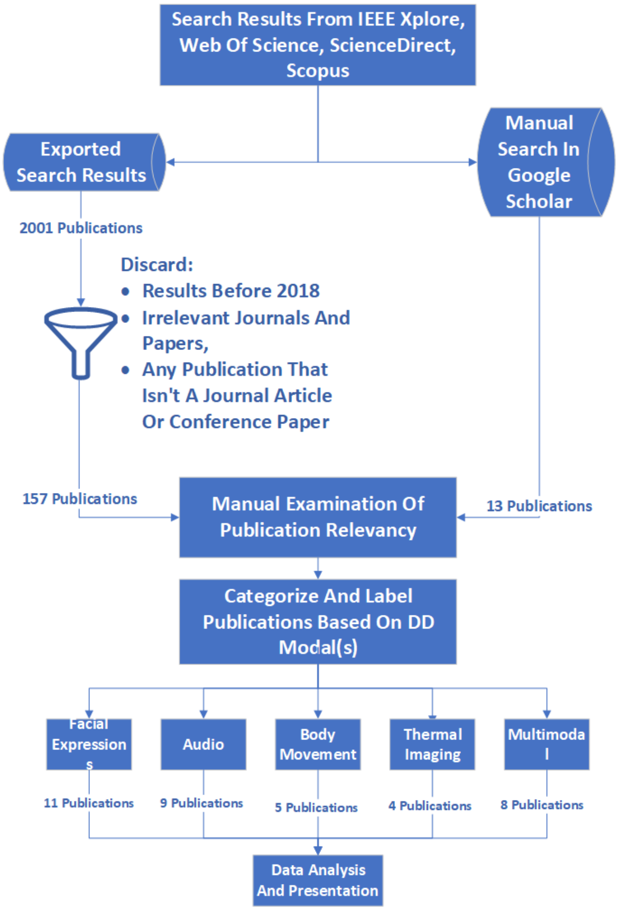

Figure 1. Articles collection and processing steps

VI. FACIAL EXPRESSIONS BASED DECEPTION DETECTION

There are two main approaches to record the facial expressions that are being targeted as a meaningful cue which lying can 'leak' through. The first is an invasive approach using fEMG (Facial Electromyography) sensors attached to the targeted muscles that are most correlated with deception, namely the frontalis, corrugator supercilii, orbicularis oculi, levator labii superioris alaeque nasi, zygomaticus, depressor anguli oris, and orbicularis oris, such as in the work presented by Dong et al. [18]. The second approach is a non-invasive video recording of the face of the subject while being interviewed, the video is later analyzed by a computer system for pattern recognition using computer vision to register all the AUs that are relevant for DD such as the work carried by Khan et al. [19].

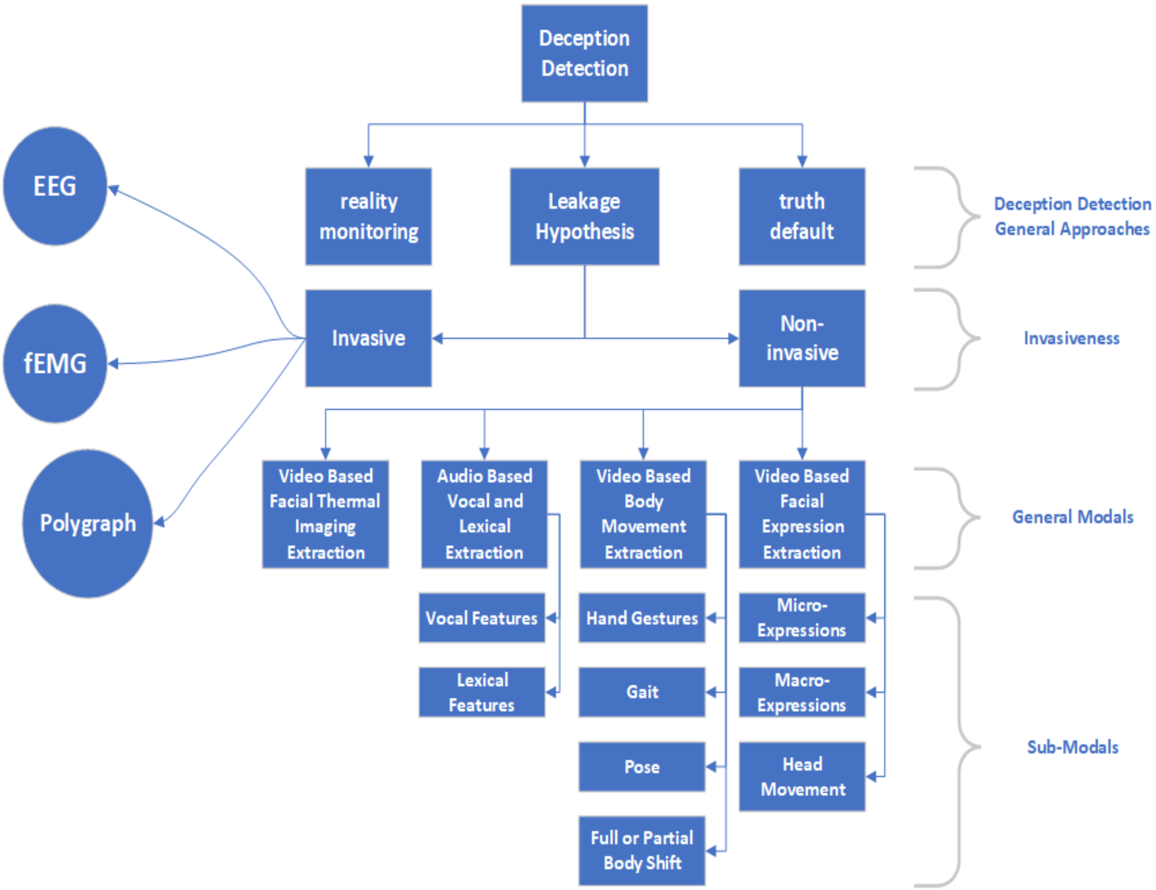

Figure 2. Deception Detection Approaches and Modals

In both cases, the patterns gathered are used as inputs in a machine learning algorithm which classifies the signal as deceitful or truthful. There are two main feature sets that can be extracted for DD, Macro and Micro expressions. The first feature set includes facial features that are more than 0.5 seconds in duration, while the second feature set tends to be shorter than 0.5 seconds [20]. They are far less voluntary than macro expressions [21] [22] which makes them very valuable for DD purposes.

Table 2 illustrates the used datasets, features and classifiers for DD based on facial expression features. Figure 3 presents a comparison among used classifiers. The most common algorithms used in recent literature are Support Vector Machine (SVM) and Random Forest (RF) with the first achieving a result of up to 77% - 83% [19] [23] accuracy in DD and the latter having similar accuracy of 78% [19]. However, when combining Facial expressions with other modals such as physiological reactions (arousal, temperature, etc.) and voice, accuracies of up to 90% can be achieved [23].

Table 2: Facial Expression features used to detect deception

Paper	Year	Dataset	Feature	Classifier	Classification Performance
[24]	2016	High-Stakes Deception Videos Collected from the Internet	Both Macro and Micro-Expressions Are Included	Random Forest	ACC=0.7692;
			Only Macro-Expressions Are Included	Random Forest	ACC=0.7385;
			Only Micro-Expressions Are Included	Random Forest	ACC=0.5692;
[25]	2019	[26]	Face	fake feature vectors and attack the classifier	ACC=0.8433; AUC=0.8411;
			Motion	fake feature vectors and attack the classifier	ACC=0.86; AUC=0.8863;
			Face, Motion	fake feature vectors and attack the classifier	ACC=0.8821; AUC=0.9057;

Chart Area

Paper	Year	Dataset	Feature	Classifier	Classification Performance
			Face, Motion, CL	fake feature vectors and attack the classifier	ACC=0.8916; AUC=0.9189;
			Face, Motion, CL, ML	fake feature vectors and attack the classifier	ACC=0.9233; AUC=0.9583;
			Face, Motion, CL, ML, AL	fake feature vectors and attack the classifier	ACC=0.9316; AUC=0.9671;
[23]	2020	[26]	Facial Affect	SVM	ACC=0.72; AUC=0.8; F1=0.67
			Facial Affect, Visual	SVM	ACC=0.74; AUC=0.86; F1=0.71
			Facial Affect, Visual (Soft Hybrid Fusion)	SVM	ACC=0.74; AUC=0.83; F1=0.72
			Facial Affect, Visual (Adaboost)	SVM	ACC=0.74; AUC=0.87; F1=0.73
[18]	2022	Experimental	Micro-Expressions from Facial Muscles	SVM	
[19]	2021	Experimental	Full Facial Features	SVM	ACC=0.77; F1=0.78
			Full Facial Features	Random Forest	ACC=0.77; F1=0.79
			Full Facial Features	ANN	ACC=0.72; F1=0.75
			Important Features (Top 24)	SVM	ACC=0.77; F1=0.78
			Important Features (Top 24)	Random Forest	ACC=0.78; F1=0.8
			Important Features (Top 24)	ANN	ACC=0.72; F1=0.74
[27]	2021	Political Videos Collected from The Internet	AU	Decision Tree	ACC=0.66;
			Emotion	Decision Tree	ACC=0.56;
			Gaze	Decision Tree	ACC=0.61;
			Pose	Decision Tree	ACC=0.53;
			AU, Gaze	Decision Tree	ACC=0.61;
			AU, Gaze, Pose Comb	Decision Tree	ACC=0.68;
			Gaze, Emotion	Decision Tree	ACC=0.65;
[28]	2018	[26]	Visual - AU	DEV Framework	ACC=0.685;
			Visual - DEV	DEV Framework	ACC=0.75;
[29]	2016	[26]	Visual (Eyebrows, Eyes, Mouth)	SVM	ACC=0.67;
[30]	2018	[26]	Visual	Multi-Layer perceptron	ACC=0.9308; AUC=0.9596;
			Micro-Expression	Multi-Layer perceptron	ACC=0.7619; AUC=0.7512;
[31]	2020	[26]	Facial Displays	SVM	ACC=0.7627; AUC=0.8581;
			Facial Displays	Random Forest	ACC=0.7627; AUC=0.927;
			Facial Displays	NN	ACC=0.8079; AUC=0.9416;
[32]	2021	Bol, RL Trail, And MU3D		Softmax classifier	

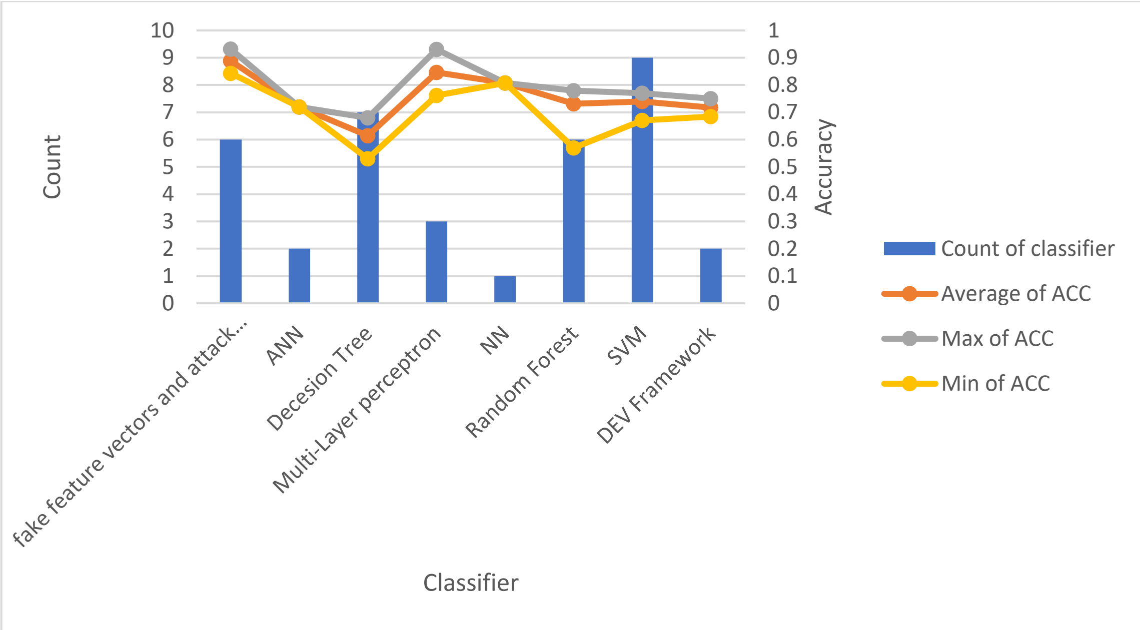

Figure 3. Classifiers used and average accuracy for each in facial expressions modal

VII. Audio Based Deception Detection

There are two main sets of features that can be extracted for DD analysis and classification, vocal and lexical features. Vocal features describe all the changes in the subject's voice such as pitch, intensity, Spectral, Cepstral (MFCC), duration, spectral harmonicity, psychoacoustic spectral and sharpness to name a few (see the InterSpeech 2013 feature set). These features can be extracted from speech and analyzed to determine their correlation with deceptive speech among other things. In addition, it is also possible to extract emotions from these features which has been shown to be a good predictor of deception [33].

Lexical features, on the other hand, includes analyzing the speech patterns in words including pauses and other verbal cues. The words can be represented in a model such as Bag-of-words for pattern and relationship analysis for the classifier algorithm. This approach has been used in two main broad fields of ADD, the first is the lexical analysis of transcribed speech from recorded audio that is the concept being investigated and discussed so far. The second is Text Based Deception Detection. This field of research also depends on lexical analysis in a way that is very similar to transcribed audio except for the original source being written instead of spoken. This introduces some differences in feature analysis such as the lack of pauses and filler words, but perhaps the most relevant and significant difference is the application area where audio based lexical analysis is mainly used in recorded interviews, trials and investigations, while text based has been utilized in detecting deception in social media, news articles, instant messaging services, emails and email spam detection to name a few [12]. Since these applications do not fall under the leakage hypotheses theory, Text Based Deception Detection is out of scope for this review.

Classification algorithms can be used to determine deception in the subject's speech based on the extracted features, SVM, Regression Vector Machine (RVM) and RF are among the top algorithms with accuracies of up to 68%, 70% and 76% respectively [34] [35] [36]. In a wholistic research approach, the audio modal can be technically categorized as a multimodal one since it's the combination of acoustic/vocal features as well as the lexical features, and recent literature rarely investigates one without the other. However, when paired for example with facial expressions such as in the works of Şen et al. [31], accuracies of up to 83% were achieved using NN classifier and 78% using RF (See Table 3 and Figure 4).

Table 3: Audio features used to detect deception

Paper	Year	Dataset	Feature	Classifier	Classification Performance
[35][1]	2021	[37], [38]	Time Difference Energy	Levenberg-Marquardt	ACC=1
			Time Difference Energy	LSTM	ACC=1
			Delta Energy	Levenberg-Marquardt	ACC=0.875
			Delta Energy	LSTM	ACC=0.8333
			Time Difference Cepstrum	Levenberg-Marquardt	ACC=0.8333
			Time Difference Cepstrum	LSTM	ACC=0.9166
			Delta Cepstrum	Levenberg-Marquardt	ACC=0.917
			Delta Cepstrum	LSTM	ACC=0.75
[39]	2018	Columbia X-Cultural Deception	LIWC (Multiple Turns)	Random Forest	ACC=0.7278; F1=0.7274
			Lexical (Multiple Turns)	Random Forest	ACC=0.7033; F1=0.7025
			LIWC+Lexical (Multiple Turns)	Random Forest	ACC=0.7166; F1=0.7158
			LIWC+Individual (Multiple Turns)	Random Forest	ACC=0.7185; F1=0.7179
			Lexical+Individual (Multiple Turns)	Random Forest	ACC=0.6995; F1=0.6986
			LIWC+Lexical+Individual (Multiple Turns)	Random Forest	ACC=0.724; F1=0.7233
[34]	2017	Columbia X-Cultural Deception	TRIGRAMS	LR	F1=0.6119
			OPENSMILE09	Random Forest	F1=0.5954
			OPENSMILE09 + TRIGRAMS	Random Forest	F1=0.581
			OPENSMILE13	DNN	F1=0.6071
			OPENSMILE09	DNN	F1=0.6271
			MFCC	BLSTM	F1=0.5464
			WE	BLSTM	F1=0.6046
			OPENSMILE09 + WE	HYBRID	F1=0.639
[32]	2021	Bol (Set-A)	Audio	Softmax Classfier	ACC=0.9008
		Bol (Set-B)	Audio	Softmax Classfier	ACC=0.93
		RL Trail	Audio	Softmax Classfier	ACC=0.9445
		MU3D	Audio	Softmax Classfier	ACC=0.958
[28]	2018	[26]	Vocal (IS09)	DEV Framework	ACC=0.715
			Vocal (IS13)	DEV Framework	ACC=0.7005
			Vocal (DEV-Vocal)	DEV Framework	ACC=0.7416
[27] **(Top Results)**	2020	Political Videos Collected From The Internet	Vocal (Glove)	Decision Tree	ACC=0.61
			Vocal (POS)	Decision Tree	ACC=0.61
			Vocal (Polarity Scores + Unigrams + POS + Glove + LIWC	Decision Tree	ACC=0.66
			Acoustic (Top 20 Features (Deception In Spoken Dialogue: Classification And Individual Differences))	Decision Tree	ACC=0.51
			Acoustic (IS09)	Decision Tree	ACC=0.53
			Acoustic (IS13)	Decision Tree	ACC=0.63
			Acoustic (IS09+IS13)	Decision Tree	ACC=0.56
[31]	2020	[26]	Pitch (Std)	SVM	ACC=0.6158; AUC=0.6507
			Pitch (Std)	Random Forest	ACC=0.7119; AUC=0.7939
			Pitch (Std)	NN	ACC=0.5141; AUC=0.7427
			Pitch (Mean)	SVM	ACC=0.5424; AUC=0.5223
			Pitch (Mean)	Random Forest	ACC=0.5311; AUC=0.5465
			Pitch (Mean)	NN	ACC=0.6102; AUC=0.5235
			Sil.Sp.Hist	SVM	ACC=0.5763; AUC=0.4159
			Sil.Sp.Hist	Random Forest	ACC=0.5932; AUC=0.7069
			Sil.Sp.Hist	NN	ACC=0.5593; AUC=0.6483
			All Acoustic	SVM	ACC=0.565; AUC=0.5864
			All Acoustic	Random Forest	ACC=0.6328; AUC=0.7059
			All Acoustic	NN	ACC=0.6102; AUC=0.6589
[40]	2021	[26]	Audio (Pitch, Frame Count, Various Durations, Among Others)	KNN	ACC=0.6; F1=0.66
[30]	2018	[26]	Audio	Multi-Layer Perceptron	ACC=0.5238

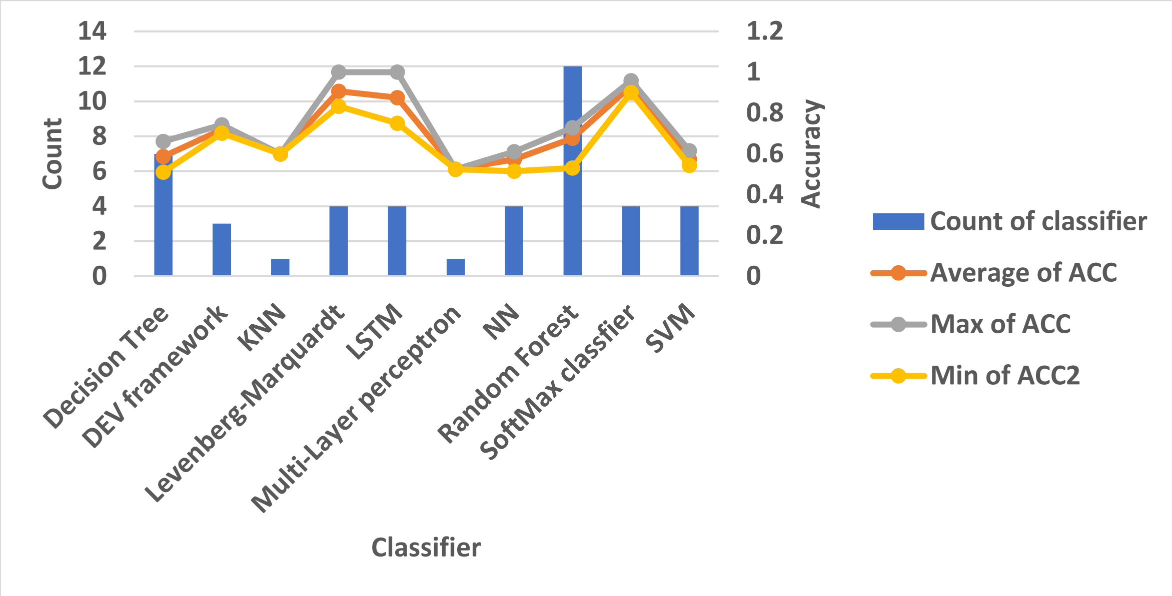

Figure 4. Classifiers used and average accuracy for each in audio modal

VIII. Body Movement Based Deception Detection

Similar to the facial expressions-based DD, Body movement-based DD tracks and analyzes the changes in body posture, hand and leg movement, head movement among other features that correlate with deception [41]. These features are extracted using computer vision and fed to a classification algorithm. Although the investigation of body movement as means to detect deception hasn't been researched as much as the other modals, recent literature has shown promising results with accuracies of up to 91% using Fisher-LSTM classifier from hand gesture features alone [42], other attempts such as T.O. Meservy et al. [43] used hand, arm and head features with accuracies of 71% using SVM classifier. Table 4 lists the used datasets, features and classifiers for DD based on body movement features. A comparison among used classifiers in this modal is shown in Figure 5. When paired with other modals however, body movement features can be a valuable addition in a multimodal approach with accuracies of up to 96% using MLP classifier with Audio, Micro Expressions, Text and Video features [32].

Table 4. Body Movement features used to detect deception

Paper	Year	Dataset	Feature	Classifier	Classification Performance
[42]	2021	[26]	Hand Gestures	Fisher-LSTM	ACC=0.9096; AUC=0.9114
[27]	2020	Politifact.Com	Pose	Decesion Tree	ACC=0.53;
[31]	2020	[26]	Hand Gestures	SVM	ACC=0.5028; AUC=0.7232
				Random Forest	ACC=0.6497; AUC=0.6671
				NN	ACC=0.6158; AUC=0.693
[40]	2021	[26]	Mostly Head and Hand Movement (In Addition To Lips Movement)	KNN	ACC=0.94; F1=0.94;
[44]	2020	Experimental	Gait	LSTM	ACC=0.7274
			Gestures	LSTM	ACC=0.6159
			Gestures, Gait	LSTM	ACC=0.7774
			Deep Features	LSTM	ACC=0.8267
			All	LSTM	ACC=0.8841

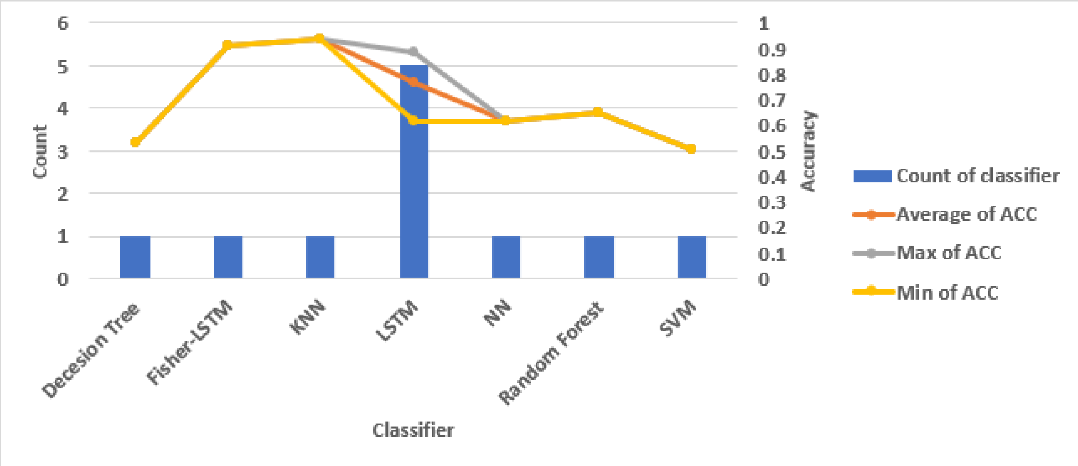

Figure 5. Classifiers used and average accuracy for each in body movement modal

IX. Thermal Imaging based Deception Detection

Thermal imaging, is a non-invasive approach to the temperature based DD found in the polygraph, where the subjects face is recorded using a thermal sensor to track changes in facial temperature throughout the interview which reflects stress levels of the subject, stress levels can be instantaneous changes which manifest in changes in the periorbital blood flow or it can be sustained which manifests in changes in blood flow in the forehead [45] with the first being more informative and reflective of deception [46] [47]. Capturing changes in facial temperature is done using an infrared camera to record the subject's face during the interview to produce a video with an added layer of thermal data on top of the RGB and Audio data being captured. Thermal data is then processed by selecting and tracking specific regions in the face such as the eyes and forehead to produce a feature vector that can be used for pattern analysis and classification. Accuracies of up to 86.88% have been reached using KNN classifier [48] and up to 91.7% using Binary Logistic Regression (LR) [47]. Table 5 lists the deployed datasets, features and classifiers for this modal, while Figure 6 presents average accuracies for each thermal imaging DD classifier. Although thermal imaging can yield promising results in the field of DD, attempts at ADD using machine learning for classification as well as pairing it with other DD modals for a multimodal approach are lacking in the literature to say the least, leaving a lot of unexplored potential for this approach as evidenced by the success of the machine learning based classifiers and multimodal attempts previously discussed.

Table 5. Thermal Imaging features used to detect deception

Paper	Year	Dataset	Feature	Classifier	Classification Performance
[42]	2021	[26]	Hand Gestures	Fisher-LSTM	ACC=0.9096; AUC=0.9114
[27]	2020	Politifact.Com	Pose	Decesion Tree	ACC=0.53;
[31]	2020	[26]	Hand Gestures	SVM	ACC=0.5028; AUC=0.7232
				Random Forest	ACC=0.6497; AUC=0.6671
				NN	ACC=0.6158; AUC=0.693
[40]	2021	[26]	Mostly Head and Hand Movement (In Addition To Lips Movement)	KNN	ACC=0.94; F1=0.94;
[44]	2020	Experimental	Gait	LSTM	ACC=0.7274
			Gestures	LSTM	ACC=0.6159
			Gestures, Gait	LSTM	ACC=0.7774
			Deep Features	LSTM	ACC=0.8267
			All	LSTM	ACC=0.8841

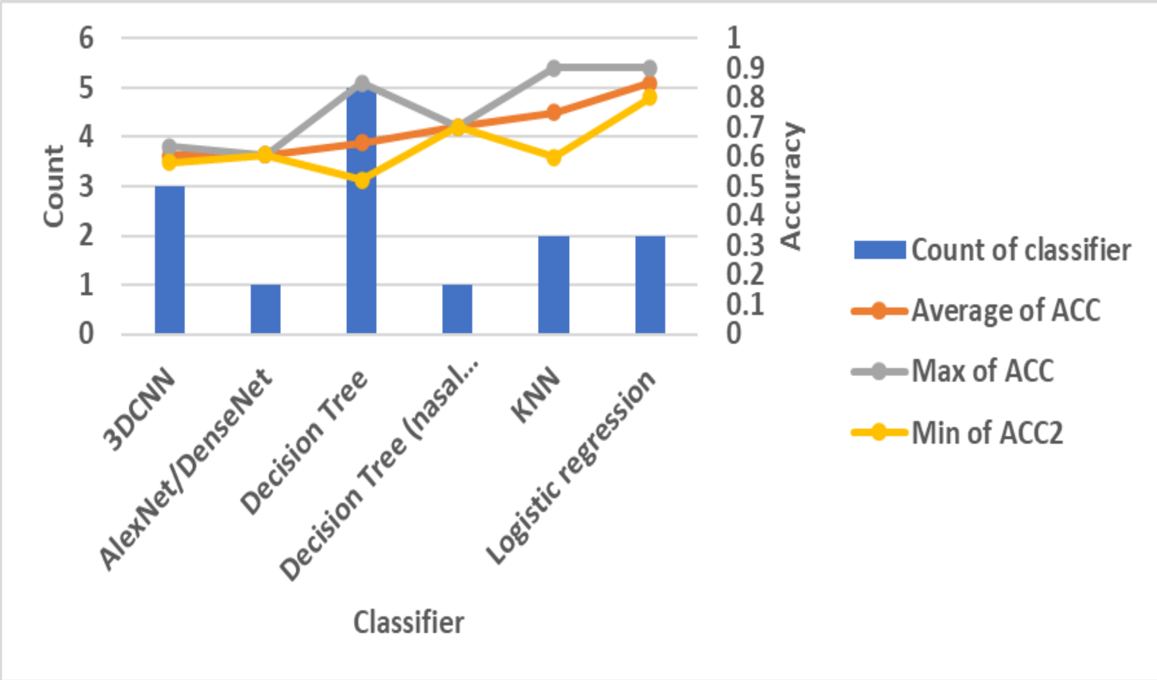

Figure 6. Classifiers used and average accuracy for each in thermal imaging modal

X. MULTIMODAL APPROACH FOR DECEPTION DETECTION

As previously discussed, attempting a multimodal approach can significantly improve the classification accuracy of DD; however, it isn't clear that it guarantees improvement to the classification performance. For instance, a recent study by Kamboj et al. [27] has achieved an accuracy of 70% using a combination of lexical, acoustic and visual features, or in the case of Şen et al. [31], 72% accuracy was achieved when combining all visual, acoustic and linguistic features as opposed to 84.18% with only visual and acoustic modals. This may be attributed to some modals having less discriminative power compared to others depending on the approach, and methodology. In addition, used classifier algorithm could play a big role on the outcome. This is evident by the research conducted by Kamboj et al. [27] and Şen et al. [31], where the low performance of the first one was rationalized by the authors to be due to acoustic features having inherently low discriminating power. Meanwhile, the second work had its highest performance when combining acoustic features with the facial features. This suggests that the multimodal approach, while proven effective in many of the recent literature such as in the case of Karnati et al. [32], who obtained an accuracy of up to 95%-98% using a Deep Convolutional Neural Networks (DCNN) based framework as a feature extractor and classifier combining video (facial expressions), audio (acoustic only) and EEG modals; It is evident that the multimodal approach is not a perfect solution out of the box. Table 6 and Figure 7 depict the deployed dataset, features, classifiers, and average accuracies for each multimodal approach.

Table 6. Multimodal approach used to detect deception

Paper	Year	Dataset	Feature	Classifier	ACC
[27] (Top Results)	2020	Politifact.Com	Visual Aus, Gaze, Pose; Acoustic; Lexical Glove	Decision Tree	ACC=0.63
			Visual Gaze, Emotion; Acoustic IS13; Lexical All	Decision Tree	ACC=0.58
			Visual Gaze, Emotion, Pos	Decision Tree	ACC=0.69
[31] (Top Results)	2020	[26]	Facial Displays, Acoustic Features	NN	ACC=0.8418
			Facial Displays, Acoustic Features, Linguistic Features	Random Forest	ACC=0.7853
			Facial Displays, Pitch, Sil.Sp.Hist	NN	ACC=0.8305; AUC=0.9166
			Facial Displays, Pitch	NN	ACC=0.8249; AUC=0.9462
			Facial Dispalys, All Acoustic Features	SVM	ACC=0.8232; AUC=0.8604
[32]	2021	Bol (Set-A)	Video, Audio	Softmax Classfier	ACC=0.9173
		Bol (Set-B)	Video, Audio	Softmax Classfier	ACC=0.9604
		RL Trail	Video, Audio	Softmax Classfier	ACC=0.9733
		MU3D	Video, Audio	Softmax Classfier	ACC=0.9814
		Bol (Set-A)	Audio, EEG	Softmax Classfier	ACC=0.9538
		Bol (Set-A)	Video, EEG	Softmax Classfier	ACC=0.9563
		Bol (Set-A)	Video, Audio, EEG	Softmax Classfier	ACC=0.9591
[40]	2021	[26]	Audio, Text	KNN	ACC=0.69; F1=0.75
			Audio, Video	KNN	ACC=0.85; F1=0.85
			Text, Video	KNN	ACC=0.65; F1=0.66
			Audio, Video, Text	KNN	ACC=0.78; F1=0.79
[54] (Top Results)	2018	[26]	Facial Expressions, Gaze, Head Movement, Hand Gestures, Verbal Features	Mvl	ACC=0.89
			Facial Expressions (Extracted Via Alexnet-FT), Gaze, Head Movement, Hand Gestures, Verbal Features	SVM	ACC=0.99
			Facial Expressions (Extracted Via Alexnet-FT), Gaze, Head Movement, Hand Gestures, Verbal Features	LMKL	ACC=0.99
			Facial Expressions (Extracted Via Alexnet-FT), Gaze, Head Movement, Hand Gestures, Verbal Features	MVL	ACC=0.98
[55]	2019	[26]	Audio, Text, Micro Expressions	SRKDA For Audio, Linear SVM For Text and Adaboost For Facial Expressions; Fuse The Results Of Each Individual Modal Using Majority Voting For Final Decision	ACC=0.97

[56]	2021	[26]	Audio, Visual (Early Fusion)	KNN	ACC=0.64; AUC=0.6; F1=0.69
			Audio, Visual (Late Fusion)	KNN	ACC=0.63; AUC=0.65; F1=0.68
			Visual, Acoustic, Verbal Features	FFCSN (An Adversial Learning Module)	ACC=0.97; AUC=0.9978
[53] (Top Results)	2019	Experimental (Males And Females, Abortion Topic)	Linguistic, Thermal	Decision Tree	ACC=0.635
		Experimental (Females, Abortion Topic)	Linguistic, Visual	Decision Tree	ACC=0.613
		Experimental (Males , Abortion Topic)	Thermal, Visual	Decision Tree	ACC=0.598
		Experimental (Females, Abortion Topic)	Thermal, Visual	Decision Tree	ACC=0.736
		Experimental (Males And Females, Abortion Topic)	Linguistic, Thermal, Visual	Decision Tree	ACC=0.63
		Experimental (Males And Females, Best Friend Topic)	Linguistic, Thermal	Decision Tree	ACC=0.611
		Experimental (Males, Best Friend Topic)	Linguistic, Visual	Decision Tree	ACC=0.608
		Experimental (Males, Best Friend Topic)	Thermal, Visual	Decision Tree	ACC=0.647
		Experimental (Female, Best Friend Topic)	Linguistic, Thermal, Visual	Decision Tree	ACC=0.569
		Experimental (Females, Mock Crime Topic)	Linguistic, Thermal	Decision Tree	ACC=0.698
		Experimental (Females, Mock Crime Topic)	Linguistic, Visual	Decision Tree	ACC=0.717
		Experimental (Males, Mock Crime Topic)	Thermal, Visual	Decision Tree	ACC=0.549
		Experimental (Females, Mock Crime Topic)	Linguistic, Thermal, Visual	Decision Tree	ACC=0.679
		Experimental (Females, All Topics)	Linguistic, Thermal	Decision Tree	ACC=0.626
		Experimental (Females, All Topics)	Linguistic, Visual	Decision Tree	ACC=0.728
		Experimental (Females, All Topics)	Thermal, Visual	Decision Tree	ACC=0.551
		Experimental (Females, All Topics)	Linguistic, Thermal, Visual	Decision Tree	ACC=0.619
		Experimental (Females, Abortion Topic)	Linguistic, Thermal, Visual	SVM	ACC=0.623
		Experimental (Female, Best Friend Topic)	Linguistic, Thermal, Visual	SVM	ACC=0.66
		Experimental (Females, Mock Crime Topic)	Linguistic, Thermal, Visual	SVM	ACC=0.604
		Experimental (Males And Females, All Topics)	Linguistic, Thermal, Visual	SVM	ACC=0.619

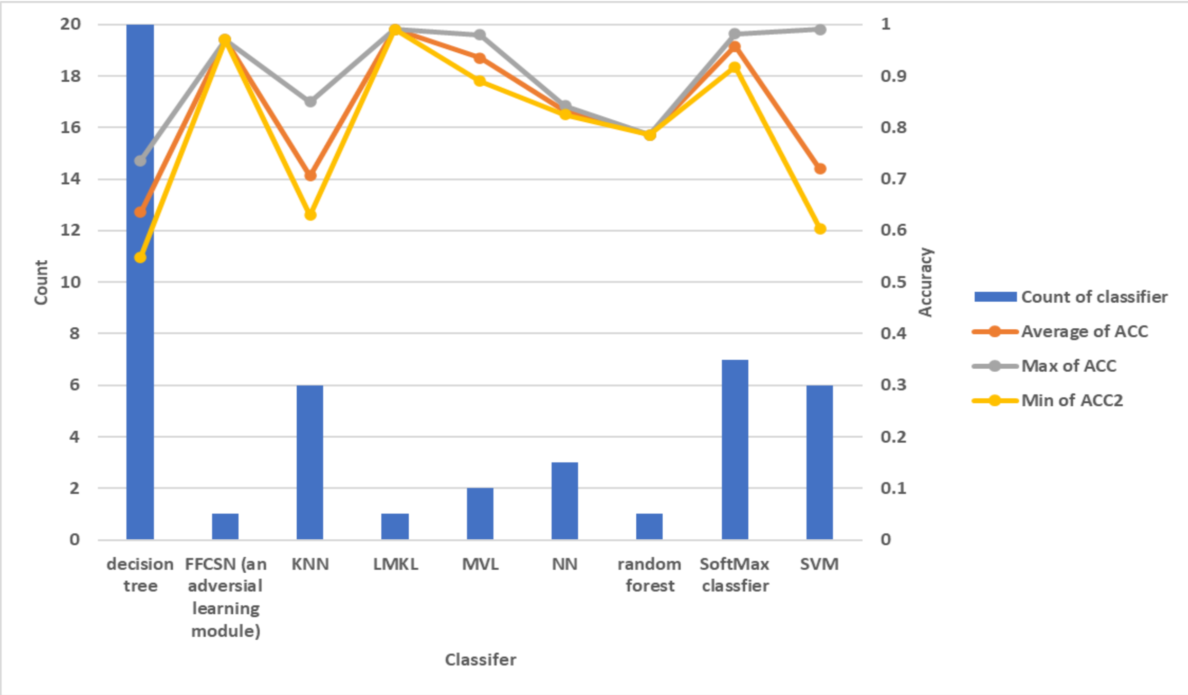

Figure 7. Classifiers used and average accuracy for each in multimodal approach

XI. CONCLUSIONS

Multimodal approach to deception detection appears to be the future since each modal can only get so far on its own in real situations with high stakes concerning homeland security or the court rooms. This is especially true when considering that all the systems implemented in the literature were trained on small data sets and/or fully controlled environments to produce the best results possible, a problem that the entire field of DD suffers from. This calls for the need to diversify the modals that the system can work with to maximize accuracy and produce reliable results regardless of the quality and amount of given data required for analysis. Furthermore, the problem of small data sets that are being worked with for training and testing is a major challenge that needs to be addressed before the field of DD can truly realize its potential. A small data set can produce results that may seem impressive in theory but are undependable in real life situations due to the developed model having a very specific set of expectations and requirements to produce ideal results. Many researchers have attempted to overcome this challenge by generating their own dataset by interviewing real subjects and recording their responses. This of course comes with its own set of challenges mentioned previously such as direct contact related issues that need to be addressed via avatar mediated interviewing for example. As well as difficulty to incentivize the participants to lie or conceal the truth with effort to mimic a real-life scenario without a reward of some form (financial or otherwise), these challenges among others have always kept the number of participants low, resulting in a small sample size to work with.

XII. AUTHORS CONTRIBUTION STATEMENT

Fahad Abdulridha: Conceptualization, Methodology, Software, Data curation, Writing- Original draft preparation. **Baraa M. Albaker:** Supervision, Writing - Review & Editing, Investigation, Validation.

XIII. DECLARATION OF COMPETING INTEREST

The authors of this work declare that to their knowledge, there are no competing financial interests nor personal relationships of any nature that would influence or cause bias in this work and the results reported in it.

XIV. DATA USE

The authors declare that the data used in this work requires no informed consent and all the data is publicly available and accessible.

REFERENCES

[1] P. V. Trovillo, "A history of lie detection," J. Crim. Law Criminol., vol. 29, 30, pp. 848–881, 104–119, 1939.
[2] J. A. Larson, Lying and its detection. Oxford, England: Univ. of Chicago Press, 1932, p. 2.

[3] L. Saxe, D. Dougherty, and T. Cross, “The validity of polygraph testing: Scientific analysis and public controversy,” Am. Psychol., vol. 40, pp. 355–366, 1985, doi: 10.1037/0003-066X.40.3.355.

[4] P. Ekman and M. O’Sullivan, “Who can catch a liar?,” Am. Psychol., vol. 46, pp. 913–920, 1991, doi: 10.1037/0003-066X.46.9.913.

[5] P. Ekman, W. V. Friesen, and K. R. Scherer, “BODY MOVEMENT AND VOICE PITCH IN DECEPTIVE INTERACTION,” vol. 16, no. 1, pp. 23–28, Jan. 1976, doi: 10.1515/semi.1976.16.1.23.

[6] J. A. Podlesny and D. C. Raskin, “Physiological measures and the detection of deception,” Psychol. Bull., vol. 84, pp. 782–799, 1977, doi: 10.1037/0033-2909.84.4.782.

[7] “Accuracy of Deception Judgments - Charles F. Bond, Bella M. DePaulo, 2006.” https://journals.sagepub.com/doi/10.1207/s15327957pspr1003_2 (accessed Dec. 13, 2022).

[8] S. V. Fernandes and M. S. Ullah, “A Comprehensive Review on Features Extraction and Features Matching Techniques for Deception Detection,” IEEE Access, vol. 10, pp. 28233–28246, 2022, doi: 10.1109/ACCESS.2022.3157821.

[9] X. Zhou and R. Zafarani, “A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities,” ACM Comput. Surv., vol. 53, no. 5, pp. 1–40, Oct. 2020, doi: 10.1145/3395046.

[10] F. Tomas, O. Dodier, and S. Demarchi, “Computational Measures of Deceptive Language: Prospects and Issues,” Front. Commun., vol. 7, p. 792378, Feb. 2022, doi: 10.3389/fcomm.2022.792378.

[11] M. Zabcikova, Z. Koudelkova, and R. Jasek, “Concealed Information Detection Using EEG for Lie Recognition by ERP P300 in Response to Visual Stimuli: a Review,” WSEAS Trans. Inf. Sci. Appl., vol. 19, pp. 171–179, Sep. 2022, doi: 10.37394/23209.2022.19.17.

[12] E. Saquete, D. Tomás, P. Moreda, P. Martínez-Barco, and M. Palomar, “Fighting post-truth using natural language processing: A review and open challenges,” Expert Syst. Appl., vol. 141, p. 112943, Mar. 2020, doi: 10.1016/j.eswa.2019.112943.

[13] J. Fan and X. Shen, “New progress in the paradigm of elicited deception : Application of human-computer interaction in deception detection,” in 2021 2nd International Conference on Information Science and Education (ICISE-IE), Chongqing, China, Nov. 2021, pp. 1558–1562. doi: 10.1109/ICISE-IE53922.2021.00345.

[14] M. Thangaraj and M. Sivakami, “Text Classification Techniques: A Literature Review,” Interdiscip. J. Inf. Knowl. Manag., vol. 13, pp. 117–135, 2018, doi: 10.28945/4066.

[15] N. Vogler and L. Pearl, “Using linguistically defined specific details to detect deception across domains,” Nat. Lang. Eng., vol. 26, no. 3, pp. 349–373, May 2020, doi: 10.1017/S1351324919000408.

[16] A. Nortje and C. Tredoux, “How good are we at detecting deception? A review of current techniques and theories,” South Afr. J. Psychol., vol. 49, no. 4, pp. 491–504, Dec. 2019, doi: 10.1177/0081246318822953.

[17] D. B. Buller and J. K. Burgoon, “Interpersonal Deception Theory,” Commun. Theory, vol. 6, no. 3, pp. 203–242, 1996, doi: 10.1111/j.1468-2885.1996.tb00127.x.

[18] Z. Dong, G. Wang, S. Lu, L. Dai, S. Huang, and Y. Liu, “Intentional-Deception Detection Based on Facial Muscle Movements in an Interactive Social Context,” Pattern Recognit. Lett., vol. 164, pp. 30–39, Dec. 2022, doi: 10.1016/j.patrec.2022.10.008.

[19] W. Khan, K. Crockett, J. O’Shea, A. Hussain, and B. M. Khan, “Deception in the eyes of deceiver: A computer vision and machine learning based automated deception detection,” Expert Syst. Appl., vol. 169, p. 114341, May 2021, doi: 10.1016/j.eswa.2020.114341.

[20] M. G. Frank and E. Svetieva, “Microexpressions and Deception,” in Understanding Facial Expressions in Communication: Cross-cultural and Multidisciplinary Perspectives, M. K. Mandal and A. Awasthi, Eds. New Delhi: Springer India, 2015, pp. 227–242. doi: 10.1007/978-81-322-1934-7_11.

[21] U. Hess and R. E. Kleck, “Differentiating emotion elicited and deliberate emotional facial expressions,” Eur. J. Soc. Psychol., vol. 20, no. 5, pp. 369–385, 1990, doi: 10.1002/ejsp.2420200502.

[22] M. G. Frank, P. Ekman, and W. V. Friesen, “Behavioral markers and recognizability of the smile of enjoyment,” J. Pers. Soc. Psychol., vol. 64, pp. 83–93, 1993, doi: 10.1037/0022-3514.64.1.83.

[23] L. Mathur and M. J. Matarić, “Introducing Representations of Facial Affect in Automated Multimodal Deception Detection,” in Proceedings of the 2020 International Conference on Multimodal Interaction, New York, NY, USA, Oct. 2020, pp. 305–314. doi: 10.1145/3382507.3418864.

[24] L. Su and M. Levine, “Does ‘lie to me’ lie to you? An evaluation of facial clues to high-stakes deception,” Comput. Vis. Image Underst., vol. 147, pp. 52–68, Jun. 2016, doi: 10.1016/j.cviu.2016.01.009.

[25] M. Ding, A. Zhao, Z. Lu, T. Xiang, and J.-R. Wen, “Face-Focused Cross-Stream Network for Deception Detection in Videos,” in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, Jun. 2019, pp. 7794–7803. doi: 10.1109/CVPR.2019.00799.

[26] V. Pérez-Rosas, M. Abouelenien, R. Mihalcea, and M. Burzo, “Deception Detection using Real-life Trial Data,” in Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, New York, NY, USA, Nov. 2015, pp. 59–66. doi: 10.1145/2818346.2820758.

[27] M. Kamboj, C. Hessler, P. Asnani, K. Riani, and M. Abouelenien, “Multimodal Political Deception Detection,” IEEE Multimed., vol. 28, no. 1, pp. 94–102, Jan. 2021, doi: 10.1109/MMUL.2020.3048044.

[28] H. Karimi, “Interpretable Multimodal Deception Detection in Videos,” in Proceedings of the 20th ACM International Conference on Multimodal Interaction, Boulder CO USA, Oct. 2018, pp. 511–515. doi: 10.1145/3242969.3264967.

[29] M. Jaiswal, S. Tabibu, and R. Bajpai, “The Truth and Nothing But the Truth: Multimodal Analysis for Deception Detection,” in 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW), Barcelona, Spain, Dec. 2016, pp. 938–943. doi: 10.1109/ICDMW.2016.0137.

[30] G. Krishnamurthy, N. Majumder, S. Poria, and E. Cambria, “A Deep Learning Approach for Multimodal Deception Detection.” arXiv, Mar. 01, 2018. Accessed: Dec. 09, 2022. [Online]. Available: http://arxiv.org/abs/1803.00344

[31] M. U. Şen, V. Pérez-Rosas, B. Yanikoglu, M. Abouelenien, M. Burzo, and R. Mihalcea, “Multimodal Deception Detection Using Real-Life Trial Data,” IEEE Trans. Affect. Comput., vol. 13, no. 1, pp. 306–319, Jan. 2022, doi: 10.1109/TAFFC.2020.3015684.

[32] M. Karnati, A. Seal, A. Yazidi, and O. Krejcar, "LieNet: A Deep Convolution Neural Network Framework for Detecting Deception," IEEE Trans. Cogn. Dev. Syst., vol. 14, no. 3, pp. 971–984, Sep. 2022, doi: 10.1109/TCDS.2021.3086011.

[33] S. Amiriparian, J. Pohjalainen, E. Marchi, S. Pugachevskiy, and B. Schuller, "Is Deception Emotional? An Emotion-Driven Predictive Approach," in Interspeech 2016, Sep. 2016, pp. 2011–2015. doi: 10.21437/Interspeech.2016-565.

[34] G. Mendels, S. I. Levitan, K.-Z. Lee, and J. Hirschberg, "Hybrid Acoustic-Lexical Deep Learning Approach for Deception Detection," in Interspeech 2017, Aug. 2017, pp. 1472–1476. doi: 10.21437/Interspeech.2017-1723.

[35] 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, doi: 10.1109/ACCESS.2021.3084200.

[36] 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, doi: 10.1016/j.neucom.2014.04.083.

[37] M. Sanaullah and M. H. Chowdhury, "Neural network based classification of stressed speech using nonlinear spectral and cepstral features," in 2014 IEEE 12th International New Circuits and Systems Conference (NEWCAS), Jun. 2014, pp. 33–36. doi: 10.1109/NEWCAS.2014.6933978.

[38] S. V. Fernandes and M. S. Ullah, "Phychoacoustic Masking of Delta and Time -Difference Cepstrum Features for Deception Detection," in 2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), Oct. 2020, pp. 0213–0217. doi: 10.1109/UEMCON51285.2020.9298117.

[39] S. I. Levitan, A. Maredia, and J. Hirschberg, "Linguistic Cues to Deception and Perceived Deception in Interview Dialogues," in Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), New Orleans, Louisiana, Jun. 2018, pp. 1941–1950. doi: 10.18653/v1/N18-1176.

[40] S. Chebbi and S. B. Jebara, "Deception detection using multimodal fusion approaches," Multimed. Tools Appl., Jun. 2021, doi: 10.1007/s11042-021-11148-9.

[41] T. O. Meservy, M. L. Jensen, W. J. Kruse, J. K. Burgoon, and J. F. Nunamaker, "Automatic Extraction of Deceptive Behavioral Cues from Video," in Terrorism Informatics: Knowledge Management and Data Mining for Homeland Security, H. Chen, E. Reid, J. Sinai, A. Silke, and B. Ganor, Eds. Boston, MA: Springer US, 2008, pp. 495–516. doi: 10.1007/978-0-387-71613-8_23.

[42] D. Avola, L. Cinque, M. De Marsico, A. Fagioli, and G. L. Foresti, "LieToMe: Preliminary study on hand gestures for deception detection via Fisher-LSTM," Pattern Recognit. Lett., vol. 138, pp. 455–461, Oct. 2020, doi: 10.1016/j.patrec.2020.08.014.

[43] T. O. Meservy et al., "Deception detection through automatic, unobtrusive analysis of nonverbal behavior," IEEE Intell. Syst., vol. 20, no. 5, pp. 36–43, Sep. 2005, doi: 10.1109/MIS.2005.85.

[44] T. Randhavane, U. Bhattacharya, K. Kapsaskis, K. Gray, A. Bera, and D. Manocha, "The Liar's Walk: Detecting Deception with Gait and Gesture." arXiv, Mar. 29, 2020. doi: 10.48550/arXiv.1912.06874.

[45] I. Pavlidis, J. Dowdall, N. Sun, C. Puri, J. Fei, and M. Garbey, "Interacting with human physiology," Comput. Vis. Image Underst., vol. 108, no. 1, pp. 150–170, Oct. 2007, doi: 10.1016/j.cviu.2006.11.018.

[46] I. Pavlidis and J. Levine, "Monitoring of periorbital blood flow rate through thermal image analysis and its application to polygraph testing," in 2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Oct. 2001, vol. 3, pp. 2826–2829 vol.3. doi: 10.1109/IEMBS.2001.1017374.

[47] D. A. Pollina et al., "Facial Skin Surface Temperature Changes During a 'Concealed Information' Test," Ann. Biomed. Eng., vol. 34, no. 7, pp. 1182–1189, Jul. 2006, doi: 10.1007/s10439-006-9143-3.

[48] B. A. Rajoub and R. Zwiggelaar, "Thermal Facial Analysis for Deception Detection," IEEE Trans. Inf. Forensics Secur., vol. 9, no. 6, pp. 1015–1023, Jun. 2014, doi: 10.1109/TIFS.2014.2317309.

[49] N. Vance et al., "Deception Detection and Remote Physiological Monitoring: A Dataset and Baseline Experimental Results," IEEE Trans. Biom. Behav. Identity Sci., vol. 4, no. 4, pp. 522–532, Oct. 2022, doi: 10.1109/TBIOM.2022.3218956.

[50] P. Kodavade, S. Bhandigare, A. Kadam, N. Redekar, and K. P. Kamble, "Lie Detection Using Thermal Imaging Feature Extraction from Periorbital Tissue and Cutaneous Muscle," in Innovations in Computer Science and Engineering, vol. 171, H. S. Saini, R. Sayal, A. Govardhan, and R. Buyya, Eds. Singapore: Springer Singapore, 2021, pp. 643–650. doi: 10.1007/978-981-33-4543-0_68.

[51] A. Ravindran, G. G. Krishna, Sagara, and S. Sarath, "Region of Interest Based Comparative Analysis in Deception Detection using Thermal Images," in 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), Kannur,Kerala, India, Jul. 2019, pp. 1600–1604. doi: 10.1109/ICICICT46008.2019.8993169.

[52] A. Ravindran, G. G. Krishna, Sagara, and S. S., "A Comparative Analysis of Machine Learning Algorithms in Detecting Deceptive Behaviour in Humans using Thermal Images," in 2019 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, Apr. 2019, pp. 0310–0314. doi: 10.1109/ICCSP.2019.8697911.

[53] M. Abouelenien, M. Burzo, V. Perez-Rosas, R. Mihalcea, H. Sun, and B. Zhao, "Gender Differences in Multimodal Contact-Free Deception Detection," IEEE Multimed., vol. 26, no. 3, pp. 19–30, Jul. 2019, doi: 10.1109/MMUL.2018.2883128.

[54] N. Carissimi, C. Beyan, and V. Murino, "A Multi-View Learning Approach to Deception Detection," in 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), Xi'an, May 2018, pp. 599–606. doi: 10.1109/FG.2018.00095.

[55] S. Venkatesh, R. Ramachandra, and P. Bours, "Robust Algorithm for Multimodal Deception Detection," in 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), San Jose, CA, USA, Mar. 2019, pp. 534–537. doi: 10.1109/MIPR.2019.00108.

[56] L. Mathur and M. J. Mataric, "Unsupervised Audio-Visual Subspace Alignment for High-Stakes Deception Detection," in ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, Jun. 2021, pp. 2255–2259. doi: 10.1109/ICASSP39728.2021.9413550.