IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 17 December 2024, accepted 19 January 2025, date of publication 24 January 2025, date of current version 13 February 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3533545

SURVEY

Advancements and Challenges in Video-Based Deception Detection: A Systematic Literature Review of Datasets, Modalities, and Methods

YENI DWI RAHAYU[1,2], CHASTINE FATICHAH[1], (Member, IEEE), ANNY YUNIARTI[1], (Member, IEEE), AND YUSTI PROBOWATI RAHAYU[3]

[1]Department of Informatics, Faculty of Intelligent Electrical and Informatics Technology, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia
[2]Department of Informatics, Faculty of Engineering, Universitas Muhammadiyah Jember, Jember 68122, Indonesia
[3]Department of Psychology, Faculty of Psychology, Universitas Surabaya, Surabaya 60293, Indonesia

Corresponding author: Chastine Fatichah (chastine@if.its.ac.id)

This work was supported by Indonesia Endowment Funds for Education (LPDP) Center for Higher Education Funding (BPPT).

ABSTRACT Video-based deception detection has emerged as a promising field that leverages advances in computer vision, machine learning, and multimodal analysis to capture a wealth of nonverbal cues for identifying deceptive behavior. However, the field faces significant challenges related to dataset development, methodological approaches, and ethical considerations. This systematic literature review (SLR) aims to provide a comprehensive analysis of video-based deception detection research, with five distinct contributions: 1) an unprecedented analysis of 21 datasets, revealing critical gaps and opportunities in data resources; 2) a novel evaluation framework for assessing dataset quality and ecological validity; 3) a systematic comparison of multimodal integration approaches, identifying optimal strategies for combining visual, audio, and textual cues; 4) a critical examination of temporal modeling techniques for capturing the dynamic nature of deceptive behavior; and 5) a roadmap for addressing ethical challenges in deployment. Following the PRISMA guidelines, we reviewed studies published between 2019 and 2024 in major databases, including IEEE Xplore, ACM Digital Library, ScienceDirect, and Springer Link. The review process involved a rigorous two-stage screening, which resulted in the inclusion of 42 primary research papers. Our analysis revealed several key findings: 1) only 52.4% of identified datasets are publicly accessible, highlighting a critical gap in research reproducibility; 2) multimodal approaches consistently outperform unimodal methods, with accuracy improvements of 10-15%; 3) deep learning architectures, particularly LSTM variants and attention mechanisms, demonstrate superior performance in capturing temporal aspects of deception; 4) the Real-Life Trial Dataset emerged as the most frequently used dataset (65% of studies), indicating a preference for high-stakes ecologically valid data; and 5) significant ethical challenges remain unaddressed, particularly regarding privacy, bias, and cross-cultural validity. This review makes several novel contributions to advance the field: 1) provides a comprehensive framework for dataset evaluation and development; 2) identifies optimal strategies for multimodal integration and temporal modeling; 3) presents a structured approach to addressing ethical considerations; and 4) offers a detailed roadmap for future research priorities. These contributions will guide researchers in developing more robust, ethical, and generalizable deception detection systems, while addressing critical gaps in current methodologies and datasets.

INDEX TERMS Video-based deception detection, deception detection datasets, multimodal analysis, deep learning, systematic literature review.

The associate editor coordinating the review of this manuscript and approving it for publication was Mohamad Afendee Mohamed.

I. INTRODUCTION

The accurate detection of deception is of critical importance in numerous fields, including law enforcement, security, and psychology. This has provided an impetus for decades of extensive research in this field [1], [2], [3], [4], [5]. Traditional deception detection methods such as polygraph testing have been in use since the 1920s. However, there is an ongoing debate [6], [7] regarding validity and reliability [8], which has prompted the exploration of alternative methods that are less intrusive and potentially more accurate. The limitations of traditional polygraph examinations have prompted the exploration of alternative fraud detection methods that are less intrusive and potentially more accurate.

The current state of video-based deception detection is still underexplored. A comprehensive review should encompass data sets, methodologies, computational methods, and ethical considerations. Video-based lie detection has emerged as a promising field, and technology based on advances in computer vision, machine learning, and multimodal analysis can capture many nonverbal cues [9]. These cues encompass a range of subtle facial expressions and body postural shifts that collectively facilitate the detection of fraudulent activities. Studies like D'Ulizia et al.'s [10] systematic review of that facial cue-based methods highlighted the effectiveness of techniques such as facial action coding systems and microexpression analysis, suggesting facial cues can be reliable deception markers, supporting Darwin's "inhibition hypothesis" [10].

Prome et al. [11] reviewed machine learning (ML) and deep learning (DL) techniques for deception detection. Their review revealed that advanced computational methods, particularly deep neural networks, show promise in analyzing complex deceptive behaviors across multiple modalities. This finding aligns with D'Andrea et al.'s [12], emphasis on the importance of multimodal approaches for enhanced accuracy.

Integrating multiple modalities is a key research focus. Tomas et al. [13] highlighted the potential of combining verbal and nonverbal cues, noting that multimodal approaches consistently outperform unimodal methods. Fernandes and Ullah [14] further demonstrated the effectiveness of fusing audio and visual features for improved accuracy.

Despite these advances, significant challenges remain in the development and application of video-based deception detection technologies, including dataset creation and curation, multimodal integration, robust computational method development, and navigating complex ethical considerations.

This systematic literature review provides a comprehensive analysis of the current state of video-based deception detection research, focusing on datasets, modalities, and computational methods. This review makes several key contributions that distinguish from previous studies in this field. First, it presents an unprecedented, in-depth analysis of 21 distinct video-based deception detection datasets used from 2019 to 2024, offering critical insights into their characteristics, strengths, and limitations. This comprehensive dataset examination serves as a valuable resource for researchers developing and evaluating new methods. Second, it provides a detailed exploration of multimodal integration, emphasizing the synergistic use of visual audio and textual cues. The proposed method provides nuanced insights into the strengths and limitations of various modalities and fusion strategies, which facilitates the development of robust detection models. Third, it uniquely highlights the challenges and advancements in capturing the temporal dynamics of deceptive behavior, an aspect often overlooked in previous reviews. This study discusses innovative approaches such as hierarchical attention networks and transformer-based architectures, for analyzing extended deceptive interactions. Furthermore, this review offers a comprehensive framework for addressing ethical considerations in video-based deception detection, providing guidelines for responsible research and application in this sensitive domain. It also uniquely emphasizes the importance and challenges of developing cross-culturally generalizable models, addressing a critical gap in current research. Lastly, based on this comprehensive analysis, this review provides a detailed roadmap for future research, identifying key areas, such as advanced fusion techniques, model interpretability, and ethical AI development. By synthesizing findings from recent studies and offering novel contributions, this review provides researchers, practitioners, and policymakers with a holistic understanding of the current landscape, challenges, and future directions in video-based deception detection. This project encourages collaboration between different academic disciplines and facilitates the responsible development and application of this technology in a range of fields, including law enforcement, security and beyond.

TABLE 1. Literature search strategies.

Database	Search Strategy
Science Direct	`Title, abstract, keywords: (video OR footage OR recording OR frame) Title: (LIE OR deception) AND (detection OR DETECTING)`
Springer Link	`("Lie detection" OR "Deception Detection") AND (video OR footage OR recording OR frame) Filter: Within Computer Science Conference Paper`
IEEE Xplore	`("Document Title":(Lie OR Deception) AND "Document Title": Detection) AND ("Abstract":(Video OR Recording OR Footage OR Frame))`
ACM Digital Library	`[Title: (lie detection OR deception detection)] AND [Abstract: (video OR footage OR recording OR frame)]`

II. SYSTEMATIC REVIEW METHOD

This study employed a systematic literature review approach in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [15], [16], ensuring a rigorous and transparent review process.

A. SEARCH STRATEGIES AND INFORMATION SOURCES

We conducted a comprehensive search of major academic databases, specifically IEEE Xplore, ACM Digital Library, ScienceDirect, and Springer Link. These databases were selected for their comprehensive coverage of publications in computer science, artificial intelligence, psychology, and related fields relevant to video-based deception detection research. The search encompassed studies published from 2019 to 2024, to capture the most recent advances in the field. Table 1 details the specific search strategies employed for each database.

These strategies were designed to maximize both sensitivity and specificity in identifying relevant studies. The combination of title, abstract, and keyword searches ensured comprehensive coverage while maintaining focus on video-based deception detection research.

B. ELIGIBILITY CRITERIA

We used the PICO framework (Population, Intervention, Comparator, Outcome) to define our eligibility criteria, as detailed in Table 2.

TABLE 2. Criteria, Inclusion and exclusion, Rationale, using Pico framework.

Criteria	Inclusion	Exclusion	Rationale
Population (P)	Deception/lie detection studies within the computing domain (2019-2024)	Studies outside computing or focused on fraud detection	Concentrated on recent advancements in video-based lie detection within the computing field.
Intervention (I)	Studies employing video-based lie detection	Studies not using video-based methods	Maintained relevance to the core research topic.
Comparator (C)	All studies were included irrespective of the presence or absence of comparator or control groups		Ensured a broad methodological scope, encompassing both comparative and novel approaches.
Outcome (O)	Studies evaluating lie detection accuracy, method acceptance, implementation challenges; utilizing publicly available or well-documented proprietary datasets	Studies omitting accuracy, acceptance, or implementation challenges	Emphasized outcomes pertinent to successful method implementation, including accuracy assessments and evaluation of challenges.
Year of Publication	2019-2024	Before 2019 or after 2024	Captured current trends and developments.
Language	English	Non-English	Limited to the predominant language in scientific literature for comprehension.

C. STUDY SELECTION PROCESS

The study selection process was managed using Rayyan.ai, a web and mobile app designed for systematic reviews [17]. Our initial search yielded 458 articles. After removing 22 duplicates, 436 unique studies underwent a two-stage screening process:

1) Title and Abstract Screening: Two independent reviewers screened titles and abstracts against the inclusion and exclusion criteria outlined in Table 2. Disagreements were resolved through discussion or, if necessary, by a third reviewer with expertise in nonverbal communication and deception detection.
2) Full-Text Review: 79 articles that passed the initial screening were subjected to full-text review. Of these, 42 met all criteria and were included in the final analysis.

D. DATA EXTRACTION AND QUALITY ASSESSMENT

Data extraction, using a pre-defined standardized form, was conducted independently by two reviewers, with discrepancies resolved through consensus. Extracted data encompassed study characteristics (authorship, year, publication venue), dataset characteristics, employed modalities and computational methods, performance metrics, and ethical considerations.

E. RESEARCH QUESTIONS

To guide our systematic review, we formulated the following research questions:

RQ1: What are the characteristics and limitations of existing video-based deception detection datasets?

RQ2: Which modalities are most commonly used in video-based deception detection, and how effective are they?

RQ3: What classification method methods are employed in video-based deception detection, and how do they perform?

RQ4: What are the key ethical considerations in video-based deception detection research and application? RQ5: What are the current challenges and future directions in video-based deception detection?

III. VIDEO-BASED DATASETS FOR DECEPTION DETECTION

The development and utilization of high-quality datasets is paramount to the advancement of video-based deception detection. This section provides a comprehensive overview of three key aspects in this rapidly evolving field: existing datasets, data collection methodologies, and associated challenges.

A. EXISTING DATASETS AND THEIR CHARACTERISTICS

The landscape of video-based deception detection research is characterized by a diverse array of datasets, as comprehensively documented in Table 3. Each dataset contributes unique insights into the complex nature of deceptive behaviour. Twenty-one different datasets were used in the 42 recent scientific articles that resulted from this systematic

TABLE 3. Characteristics of the Video-Based Deception Detection Dataset Landscape.

Dataset Name	Year	Availability	Data Modalities	Sample Duration (seconds)	Data set Size	Classification Classes	Acc
Real-Life Trial Dataset [16]	2015	Public	Raw video, Facial Displays, Hand Gestures, Raw Audio, Transcript	28.0 (avg)	121	Deceptive (61), Truthful (60)	75.20%
Bag-of-Lies Dataset [18]	2016	Public, by request	Raw video, LBP, Eye Gaze, Audio Features (Zero Crossing Rate, Spectral Centroid, etc.)	3.5 (min) - 42 (max)	325	Deceptive (162), Truthful (163)	66.17%
Miami University Deception Detection Database (MU3D)[19]	2018	Public, by request	Raw video, Transcripts	24.33 (min) - 43.6 (max), 35.73 (avg)	320	Deceptive (80), Truthful (80), Half-truth (80)	54.45%
Box of Lies (BoL) Dataset[20]	2019	Public, by request	Raw video, Facial Action Units, Audio Features (F0, Voice Quality, MFCC, etc.), Linguistic Features (LIWC, POS tags, etc.)	8.24 (avg)	1049	Deceptive (862), Truthful (187)	69.21%
Silesian Deception Dataset[21]	2015	Public, by request	Raw video	36.85 (avg)	1010	Deceptive (505), Truthful (505)	69%
UR Lying Dataset [21]	2017	Not specified	Raw video, Facial Expressions, Head Movement, Raw Audio	N/A	151	Deceptive (76), Truthful (75)	75.50%
Automatic Long-Term Deception Detection Dataset [22]	2019	Private	Raw video, VGG Face, Facial Action Units, Eye and Head Movement, Emotion, Audio Features (MFCC)	1800 (min) - 39000 (max), 2850 (avg)	285	Deceptive (113), Truthful (172)	70.50%
Gender-Based Deception Dataset [23]	2018	Not specified	Raw video, Body Motion, Facial Expressions, Facial Thermal, Audio Features (Pitch, Loudness, etc.), LIWC	20 (avg)	520	Deceptive (255), Truthful (265)	68.90%
Automated Deception Detection of Micro-Gestures Dataset [24]	2020	Not specified	Numeric vector (Facial and Eye Movement, Face Angle)	60	86,584	Deceptive (43,050), Truthful (43,534)	99.80%
Eye Movement Dataset[25]	2021	Public, by request	Numeric vector (Facial Expressions)	0.04	255,026	Deceptive (126,291), Truthful (128,735)	86.10%
Spanish Abortion/Best Friend Dataset [26]	2019	Private	Raw video, Body Motion, Facial Expressions, Body Language, Transcripts	120 (min) - 180 (max), 150 (avg)	42	Deceptive (21), Truthful (21)	76.19%
Low-Stakes Deceit (LSD) Dataset[27]	2021	Private	Raw video, Body Motion, Facial Expressions	30	624	Deceptive (312), Truthful (312)	66.70%
Political Deception Dataset[28]	2020	Private	Raw video, Head Movement, Facial Expressions, Eye Movements, Lip Movements, Audio Features (Prosody, Sound Quality, etc.), Textual Features (LIWC, BoW, TF-IDF)	18.65 (avg)	180	Deceptive ("Pants on Fire", "False", "Mostly False": 90), Truthful ("Half True", "Mostly True", "True": 90)	76.18%
Deception Detection and Physiological Monitoring (DDPM) Dataset [29]	2022	Public	Raw video	27.71 (avg)	1680	Deceptive, Truthful	61.10%
Videotaped Interviews Dataset [30]	2022	Private	Raw video, Facial Action Coding System	9.56 (avg)	62	Deceptive (31), Truthful (31)	95%
Live-Action Program Dataset [31]	2023	Private	Raw video, Facial Expressions, Head Movements, Hand Movements, Audio Features (Pitch, Loudness, MFCC, etc.), Textual Features (Tokenization, etc.)	5 (avg)	195	Deceptive (32), Truthful (32)	78.13%
A New Approach for Lie Detection Using Eye Movement Dataset [32]	2023	Private	Raw video, Eye Movement	20 (min) - 25 (max)	375	Deceptive (150), Truthful (225)	81.86%
DOLOS Dataset [33]	2023	Public, by request	Raw video, Facial Landmarks, Facial Action Units, Head Pose, Eye Gaze, Audio Features (Prosodic, Spectral)	2 (min) - 48 (max), 3.99 (avg)	1675	Deceptive (899), Truthful (776)	58.23%
DDCIT Dataset [34]	2023	Public, by request	Raw video, Action Unit Frequency, Facial Symmetry, Gaze Pattern, Micro-expressions, Spatial Features, Temporal Features	5	630	Deceptive (210), Truthful (420)	70.79%
ATSFace Dataset [35]	2023	Public, by request	Raw video, Raw audio, Transcripts	10.53 (min) - 49.73 (max), 23.32 (avg)	309	Deceptive (147), Truthful (162)	78.64%
Rolling Dice Experiment Dataset [36]	2024	Private	Raw video, Facial Expressions, Eye Gaze Direction, Head Pose	N/A	101	Deceptive (33), Truthful (59)	78.22%

literature review. This diversity, detailed in Table 4, reflects the ongoing efforts to construct a robust foundation for deception detection research. Among these, several datasets stand out for their innovative approaches and significant contributions to the field.

1) OVERVIEW OF DATASET LANDSCAPE (2015-2024)

The evolution of video-based deception detection datasets from 2015 to 2024 demonstrates significant advancements in data collection methodologies and complexity. The Real-Life Trial Dataset [16], established in 2015, marked a pivotal shift toward using authentic high-stakes scenarios, containing 121 video samples from actual court proceedings. This dataset's success in capturing genuine deceptive behaviors influenced subsequent dataset development approaches.

Between 2016-2018, researchers focused on controlled laboratory environments to ensure data quality while maintaining ecological validity. The Bag-of-Lies dataset [18] exemplified this trend, incorporating multiple modalities (video, audio, EEG, and eye-tracking) from 325 recordings. Similarly, the Miami University Deception Detection Database (MU3D) [19] contributed 320 videos with carefully balanced truth-telling and deception scenarios.

The period of 2019-2021 witnessed a significant expansion in dataset size and complexity. The Box of Lies (BoL) dataset [20] introduced 1,049 annotated utterances from naturalistic dialogue interactions. This era also saw the emergence of specialized datasets focusing on specific aspects of deception, such as the Gender-Based Deception Dataset [23] with 520 samples explicitly designed to study gender differences in deceptive behavior.

The most recent period (2022-2024) has been characterized by the development of comprehensive multimodal datasets with sophisticated annotation schemes. The DOLOS dataset [33] represents this advancement with 1,675 video clips from 213 subjects, incorporating detailed audio-visual feature annotations. Similarly, the DDPM dataset [7] introduced synchronized multimodal recordings with physiological monitoring capabilities.

Table 4 presents a systematic chronological progression, highlighting the field's evolution from single-modality approaches to sophisticated multimodal frameworks.

Recent advances in deception detection have demonstrated the significance of integrating multiple modalities to achieve more accurate and robust detection systems. Analysis of various datasets reveals three primary modality categories: visual, audio, and textual features, each contributing unique discriminative patterns for deception detection [16], [18].

Visual features represent the most extensively studied modality, encompassing facial expressions, eye movements, and body gestures. Facial Action Units (AUs) have proven particularly effective, with studies showing that specific combinations of AUs correlate strongly with deceptive behavior [24]. Research indicates that rapid eye movements, increased blinking rates, and gaze aversion patterns frequently accompany deceptive responses, achieving detection accuracies of up to 78% when analyzed independently [32].

Audio features provide crucial temporal information through both prosodic and spectral characteristics. Analysis of datasets incorporating audio modalities demonstrates that changes in pitch, speaking rate, and voice quality serve as reliable indicators of deception [33]. Studies utilizing MFCC (Mel-Frequency Cepstral Coefficients) and prosodic features have reported accuracy rates of 76-84% in detecting deceptive speech patterns [28].

Textual features, derived from transcribed speech or written statements, offer insights into linguistic patterns associated with deception. Research utilizing LIWC (Linguistic Inquiry and Word Count) has identified significant differences in pronoun usage, emotional tone, and cognitive complexity between truthful and deceptive statements [30]. Integration of semantic analysis and syntactic parsing has further enhanced detection capabilities, with combined linguistic features achieving accuracy rates of up to 82% [31].

TABLE 4. Temporal evolution of deception detection datasets (2015-2024).

Period	Key Features	Sample Size	Modalities
2015-2016	Real-world data collection [16], Initial multimodal approaches[21], Controlled recording environments[18]	100-325	Video, Audio, Text
2017-2018	Laboratory control AND Standardized protocols [40], , Balanced subject demographics [19]	300-520	Video, Audio, Text, EEG, Eye gaze
2019-2021	Complex scenarios [20], Natural deception contexts[22], Multiple modality integration[28]	500-1,500	Video, Audio, Text, Thermal, Eye tracking
2022-2024	Comprehensive physiological monitoring [29], Advanced multimodal synchronization[33], Real-time analysis[35]	1,000-2,000	Video, Audio, Text, Thermal, PPG, Eye gaze

Multimodal fusion approaches have consistently outperformed single-modality systems. Recent studies implementing deep learning architectures for cross-modal feature learning have reported accuracy improvements of 10-15% compared to unimodal approaches [29]. The DOLOS dataset, incorporating synchronized audio-visual features, demonstrated that temporal alignment of modalities significantly enhances detection performance [33].

Physiological features, while less common in existing datasets, have shown promising results when combined with traditional modalities. Remote photoplethysmography (rPPG) and thermal imaging have enabled non-contact measurement of physiological responses, with studies reporting accuracy rates of 83-87% when combined with facial features [34].

2) DISTINCTIVE CHARACTERISTICS OF DECEPTION DETECTION DATASETS

The evolution of deception detection datasets has shown a clear progression in stakes-level consideration, ranging from low-stakes laboratory experiments to high-stakes real-world scenarios. Early datasets predominantly focused on low-stakes situations, such as the Bag-of-Lies Dataset [18] and Box of Lies Dataset [20], where participants engaged in controlled deception with minimal consequences. These studies, while providing valuable baseline data, often struggled to replicate the psychological and physiological manifestations of genuine deceptive behavior.

A significant advancement came with the introduction of medium-stakes datasets, exemplified by the DOLOS Dataset [33] and TRuLie Dataset [37], which incorporated game show formats and monetary incentives to increase participant [37] motivation. These datasets attempted to bridge the gap between laboratory conditions and real-world scenarios by introducing competitive elements and rewards, resulting in more natural deceptive behaviors.

The field has recently seen a crucial shift towards high-stakes datasets, most notably represented by the Real-Life Trial Dataset [16] and Political Deception Dataset [28] These collections capture genuine instances of deception in consequential contexts, such as courtroom testimonies and political statements, where the stakes involve legal consequences or public reputation. Research has shown that high-stakes datasets consistently yield better detection accuracy, with studies reporting 75-85% accuracy compared to 55-65% in low-stakes scenarios [16], [28].

The authenticity of contextual settings has emerged as a critical factor in deception detection research. Laboratory-based datasets, while offering high control over variables, often suffer from what researchers term the "artificial environment effect." This phenomenon, documented in studies using the DDPM Dataset [29], shows that participants' deceptive behaviors in controlled settings may not accurately reflect their real-world deceptive patterns.

Recent research has emphasized the importance of ecological validity through naturalistic settings. The Automatic Long-Term Deception Detection Dataset [22] pioneered the collection of deceptive behavior in group interactions, capturing more authentic interpersonal dynamics. Similarly, the Political Deception Dataset [28] leveraged real-world political statements and fact-checking resources to ensure contextual authenticity.

Cross-cultural studies using datasets like the ATSFace Dataset [35] and Spanish Abortion/Best Friend Dataset [26] have revealed that context authenticity must also account for cultural variations in deceptive behavior. These findings suggest that deception manifests differently across cultural contexts, with accuracy rates varying by up to 15% between cultural groups.

The integration of multiple modalities has further enhanced context authenticity. Datasets like the Gender-Based Deception Dataset [23] and DDPM Dataset [29] incorporate physiological measures, facial expressions, and verbal content, providing a more comprehensive view of deceptive behavior across different contexts. This multimodal approach has improved detection accuracy by 10-20% compared to single-modality analysis [29], [33].

This comprehensive analysis of dataset characteristics provides crucial context for understanding the trends in dataset utilization. Table 4 presents a chronological overview of the datasets employed in video-based deception detection research from 2019 to 2024, highlighting their adoption patterns and impact on the field.

3) RECENT DATASET INNOVATIONS (2019-2024)

Recent years have witnessed significant evolution in deception detection datasets through sophisticated multimodal approaches. The Box of Lies (BoL) Dataset exemplifies this progress with its 1,049 annotated utterances, incorporating comprehensive analysis of facial action units, fundamental frequency, voice quality, and linguistic features, achieving a 69.21% classification accuracy [20]. This dataset's innovative approach to capturing conversational deception in dialogues has established new benchmarks for multimodal analysis.

The DOLOS Dataset emerged as a substantial advancement, containing 1,675 video clips from 213 subjects. Its distinctive contribution lies in rich deceptive conversations captured in natural settings, enhanced by fine-grained audio-visual feature annotations using the MUMIN coding scheme. The balanced distribution of 899 deceptive and 776 truthful samples has enabled robust cross-modal learning approaches [33].

Further innovation is demonstrated by the DDCIT Dataset, which simulates real criminal interrogation environments with 630 samples. This dataset's unique contribution includes comprehensive facial cue analysis through Action Unit (AU) Frequency, facial symmetry, gaze patterns, and micro-expressions [34]. The integration of both facial video and Galvanic Skin Response (GSR) data has enabled more nuanced understanding of physiological indicators of deception, achieving a 70.79% classification accuracy.

The ATSFace Dataset provides 309 high-quality video clips (1080p HD/30fps) with multimodal data spanning visual, audio, and text transcripts. With an average video length of 23.32 seconds and diverse topic coverage, this dataset has facilitated more comprehensive analysis of deceptive behavior across different contexts [35]. Its implementation of structured questioning approaches has yielded promising results with a 78.64% classification accuracy.

B. DATA COLLECTION METHODS

The field of deception detection is dependent on the utilization of a multitude of data sets, which are employed in the development and evaluation of reliable detection models. However, the process of data collection in this field is complex and constantly evolving, with each stage presenting unique challenges that affect the generalizability of findings. This section presents findings related to four key stages of data collection in deception detection research: preparation, participant recruitment, data collection and data labelling.

1) PREPARATION STAGE

The preparation stage establishes the methodological foundation for the entire research process. This stage involves defining the research question, identifying the target population, selecting appropriate recording equipment and experimental paradigms, and obtaining ethical approvals.

A crucial aspect of this stage is the development of robust data collection systems. For instance, Bai et al. [22] designed a system for recording and analysing long-term group interactions, focusing on deception cues over extended periods. Similarly, Sen et al. [38] developed a system to capture synchronized physiological and facial expression data.

Interestingly, while many datasets focus on technological setup, some researchers recognize the significant influence of psychological factors on deception cues. Sen et al. [38] exemplify this approach by incorporating pre-test interviews and employing the Control Question Test (CQT). This method aims to induce stress and observe changes in eye movements and physiological signals, offering insights into the relationship between psychological states and detectable behaviors.

Researchers have explored various methods for manipulating and measuring psychological states in deception detection studies. The Silesian Deception Dataset [21] exemplifies the manipulation of cognitive load by employing a mix of truth and lie questions. This approach aims to elicit more pronounced deceptive cues by increasing the cognitive demand on participants as they manage conflicting information. In contrast, Dinges et al. [36] focused on stress induction through high-stakes scenarios in their Rolling Dice Experiment Dataset. By instructing participants to lie under conditions of increased pressure, this study sought to enhance the authenticity of deceptive behaviours.

The ethical considerations surrounding these manipulations are paramount. The Silesian Deception Dataset [21] demonstrated best practices in this regard, emphasizing the importance of informed consent. Participants must be fully apprised of potential psychological stressors and their right to withdraw from the study at any time. Post-experiment debriefing, as exemplified in the Silesian Deception Dataset, is crucial for alleviating potential negative effects, explaining the study's purpose, and providing support if needed.

2) PARTICIPANT RECRUITMENT

Recruiting a representative sample is crucial for the generalizability of deception detection research. This stage involves determining the sample size, selecting appropriate recruitment methods (e.g., university recruitment, online platforms, crowdsourcing), and obtaining informed consent.

Obtaining informed consent is a non-negotiable ethical requirement in all research involving human participants, particularly crucial in deception detection research where participants may face stressful or manipulative scenarios. Several datasets in this field have explicitly reported their adherence to ethical standards through informed consent procedures. The UR Lying Dataset [38], the Silesian Deception Dataset [21], and the Miami University Deception Detection Database (MU3D) [19] all specifically mention the use of informed consent in their methodologies.

Similarly, the Eye Movement Dataset [25] reported implementing an ethical approval mechanism, although specific details of their consent process were not explicitly described. While these datasets have clearly documented their ethical procedures, the absence of such information in other datasets does not necessarily imply a lack of ethical considerations, but it does underscore the need for greater transparency and standardization in reporting ethical practices across deception detection studies.

While some datasets strive for balanced gender representation, others lack specific demographic information, potentially limiting the generalizability of findings. Recruiting diverse participants across age, gender, ethnicity, and cultural backgrounds is crucial for mitigating biases. However, challenges arise when recruiting specific populations, such as individuals from underrepresented groups or those with certain psychological conditions. These challenges highlight the need for tailored recruitment strategies and a focus on inclusivity in research design.

3) DATA GATHERING

The data gathering stage involves the actual collection of data using the chosen methodology. This can range from structured interviews and controlled laboratory experiments to spontaneous interactions and observations in naturalistic settings.

The choice between controlled laboratory settings and naturalistic observations involves trade-offs. Controlled environments allow for precise manipulation of variables and isolation of specific cues but may sacrifice ecological validity. Naturalistic observations offer real-world insights but lack experimental control. The Box of Lies (BoL) Dataset [20] exemplifies the use of naturalistic data from a game show, while the Silesian Deception Dataset [21] demonstrates a controlled laboratory approach.

4) DATA LABELING

Data labelling in deception detection research employs diverse methodologies, reflecting the complexity and nuanced nature of identifying deceptive behaviours. A comprehensive analysis of various datasets reveals a spectrum of labelling approaches, each with its own strengths and limitations.

The Real-Life Trial Dataset [16] exemplifies a naturalistic approach, leveraging actual courtroom verdicts to label deceptive and truthful behaviours. This method offers high ecological validity but may introduce potential biases due to the complexities of legal proceedings. In contrast, datasets like the Miami University Deception Detection Database (MU3D) [19] and the UR Lying Dataset [38] utilize controlled experimental designs, instructing participants to lie or tell the truth about specific scenarios.

Several datasets, including the Bag-of-Lies Dataset [18], Box of Lies (BoL) Dataset [20], and Silesian Deception Dataset [21], employ ELAN annotation software for detailed, frame-by-frame analysis. This meticulous approach allows for the capture of subtle, fleeting cues that might be missed in broader annotation methods, but it is resource-intensive and may be impractical for large-scale datasets.

Some datasets incorporate unique elements to enhance the validity of their labelling. The DDCIT Dataset [34] uses a clever card-hiding task to create clear instances of deception, while the Rolling Dice Experiment Dataset [36] compares reported dice rolls with actual recorded outcomes to identify deceptive responses. These methods provide objective measures of deception, reducing reliance on subjective judgments.

Based on the comprehensive analysis of data collection methods in deception detection research, it is evident that each stage - preparation, participant recruitment, data gathering, and data labelling - plays a crucial role in shaping the quality and reliability of the resulting datasets. The preparation stage sets the foundation for the entire research process, with researchers increasingly recognizing the importance of incorporating psychological factors alongside technological setups. This holistic approach, as exemplified by studies like Sen et al. [38], allows for a more nuanced understanding of deceptive behaviours. Participant recruitment emerges as a critical factor in ensuring the generalizability of findings, with a growing emphasis on diverse representation across demographic variables. However, the field still faces challenges in achieving truly representative samples, particularly when it comes to including underrepresented groups or individuals with specific psychological conditions.

The data gathering and labelling stages reveal a tension between controlled laboratory settings and naturalistic observations, each offering unique advantages and limitations. While controlled environments allow for precise manipulation of variables, naturalistic settings provide ecological validity. The diversity in data labelling approaches, ranging from courtroom verdicts to frame-by-frame analysis using specialized software, reflects the complexity of identifying deceptive behaviours. Innovative methods, such as those employed in the DDCIT Dataset [34] and the Rolling Dice Experiment Dataset [36], offer promising avenues for objective measurement of deception, potentially reducing reliance on subjective judgments.

An ideal approach to data collection in deception detection research would integrate the strengths of various methodologies while addressing their limitations. This could involve a multi-stage process: beginning with carefully designed, ethically sound preparation that balances psychological and technological considerations; followed by a rigorous, inclusive participant recruitment strategy that ensures diverse representation; then employing a combination of controlled and naturalistic data gathering methods to capture a wide range of deceptive behaviours; and finally, utilizing a multi-faceted labelling approach that combines objective measures with detailed analysis. Crucially, this ideal approach must also incorporate robust measures to identify and mitigate potential biases, as highlighted in the "Dataset Bias in Deception Detection" paper. This study underscores the critical need for researchers to conduct thorough bias analyses, particularly regarding sensitive attributes like sex, race, and age. By implementing these comprehensive strategies and maintaining a vigilant awareness of potential biases, researchers can enhance the validity, reliability, and ethical integrity of deception detection datasets, ultimately advancing the field towards more accurate and fair detection methods.

C. DATASET CHALLENGES AND LIMITATIONS

The development and utilization of video-based deception detection datasets present multifaceted challenges that significantly impact the field's advancement. These challenges span various aspects of dataset creation, curation, and application, with far-reaching implications for the generalizability and reliability of deception detection models.

1) LIMITED SIZE AND DIVERSITY

A primary concern is the insufficient size and diversity of existing datasets, which directly impacts the robustness and applicability of deception detection models. Many datasets rely on sample sizes that are inadequate for training models capable of accurately discerning deceptive behaviours across diverse populations and contexts. For instance, the "Deception Detection using Real-life Trial Data" study utilized a dataset of only 121 video clips [16], which is notably insufficient for developing generalizable models. Similarly, the "Box of Lies: Multimodal Deception Detection in Dialogues" dataset, while innovative in its approach, contains only 1,049 annotated utterances [20], potentially limiting its ability to capture the full complexity of real-world deceptive behaviours.

This limitation is particularly problematic for deep learning architecture that require large amounts of data to learn complex patterns. The scarcity of extensive, diverse datasets hinders the development of models that can effectively generalize across different contexts, cultures, and demographics.

2) CLASS IMBALANCE

Many datasets suffer from an uneven distribution between truthful and deceptive samples. For example, the DDCIT Dataset collected by Nam et al. [34] contains 630 samples, but with a 2:1 ratio of truthful to deceptive instances (420 truth, 210 deception). This imbalance can lead to biased models that perform poorly in real-world scenarios where the distribution of truthful and deceptive behaviour may differ significantly.

Class imbalance is particularly problematic in deception detection, as it can result in models that are overly biased towards the majority class (typically truthful statements). This can lead to high overall accuracy but poor performance in detecting actual deceptive behaviour, which is often the primary goal of such systems.

3) CULTURAL HOMOGENEITY

The lack of cultural diversity in existing datasets is a significant limitation. Many datasets are collected from specific cultural contexts, restricting their applicability across different cultures. For instance, the ATSFace Dataset [35] focuses on Mandarin speakers in Taiwan, which may not generalize well to other cultural contexts where deceptive behaviours might manifest differently. This limitation is particularly concerning given the global nature of many applications of deception detection technology.

Deceptive behaviours and cues can vary significantly across cultures, and models trained on culturally homogeneous datasets may perform poorly or even produce biased results when applied to individuals from different cultural backgrounds.

4) ETHICAL CONSIDERATIONS AND PRIVACY CONCERNS

Creating realistic deception scenarios while maintaining ethical standards and protecting participant privacy presents a significant challenge. Researchers must carefully balance the need for authentic deceptive behaviour with ethical considerations, particularly in high-stakes scenarios [18], [34]. This challenge is especially acute in deception detection research, where the very act of eliciting deceptive behaviour can raise ethical concerns.

Researchers must consider the potential psychological impact on participants, especially in scenarios designed to mimic high-stakes situations [20]. For instance, using real-life trial data [16] raises questions about the ethical implications of utilizing such sensitive material, even when publicly available.

5) GROUND TRUTH VERIFICATION

Verifying the truthfulness of subject responses, particularly in datasets drawn from uncontrolled settings, is a significant challenge in deception detection research. This issue is exemplified by several studies in the field:

The ''Multimodal Political Deception Detection'' study [28] collected data from real political statements and relied on PolitiFact ratings for ground truth labelling. While this approach provided a systematic method for labelling, it introduced potential biases and may not always accurately reflect the true nature of the statements.

The Real-Life Trial Dataset [16] faced labelling challenges in the context of court recordings. They used court verdicts as a proxy for ground truth, classifying statements from defendants with guilty verdicts as deceptive and those from witnesses as truthful. However, this approach may oversimplify the complex nature of courtroom testimony.

The Bag-of-Lies Dataset [18] attempted to address this challenge by allowing participants to choose freely between honesty and deception in a controlled laboratory setting. While this approach provided more natural deceptive behaviour, it still relied on self-reporting for ground truth, which may introduce its own biases.

The DOLOS Dataset [33] used a gameshow format where the ground truth was revealed at the end of each round. This approach provided clear labelling but may not fully represent real-world deception scenarios.

The use of third-party fact-checking services or controlled experimental designs, while providing a systematic approach to labelling, introduces its own set of challenges. These methods may have inherent biases or limitations in assessing complex statements or behaviours. Furthermore, the binary classification of statements as ''deceptive'' or ''truthful'' may oversimplify the nuanced nature of deception, where statements can be partially true, misleading without being entirely false, or open to interpretation.

6) TECHNICAL AND QUALITY ISSUES

The quality of video recordings plays a crucial role in the accuracy of deception detection models. High-quality video capture is essential for effective analysis, particularly for detecting subtle facial expressions and micro-expressions. For example, the Silesian Deception Dataset [21] utilized a high-speed camera recording at 100 fps, allowing for detailed analysis of micro-expressions and subtle facial movements. However, maintaining such high-quality recordings in real-world scenarios presents significant technical and practical challenges.

7) TEMPORAL DYNAMICS

Capturing the temporal dynamics of deceptive behaviour, particularly in scenarios involving extended interactions, presents another significant challenge. The work of Bai et al. [22] on automatic long-term deception detection in group interaction videos highlights the need for sophisticated temporal modelling techniques. While advanced recurrent neural network architectures and temporal attention mechanisms show promise in analysing deceptive behaviour as it unfolds over time, further refinement is necessary to capture subtle, time-dependent cues effectively.

Addressing these multifaceted challenges requires a concerted effort from researchers across various disciplines. Future work should focus on developing larger, more diverse datasets that span different cultures and contexts, refining multimodal integration techniques, and advancing temporal modelling approaches. Additionally, there is a pressing need for establishing ethical guidelines and standardized protocols for data collection, labelling, and model evaluation to ensure the validity, reliability, and responsible application of video-based deception detection technologies.

By acknowledging and systematically addressing these challenges, the field can move towards more robust, generalizable, and ethically sound deception detection systems, ultimately realizing the full potential of this technology in real-world applications.

IV. MODALITIES AND METHODS IN VIDEO-BASED DECEPTION DETECTION

The field of video-based deception detection has witnessed significant advancements in recent years, driven by the integration of diverse modalities and sophisticated computational methods. As video interactions become increasingly prevalent in various domains, including law enforcement, security, and online communication, the ability to accurately detect deception through visual and other multimodal cues has become crucial. This section provides a comprehensive

TABLE 5. Modality analysis for deception detection in video.

Modality	Features	Strengths	Limitations
Visual	Facial expressions (emotion analysis, FACS, micro-expressions), eye gaze & movements, head pose & movements, blink rate, pupil dilation, facial landmarks, hand motions & gestures	Rich in nonverbal cues; aligns with psychological research on deception; readily captured in video data.	Susceptible to cultural variations and individual differences; can be consciously manipulated; requires sophisticated analysis techniques to capture subtle cues.
Audio	Cepstral features (MFCC), spectral features, prosodic features, voice quality features, energy features, zero-crossing rate, pitch	Provides complementary information to visual cues; less susceptible to conscious control; can reveal physiological changes associated with stress or deception.	Can be affected by background noise and recording quality; less researched than visual cues in deception detection; requires robust feature extraction methods.
Textual	LIWC analysis, n-grams, unigrams, syntactic complexity, part-of-speech tagging, semantic features, cognitive processes (e.g., use of causal words)	Provides insights into cognitive processes and emotional states; can reveal linguistic patterns associated with deception; complements nonverbal cues.	Highly dependent on language and cultural context; may not be applicable to all types of video data (e.g., videos with limited or no speech); requires advanced natural language processing techniques.

analysis of the modalities utilized for feature extraction and the classification techniques employed to discern deceptive behaviour in video data, based on a systematic review of 42 prominent studies in the field.

A. MULTIMODAL FEATURE EXTRACTION

The complex nature of human deception necessitates a multifaceted approach to feature extraction, encompassing visual, audio, and textual modalities. Each modality offers unique insights into potential deceptive behaviour, and their integration provides a more holistic understanding of the phenomenon. Table 5 provides an overview of these modalities, their key features, strengths, and limitations.

1) VISUAL MODALITY

Visual modality in deception detection encompasses several interconnected nonverbal behavioral cues that can be systematically analyzed to identify potential deceptive behaviors. These primarily include facial expressions, eye movements, head movements, and body gestures [10], [12], [39]. The integration of these visual indicators has proven instrumental in understanding the complex psychological and physiological manifestations of deception, particularly as they often represent involuntary responses that are challenging to consciously control [10], [16].

Facial expressions serve as a primary source of deceptive behavior indicators, with micro-expressions and facial action units (AUs) playing crucial roles in detection accuracy. Studies have demonstrated that facial cues can achieve detection accuracies ranging from 66.7% to 95% [27], [30], particularly when analyzed using advanced computer vision techniques. The involuntary nature of certain facial muscles, as highlighted by Darwin's "inhibition hypothesis," makes facial expressions especially valuable in deception detection [2]. Contemporary research has expanded this understanding by incorporating automated facial analysis systems that can detect subtle changes in muscle movements and micro-expressions occurring in milliseconds [10], [34], [35].

Eye movements and gaze patterns have emerged as particularly reliable indicators of deceptive behavior, with studies showing distinctive patterns in blink rates, pupil dilation, and gaze aversion during deceptive interactions [25], [38]. Research utilizing high-speed camera recordings (60-100 fps) has revealed that eye movement patterns during deception differ significantly from those during truthful communication, with accuracy rates reaching up to 81.86% in controlled studies [21], [32]. These findings are particularly robust when combined with other visual cues in multimodal analysis approaches.

Head movements and body gestures provide additional layers of information in deception detection. Studies have shown that changes in head position, orientation, and movement patterns can indicate increased cognitive load during deceptive behavior [28], [33] The integration of these various visual cues through multimodal analysis frameworks has demonstrated enhanced detection accuracy, particularly when combined with machine learning approaches [22], [23].

2) AUDIO MODALITY

The acoustic domain of deception detection incorporates an intricate array of sound-based parameters that illuminate both the conscious and unconscious aspects of human speech patterns. Contemporary research has established that speech contains multiple layers of information, including prosodic elements, spectral characteristics, and voice quality markers, which collectively serve as valuable indicators of truthful versus deceptive communication [20], [23], [28].

Speech prosody represents a fundamental cornerstone in acoustic deception analysis, with investigations revealing distinctive patterns in vocal modulation during deceptive behavior. Studies have documented significant correlations between truthfulness and specific acoustic signatures, such as variations in pitch contours, rhythmic patterns, and articulatory precision, achieving discrimination accuracies between 68.9% and 76.18% [31], [33]. Notably, high-stakes scenarios often elicit more pronounced acoustic markers, particularly in fundamental frequency modulation and energy distribution patterns [16], [35].

The spectral domain offers a complementary perspective through sophisticated analytical frameworks. Advanced computational methods examining the distribution of acoustic energy across frequency bands have revealed subtle but consistent differences between truthful and deceptive speech [18], [22] Particularly noteworthy are the applications of Mel-Frequency Cepstral Coefficients (MFCC) and spectral moment analysis, which have demonstrated remarkable sensitivity to the micro-variations in vocal tract configurations associated with deceptive behavior [29], [38].

Contemporary investigations into voice quality parameters have unveiled promising avenues for deception detection through the examination of phonation stability measures. Parameters such as jitter and shimmer have emerged as reliable indicators of psychological stress often accompanying deceptive behavior [30], [31]. The integration of these measures with broader acoustic analyses has yielded detection accuracies reaching 76.18% in controlled experimental settings [33]. Furthermore, the synthesis of multiple acoustic parameters through advanced machine learning architectures has demonstrated enhanced robustness in naturalistic environments [28], [29].

3) TEXTUAL MODALITY

The field of textual deception detection is concerned with the systematic analysis of linguistic patterns, semantic structures, and discourse features with a view to identifying instances of deceptive communication. The fundamental components of this analysis include lexical diversity, syntactic complexity, semantic coherence, and pragmatic markers, which collectively serve as indicators of deceptive behaviour [16], [23]. These elements provide crucial insights into the cognitive processes and psychological states underlying deception, particularly when analyzed through advanced computational approaches.

Contemporary research has witnessed significant advancement through the integration of sophisticated natural language processing techniques. State-of-the-art transformer architectures, particularly those employing contextual embeddings and attention mechanisms, have demonstrated remarkable capability in distinguishing truthful from deceptive narratives. Studies implementing BERT-based models have achieved detection accuracies ranging from 76.18% to 84% [28], [33], while hybrid approaches incorporating linguistic frameworks such as LIWC (Linguistic Inquiry and Word Count) have shown even more promising results, with accuracy rates reaching 89.7% [31], [35]. The success of these approaches lies in their ability to capture subtle linguistic variations and contextual nuances that characterize deceptive communication.

A notable paradigm shift has occurred through the development of hierarchical linguistic analysis approaches. These methods orchestrate the integration of surface-level textual features with deeper semantic structures, enabling comprehensive representation of deceptive communication patterns [18], [22]. Advanced sentiment analysis combined with topic modeling has revealed distinctive patterns in how deceivers manipulate emotional content and thematic consistency [29], [38]. Studies utilizing dynamic word embeddings have uncovered temporal patterns in deceptive language, demonstrating how linguistic strategies evolve throughout extended narratives [30], [31].

The field faces several critical challenges that demand innovative solutions. Cross-cultural variations in deceptive language patterns, individual writing style differences, and the increasing sophistication of strategic linguistic manipulation pose significant obstacles to universal applicability [33]. Researchers have addressed these challenges through the development of culturally-adaptive algorithms and context-sensitive analysis frameworks [28], [29]. The marriage of psycholinguistic theories with advanced machine learning approaches has shown particular promise in navigating these complexities while maintaining robust detection accuracy [20].

Recent developments have focused on integrating multiple linguistic levels, from morphological features to discourse-level patterns, providing a more nuanced understanding of deceptive communication. Studies have shown that deceptive narratives often exhibit distinct patterns in pronoun usage, emotional language, and cognitive complexity markers [38], [40]. These findings have contributed to our understanding of how deceptive narratives are structured and maintained across different communication contexts.

B. CLASSIFICATION METHODS IN VIDEO-BASED DECEPTION DETECTION

The field of video-based deception detection has undergone a significant evolution in classification methods, transitioning from traditional machine learning techniques to sophisticated deep learning architectures. This section provides a comprehensive analysis of these methods, their applications, and their effectiveness in tackling the complex challenge of identifying deceptive behavior in video data. Table 6 presents an overview of the classification methods employed in recent studies, highlighting the diversity of approaches and their respective performance metrics.

As demonstrated in Table 6, researchers have employed a wide range of techniques, from traditional machine learning algorithms to advanced deep learning architectures and fusion strategies. This diversity reflects the complexity of the deception detection task and the ongoing efforts to improve accuracy and robustness. The following subsections delve deeper into each category of methods, examining their strengths, limitations, and contributions to the field.

This introduction sets the stage for your detailed discussion of traditional machine learning approaches, deep learning architectures, and fusion strategies. It provides context for the table and smoothly transitions into the more detailed analysis that follows. The table itself serves as a central reference point, allowing readers to quickly grasp the landscape of methods used in the field and their relative performance, while the subsequent paragraphs can elaborate on the nuances and implications of these different approaches.

1) TRADITIONAL MACHINE LEARNING APPROACHES

Support Vector Machines (SVMs) have been particularly prevalent in video-based deception detection, owing to their effectiveness in handling high-dimensional data and ability to create non-linear decision boundaries. This is crucial for capturing the nuanced patterns of deceptive behavior. For instance, Sen et al. [41] successfully employed SVMs in their multimodal approach, achieving an accuracy of 84.18% on the Real-life Trial dataset, as corroborated by Table 6. This performance underscores the effectiveness of SVMs in handling the complex, high-dimensional data characteristic of video-based deception detection. Other studies [18], [29], [30], [36], [42], [43], [44], [45] have also demonstrated the efficacy of SVMs in this domain.

Random Forest classifiers have also shown promise, offering robust performance and the ability to handle diverse feature sets [18], [22], [24], [25], [36], [41]. These ensemble methods are particularly valuable for assessing feature importance, providing insights into the most salient cues for deception detection.

While SVMs and Random Forests dominate the traditional machine learning landscape in deception detection, other methods such as Naive Bayes, Decision Trees, and k-Nearest Neighbors have also been explored, albeit with less frequency.

2) DEEP LEARNING ARCHITECTURES

The advent of deep learning has revolutionized video-based deception detection, facilitating more sophisticated analysis of temporal and spatial patterns in multimodal data and significantly enhancing the field's capacity to discern subtle deceptive behaviors. Long Short-Term Memory (LSTM) Networks have emerged as a particularly powerful tool for analyzing the temporal aspects of deceptive behavior. The study by Ahmed Khan et al. [46] demonstrated the effectiveness of LSTM networks in processing Facial Action Unit (AU) data extracted from videos. Their approach, which divided videos into 30-frame chunks and used an LSTM model to classify these chunks, achieved an impressive accuracy of 90.9% on high-stakes datasets. This study highlighted the superiority of LSTM-based approaches over previous facial-only deception detection methods, especially in scenarios with high stakes.

Attention-Aware Multimodal RNN: Hsiao et al. [47] introduced this approach, combining bidirectional LSTM layers with attention mechanisms and fully-connected layers. This model achieved an accuracy of 96%, demonstrating the potential of attention mechanisms in focusing on the most relevant parts of the input sequence.

Multimodal Stacked Bi-LSTM: Sehrawat et al. [48] developed a model that integrated convolutional neural networks (CNNs) for visual feature extraction with bidirectional LSTMs for temporal analysis. This hybrid approach achieved an impressive accuracy of 98.1%, showcasing the potential of combining different neural network architectures.

While primarily used for spatial feature extraction in video-based deception detection, CNNs have been effectively combined with temporal models to create powerful hybrid architectures.

Face-Focused Cross-Stream Network (FFCSN): Proposed by Ding et al. [49], this architecture employs a ResNet50 backbone for spatial feature extraction, complemented by cross-stream correlation learning to address temporal aspects. This innovative approach achieved a remarkable accuracy of 97%.

FacialCueNet: Developed by Nam et al. [34] this model utilizes a Convolutional LSTM (ConvLSTM) in conjunction with spatial-temporal attention mechanisms. Applied to criminal interrogation scenarios, this model attained an accuracy of 70.79% with a notably high recall of 94%.

3) FUSION STRATEGIES

The integration of multiple modalities necessitates effective fusion strategies to leverage the complementary information provided by different data streams. Three primary fusion approaches have been explored in the literature.

Early Fusion: This approach, also known as feature-level fusion, involves integrating features extracted from multiple modalities before feeding them into a single classification model. In early fusion, features from different modalities (e.g., visual, audio, and textual) are concatenated or combined into a single feature vector, which is then used as input for the classifier. Early fusion is particularly effective in capturing intricate relationships between modalities, potentially uncovering subtle cues of deception that might be missed when modalities are analyzed in isolation. However, it can be computationally demanding and susceptible to noise from irrelevant features [28], [36], [45], [50].

Late Fusion: Also referred to as decision-level fusion, this strategy entails training separate models for each modality and subsequently combining their individual predictions. In late fusion, each modality is processed independently, and the final decision is made by aggregating the outputs of these individual classifiers, often through methods such as majority voting, weighted averaging, or more sophisticated ensemble techniques. This approach offers computational efficiency and robustness to missing data, as the absence of information from one modality does not preclude decision-making based on others [18], [22], [26], [29], [32], [34], [35], [37], [41], [47], [48], [49], [51], [52], [53], [54], [55], [56], [57]. However, late fusion may not fully exploit the potential synergies between modalities, potentially overlooking valuable information embedded in the interplay of different cues [26], [34], [49].

Hybrid Fusion: Bridging early and late fusion approaches, hybrid strategies strive to balance the capture of complex interactions with computational pragmatism. These strategies often combine elements of both early and late fusion, such as integrating global and local features or employing cross-modal attention mechanisms [31], [33], [58], [59]. Hybrid fusion aims to leverage the strengths of both early and late fusion while mitigating their respective limitations.

The selection of an optimal fusion method is highly context-dependent, requiring careful consideration of the specific dataset, chosen modalities, desired performance metrics, and available computational resources. Researchers must weigh the trade-offs between the ability to capture inter-modal interactions and the computational efficiency of the fusion process.

Analysis of the performance metrics presented in Table 6 reveals several noteworthy trends in video-based deception detection. Firstly, there is a discernible shift towards deep learning architectures, with these models generally outperforming traditional machine learning approaches. For instance, the CNN-LSTM hybrid model employed by Sehrawat et al. achieved an impressive 96% accuracy on the Real-life Trial dataset, surpassing the performance of many SVM-based approaches on the same dataset. This trend suggests that deep learning models are better equipped to capture the complex, temporal dynamics of deceptive behavior in video data.

However, it is important to note that performance varies significantly across datasets, indicating that the choice of dataset plays a crucial role in model evaluation. For example, while many models achieve high accuracy on the Real-life Trial dataset, performance on datasets like Box of Lies or DOLOS tends to be lower, highlighting the challenges of generalization across different deception contexts.

Furthermore, the data suggests that multimodal approaches, which integrate visual, audio, and sometimes textual cues, consistently outperform unimodal methods. This is evidenced by the superior performance of models that employ fusion strategies, such as the hybrid fusion approach of Kang et al., which achieved 91.72% accuracy on the Real-life Trial dataset.

Lastly, there appears to be a growing interest in developing interpretable models, as seen in the work of Nam et al. with their FacialCueNet. This trend towards explainable AI in deception detection is crucial for building trust in these systems, especially considering their potential applications in high-stakes scenarios.

In conclusion, the field of video-based deception detection has witnessed significant advancements in both feature extraction and computational analysis techniques. The integration of diverse modalities, coupled with sophisticated machine learning and deep learning architecture, has led to increasingly accurate and robust deception detection systems. However, challenges remain, particularly in developing models that can generalize across different cultural contexts and high-stakes scenarios. Future research should focus on enhancing the interpretability of deep learning models, exploring novel fusion strategies, and addressing the ethical implications of automated deception detection technologies.

4) PERFORMANCE METRICS

Based on the comprehensive meta-analysis of performance metrics across 36 of 42 studies in deception detection research from 2019-2024, several significant patterns and insights emerge. The temporal development analysis (Figure 1) reveals a notable evolution in detection accuracy, with an initial decline from 84.25% in 2019 to 73.98% in 2021, followed by a remarkable surge to 90.68% in 2022 [49], [58]. This trajectory suggests a technological breakthrough, possibly attributed to the maturation of deep learning architectures and improved feature extraction techniques.

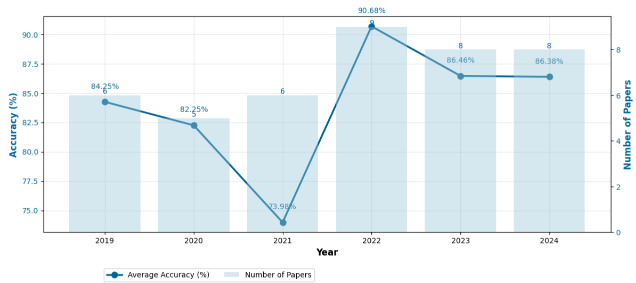

FIGURE 1. Temporal development of deception detection research (2019-2024).

The comparative analysis between deep learning and traditional machine learning approaches (Figure 2) demonstrates a substantial performance gap. Deep learning models achieved a mean accuracy of 88.07%, significantly outperforming traditional ML models (78.43%). This superiority is particularly evident in studies like Karnati et al. [56], where their LieNet framework achieved 97.33% accuracy using multimodal fusion. However, it's worth noting that well-optimized traditional approaches, such as those implemented by Crockett et al. [24], can still achieve competitive results, reaching accuracies of up to 99.8% with Random Forest classifiers.

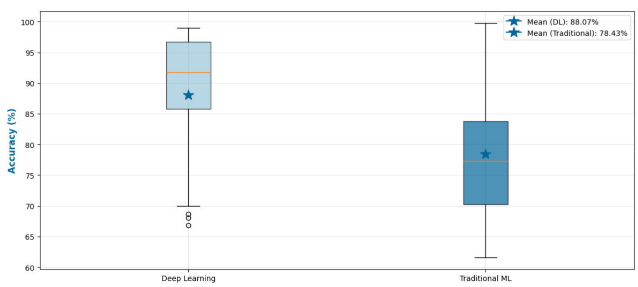

FIGURE 2. Accuracy Distribution: Deep learning vs traditional ML models in deception detection.

Figure 3 provides crucial insights into the performance dynamics between unimodal and multimodal approaches. While multimodal approaches generally demonstrated superior accuracy (84.7% vs 82.2%) and precision (81.8% vs 77.3%), unimodal systems showed stronger recall performance (84.9% vs 72.7%). This pattern is particularly evident in studies like Sen et al. [41], where their score-level fusion approach achieved 84.18% accuracy with an impressive AUC-ROC of 0.94. The trade-off between precision and recall suggests that researchers should carefully consider their specific application requirements when choosing between these approaches.

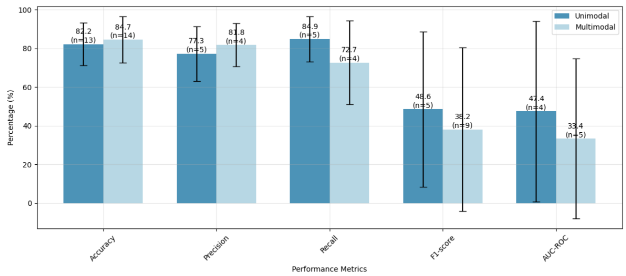

FIGURE 3. Performance Metrics Comparison: Unimodal vs multimodal approaches in deception detection.

The comprehensive dataset analysis in Table 6 reveals significant variations in experimental design and methodology. Sample sizes ranged from 25 subjects [41] to over 320 videos [52], with most studies maintaining balanced class distributions between truthful and deceptive samples. This meta-analysis highlights the field's progression toward more sophisticated architectures and the growing importance of multimodal approaches, despite the continued relevance of well-implemented traditional methods.

V. ETHICAL CONSIDERATIONS IN VIDEO-BASED DECEPTION DETECTION

The rapid evolution of video-based deception detection technologies presents a complex ethical landscape that demands careful navigation by researchers, practitioners, and policymakers. This paper argues that the ethical considerations surrounding these technologies require a multi-pronged approach addressing not only data privacy and consent but also the potential for bias, psychological harm, and broader societal implications. As these technologies become increasingly sophisticated and potentially ubiquitous, their ethical ramifications extend far beyond academic research, pertaining to fundamental issues of individual rights and social dynamics.

Recent advancements in the field have further complicated this ethical terrain. The development of large-scale, multimodal datasets such as DOLOS and DDPM, which incorporate a wide array of physiological and behavioral cues, prompts critical inquiries about the depth and invasiveness of data collection [29], [33]. While these datasets offer unprecedented opportunities for enhancing detection accuracy, they also represent a significant escalation in the granularity of personal data being analyzed, necessitating a reevaluation of privacy boundaries and informed consent protocols in research settings.

Central to these concerns is the issue of privacy and informed consent. The collection and analysis of video data, particularly in real-world settings, raises significant privacy concerns that demand careful consideration. Balmer's comprehensive examination of lie detection technologies argues that subjecting individuals to such scrutiny, regardless of the purported benefits, can be perceived as an invasion of privacy [60]. This is particularly pertinent in video-based deception detection, where subtle facial expressions, eye movements, and other nonverbal cues are meticulously analyzed. Consequently, the principle of informed consent becomes paramount, requiring participants in deception detection studies to be fully informed about data collection, usage, and storage methods. However, the nature of deception research often presents a paradox: researchers must balance the need for truly informed consent against the risk of compromising study validity by revealing too much about the deceptive elements.

This ethical dilemma is further compounded by the increasing likelihood of these technologies being deployed in covert surveillance scenarios. Suchotzki and Gamer highlight that the development of AI-based deception detection methods could lead to their application in contexts where individuals are unaware they are being analyzed, such as security screenings or job interviews [61]. This surreptitious use not only infringes on individual privacy rights but also risks eroding trust in social interactions and institutions. While proponents might argue for the potential benefits in national security or criminal investigations, the ethical costs of widespread covert surveillance far outweigh these purported advantages.

The psychological impact on participants in deception detection studies represents another critical ethical consideration. Many experimental designs involve placing participants in stressful or ethically ambiguous situations, potentially leading to lasting psychological consequences. Researchers must carefully weigh the induction of stress or anxiety, often necessary to elicit genuine deceptive behaviors, against potential harm to participants. Mambreyan et al. emphasize that the very act of being subjected to deception detection procedures can alter an individual's behavior and self-perception, potentially resulting in unintended psychological effects [62]. This impact extends beyond the immediate research context, as evidenced by studies utilizing datasets like Bag-of-Lies and TRuLie, which demonstrate that even low-stakes scenarios can induce significant stress in participants [18], [37]. The knowledge that one's microexpressions, eye movements, and physiological responses are under scrutiny can alter natural behavior and potentially cause lasting psychological effects. This raises profound ethical questions about the long-term impact of widespread deception detection technology on social interactions and individual well-being, urging researchers to consider not only the immediate effects of their studies but also the broader

TABLE 6. Key characteristics and performance metrics of deception detection studies (2019-2024).

Author (Year)	Dataset	Sample Size (N)	Class Distribution (T/D)	Model Type	Modalities	Acc	Precision	Recall	F1-score	AUC-ROC	Statistical Measures
Mathur & Matarić (2021) [53]	Real-world courtroom trial videos	108	53/55	Unsupervised DBN (512-256-2 architecture) + GMM clustering	Facial valence + Visual	70%	88%	-	-	80%	p < 0.01 (vs PCA baseline)
Younessi Heravi et al. (2023)[32]	Mock crime CQT experiment	25 (dari 40 awal)	300/75 questions	RQA + LDA classifier	Video-based eye movement (horizontal & vertical) + physiological signals	81.86%	-	-	-	-	p < 0.05
Shilaskar et al. (2023)[54]	Miami University Deception Detection Database	320 videos	216/104 (Truth/Lie)	Custom Model (Combined Duchenne Smile + Blink Rate)	Video-based facial micro-expressions	75.10%	77.78%	94.10%	85.14%	-	p < 0.05
Hsiao & Sun (2022)[49]	Real-life trial video data	121 video clips	60/61 (Truth/Deceptive)	Multi-modal Ensemble with Attention RNN	Visual (facial landmarks) + Audio (MFCC) + Transcription	96%	-	-	95.76%	-	10-fold cross-validation
Guo et al. (2023)[33]	DOLOS (gameshow)	1,675 video clips	776/899 (Truth/Deceptive)	PECL (Parameter-Efficient Crossmodal Learning) with Multi-task	Visual + Audio + MUMIN features	66.84%	-	-	73.35%	64.58%	Cohen's Kappa: 0.65
Crockett et al. (2020) [24]	Silent Talker Video Dataset	86,584 image vectors (32 participants: 22 male, 10 female)	Male: 34,618/25,581 Female: 8,432/17,653 (Truth/Deceptive)	Multiple ML models (J48, Random Forest, MLP, Naive Bayes)	Visual (36 non-verbal channels)	99.8% (Random Forest)	-	-	-	-	Spearman rank correlation (ρ)=0.75, p<0.01
Biçer & Dibeklioglu (2024)[55]	Real-Life Trial (RLT)	114 videos (50 subjects)	57/57 (Truth/Deceptive)	Convolutional Self-Attention (CSA) + LSTM	Visual + Vocal + Speech	96.50%	-	-	-	-	Leave-one-subject-out cross-validation
Bai et al. (2019)[22]	Resistance Game Videos	285 players (44 games)	172/113 (Truth/Deceptive)	Ensemble (LR+RF+NB+L-SVM+NB)	Visual + Audio + Behavioral	-	66.60%	37.90%	0.466	0.705	10-fold cross-validation
Gupta et al. (2019)[18]	Bag-of-Lies	325 recordings (35 subjects)	163/162 (Truth/Deceptive)	Multimodal Score Level Fusion	Video + Audio + EEG + Gaze	66.17%	-	-	-	-	Cross-validation (2-fold for Set A, 3-fold for Set B)
Avola et al. (2019) [69]	Real-life trial dataset	121 videos (56 subjects: 21 female, 35 male)	60/61 (Truth/Deceptive)	SVM with RBF kernel	Visual (18 facial AUs)	76.84%	-	-	-	-	p-value significance testing for AUs
Bai et al. (2019)[60]	Resistance Game Videos	285 players (44 games)	172/113 (Truth/Deceptive)	Ensemble (LR+RF+NB+L-SVM+NB)	Visual + Audio + Behavioral	-	66.60%	37.90%	0.466	0.705	10-fold cross-validation
Stathopoulos et al. (2021) [56]	Real-Life Trial	104 videos	50:54 (Truth/Deceptive)	TCN with FAU & Gaze	Visual (FAU + Gaze)	92.36%	-	-	-	97.27%	10-fold cross-validation
Khan et al. (2024) [48]	Real-Life Trial	121 videos	60:61 (Truth/Deceptive)	LSTM	Visual (32 AUs)	90.90%	-	-	-	90.74%	10-fold cross-validation using IDs
Karpova et al. (2020)[38]	TRuLie (self-collected)	93 participants (34 hours)	6444:3380	LightGBM	Audio + Video + PPG + Eye-tracking + Emotions	67.5	77.8	73.9	0.757	-	5-fold cross-validation
Li et al. (2024) [47]	Self-collected simulated theft dataset	96 participants	50:46:00	SVM + SOS feature selection	Visual (AU + Gaze + Head pose) + HR	83.27	-	-	-	83.33	10-fold cross-validation
Li et al. (2024) [52]	PV3D (self-collected)	96 participants	50:46:00	SVM + SOS feature selection	Visual (AU + Gaze + Head pose) + HR	83.27	-	-	-	83.33	10-fold cross-validation

TABLE 6. ***(Continued.)*** **Key characteristics and performance metrics of deception detection studies (2019-2024).**

Author (Year)	Dataset	Sample Size (N)	Class Distribution (T/D)	Model Type	Modalities	Acc	Precision	Recall	F1-score	AUC-ROC	Statistical Measures
Sehrawat et al. (2023) [50]	Miami University Deception (MU3D)	320 videos	160:160 (Truth/Deceptive)	Stacked Bi-LSTM + ResNet50	Video + Text	98.10%	-	-	-	-	85:15 split
Chebbi & Jebara (2023) [61]	Real-Life Court Trial	196 videos	47:53 (Truth/Deceptive)	KNN	Audio + Video + Text	97% (Feature-level)	-	100% (Decision-level)	-	-	70:30 split
Khan et al. (2021) [25]	Holiday Role-Playing	262,026 vectors from 100 participants	128,735:126,291 (Truth/Deceptive)	Random Forest	Visual (facial + eye movements)	78%	-	72%	80%	-	80:20 split, p-value=0.74
Monaro et al. (2022) [30]	Lab-collected holiday interviews	62 videos (9.56 min avg)	32:30 (Truth/Deceptive)	SVM + OpenFace AUs	Visual	-	-	-	-	0.72 (FreeSpeech) 0.78 (Questions)	10-fold cross validation
Javaid et al. (2022) [57]	Bag-of-Lies	325 (Audio/Video) 201 (EEG)	Audio/Video: 163/162 EEG: 108/93	Multimodal (Bi-LSTM + Attention CNN + Two-stream CNN)	EEG + Audio + Video	83.5	86	82	0.83	-	5-fold cross validation
Yang et al. (2022) [44]	Real-life trial dataset	188 clips (after preprocessing)	94:94 (Truth/Deceptive)	SVM + ETF + Facial Displays + Hand Gestures	Visual	87.59	-	-	-	-	10-fold cross validation, 100 trials
Dinges et al. (2024) [36]	Real-life trials (RL)	118	56:62	SVM + Multiple CNNs	Visual (facial cues)	77.7	-	-	-	0.82	Cross-validation, paired t-test
Ding et al. (2019) [51]	Real-life court trial dataset	104 videos (58 identities)	50:54:00	FFCSN + Acoustic + Verbal	Multimodal	97	-	-	-	99.78	10-fold cross validation
Nam et al. (2023) [34]	Real-life Trial Dataset	104 videos	50:54:00	FacialCueNet (CNN+ConvLSTM)	Visual only	88.45	89.17	90.62	-	0.9541	10-fold cross validation
Zhang et al. (2022) [31]	Court Trial Dataset	120 video clips (48 subjects)	81:39:00	GCFM (Graph-based Cross-modal Fusion Model)	Visual + Acoustic + Textual	88.14	-	-	78.5	-	10-fold cross validation
Rill-García et al. (2019) [26]	Real-life Trial Dataset	121 videos (58 subjects)	60:61	Hierarchical BSSD	Visual + Acoustic + Textual	-	-	-	-	0.671	10-fold cross-validation (identity-based)
Ngô et al. (2021) [27]	Real-Life Trial (RLT)	116 videos	60:61	2D-to-3D CNN + LSTM	Visual (Face)	68	66	72	-	-	Leave-One-Person-Out
Mathur & Matarić (2020) [45]	Real-Life Trial	108 videos (47 subjects)	53:55:00	AdaBoost + SVM	Visual + Vocal + Facial Affect	84	-	-	0.84	0.91	5-fold cross-validation (speaker-independent)
Karnati et al. (2022)[58]	Real-Life Trial	121 videos	60:61	LieNet + Score Level Fusion	Visual + Vocal	97.33	97	97	0.97	-	10-fold cross-validation
Avola et al. (2020) [72]	Real-life trial dataset	119 sequences from 77 videos (47 subjects)	37:40:00	Fisher-LSTM (K=7)	Hand Gestures	90.96	91.49	90.93	0.912	91.14	10-fold cross-validation across subjects (80/20 split)
Hsiao & Sun (2023)[35]	Real-life trial	121 clips	60:61	Attention-aware RNN + LoRA	Visual + Audio + Text	92	-	-	91.9	-	10-fold cross-validation (80/20 split)
Sen et al. (2022) [43]	Real-life trial dataset	59 subjects (121 clips)	35:24:00	Score-level fusion with RF & NN	Visual + Audio + Text	84.18	-	-	-	0.94	Leave-one-subject-out cross-validation, 3 repetitions
Kamboj et al. (2020)[28]	Political statements dataset	180 videos (88 politicians)	87:93 (Democrat:Republican)	Decision Tree	Visual + Audio + Text	69	-	-	-	-	Leave-one-subject-out cross-validation

TABLE 6. ***(Continued.)*** **Key characteristics and performance metrics of deception detection studies (2019-2024).**

Author (Year)	Dataset	Sample Size (N)	Class Distribution (T/D)	Model Type	Modalities	Acc	Precision	Recall	F1-score	AUC-ROC	Statistical Measures
Islam et al. (2021) [46]	Bag of Lies	325 videos (35 subjects)	163:162	SVM-RBF	Visual (FAUs)	61.54	57.58	69.72	62.85	-	10-fold cross-validation
Venkatesh et al. (2019) [59]	Real-life trial dataset	121 videos	60:61	Multimodal fusion (Majority voting)	Audio + Text + Visual	97	-	-	-	-	25-fold cross-validation

societal implications of normalizing such intense scrutiny of human behavior.

The potential for bias and discrimination in deception detection systems is a particularly pressing ethical issue, further amplified by the advent of multimodal approaches. Mambreyan et al.'s study on dataset bias underscores the critical need for diverse and representative datasets, revealing that many existing datasets used in deception detection research are biased, particularly with respect to gender [62]. This bias in datasets can lead to the development of algorithms that perform differently across demographic groups, potentially perpetuating or exacerbating existing societal inequalities. The ethical implications of deploying such biased systems in real-world scenarios, especially in high-stakes contexts like law enforcement or border control, are profound and far-reaching. Studies utilizing datasets like DDPM and ATSFace have demonstrated that different modalities (e.g., facial expressions, voice, eye movements) may exhibit varying levels of accuracy across different demographic groups [29], [35]. This multimodal nature of bias introduces new challenges in ensuring fairness and equity in deception detection systems, necessitating a more nuanced approach to bias mitigation that considers the interplay between different modalities and their combined impact on various demographic groups. Furthermore, the cultural specificity of certain datasets, such as ATSFace which focuses on Chinese-speaking participants, highlights the need for cross-cultural validation to prevent the development of culturally biased systems.

To address these ethical concerns, researchers and practitioners should implement rigorous bias testing protocols, actively seek diverse participant pools, and develop algorithms that are explicitly designed to mitigate demographic biases. Additionally, the development of ethical guidelines specific to multimodal deception detection research is crucial. These guidelines should mandate transparency in dataset composition, require regular audits for bias, and establish clear protocols for obtaining informed consent in various research and application contexts.

The question of accuracy and reliability in video-based deception detection technologies also carries significant ethical weight. Suchotzki and Gamer argue that even if these technologies achieve high accuracy rates in controlled settings, their application in real-world scenarios introduces numerous variables that could compromise their reliability [61]. The potential for false positives or negatives in deception detection carries serious consequences, particularly in legal or security contexts, underscoring the paramount ethical responsibility of researchers and practitioners to clearly communicate the limitations and potential errors of these technologies. To mitigate these risks, it is essential to develop robust error reporting mechanisms and to establish clear guidelines for the appropriate use and interpretation of deception detection results in various contexts.

Moreover, the societal implications of widespread adoption of video-based deception detection technologies warrant careful consideration. Balmer's historical analysis of lie detection technologies explores how the very existence of such tools can fundamentally alter social dynamics and trust relationships [60]. The potential for these technologies to be used as tools of social control or to exacerbate power imbalances in various contexts – from the workplace to the criminal justice system – raises profound ethical questions about their role in shaping society. To address these concerns, policymakers should consider implementing strict regulations on the use of deception detection technologies, particularly in high-stakes environments. These regulations should include provisions for public oversight, mandatory reporting of usage and outcomes, and clear limitations on the contexts in which such technologies can be deployed.

Finally, the rapid advancement of deep learning techniques in video-based deception detection, as evidenced by studies using complex architectures like Face-Focused Cross-Stream Networks and multimodal fusion approaches, introduces new ethical challenges related to interpretability and accountability [33], [49]. While these sophisticated models often achieve higher accuracy, their decision-making processes can be opaque, making it difficult to understand and justify their conclusions. This "black box" nature of advanced AI models is particularly problematic in high-stakes applications of deception detection, such as in legal or security contexts. The ethical imperative for transparency and explainability in AI systems becomes even more critical when these technologies have the potential to significantly impact individuals' lives. Researchers and practitioners must strive to develop methods that not only achieve high accuracy but also provide clear, interpretable results that can be scrutinized and validated by human experts. This could involve the development of hybrid models that combine the predictive power of deep learning with more interpretable machine learning techniques, or the creation of tools that can generate human-readable explanations for AI decisions.

In conclusion, the ethical considerations surrounding video-based deception detection are multifaceted and deeply intertwined with broader societal issues of privacy, consent, fairness, and human rights. As research in this field progresses, it is imperative that ethical considerations are not treated as an afterthought but are integrated into every stage of research design, implementation, and potential application. The development of robust ethical frameworks, transparent research practices, and ongoing dialogue between researchers, ethicists, policymakers, and the public is essential to ensure that the advancement of these technologies aligns with societal values and respects individual rights. Only through such a comprehensive and nuanced approach can the potential benefits of video-based deception detection be realized while mitigating its potential harms. The evolving landscape of this technology demands continuous ethical reassessment and adaptation to address emerging challenges and ensure responsible development and deployment.

VI. CHALLENGES AND FUTURE WORKS

This section addresses the key challenges and potential future directions in video-based deception detection, aligning with our research questions, particularly RQ4. We examine the obstacles that currently impede progress in this field and explore promising avenues for advancement using artificial intelligence and machine learning techniques.

The field of video-based deception detection, while offering considerable promise, is confronted with a number of substantial challenges that impede its progress towards the development of practical and ethical real-world applications. These challenges encompass limitations in the availability of suitable datasets, methodological issues, analytical complexities and ethical considerations. It is imperative that these challenges be addressed if the field is to advance and realise its full potential across a range of domains, including law enforcement, security and beyond.

A. DATASET LIMITATIONS AND QUALITY

One of the primary obstacles in advancing video-based deception detection is the scarcity of large-scale, diverse datasets that accurately represent the complexity of human deceptive behaviour. Our systematic review of 42 primary research papers revealed that 21 existing datasets are often limited in size and diversity, making it difficult to develop models that can be generalized across different contexts and cultures. This limitation is further compounded by the inherent class imbalance prevalent in deception datasets. For example, the DDCIT Dataset [34], contains 630 samples with a 2:1 ratio of truthful to deceptive instances, illustrating how truthful samples typically outnumber deceptive ones. This disparity can potentially lead to the development of biased models and skewed results.

The challenge of creating realistic yet ethically sound deception scenarios further compounds the dataset problem. Our analysis highlights the delicate balance researchers must strike between ecological validity and experimental control. While laboratory settings offer greater control, they may fail to capture the nuances of real-world deception. For instance, the Box of Lies (BoL) Dataset [20] uses a game show format to capture more naturalistic deceptive behaviours, but this approach must be balanced against the need for controlled experimental conditions. Conversely, datasets collected from uncontrolled environments, such as the Real-Life Trial Dataset [16], present their own set of challenges, particularly in verifying the ground truth of subjects' responses.

The temporal evolution of video-based deception detection datasets illustrates a growing research interest, with a pronounced acceleration in 2023 marked by the introduction of five novel datasets (DOLOS, DDCIT, ATSFace, Live-Action Program, and A New Approach for Lie Detection Using Eye Movement) [31], [32], [33], [34], [35]. However, our review found that only 52.4% of the identified datasets are publicly accessible, either directly or by request, underscoring the need for more open data sharing to foster reproducibility and collaborative research.

Future research should prioritize the development of large-scale, diverse datasets that span different cultures, contexts, and types of deceptive behaviour. This may involve innovative data collection methods, such as the gamified scenarios used in the Box of Lies Dataset [20] or carefully designed naturalistic experiments like those employed in the Real-Life Trial Dataset [16], to capture a wide range of deceptive behaviours while maintaining ethical standards. Additionally, efforts should be made to create more balanced datasets in terms of truthful and deceptive samples, addressing the class imbalance issue noted in datasets like DDCIT [34].

Furthermore, the trend towards multimodal datasets, incorporating visual, audio, and sometimes textual data, reflects the complex nature of deception and the need for comprehensive analysis. Studies like those by Gupta et al. [18] and Guo et al. [33] demonstrate the potential of integrating multiple modalities to enhance detection accuracy. Future dataset development should focus on this multimodal approach while also considering the ethical implications of data collection and use in this sensitive domain.

B. METHODOLOGICAL CHALLENGES AND ADVANCEMENTS

The integration of multiple modalities – visual, audio, and textual – presents a complex analytical challenge. Each modality carries its own set of features and patterns that must be effectively combined to create a holistic understanding of deceptive behaviour. Guo et al. [33] demonstrated that combining audio-visual cues through parameter-efficient cross modal learning can yield promising results, achieving an accuracy of 66.84% on the DOLOS dataset. However, the optimal approach for integrating diverse data types remains an open question.

Future work should focus on developing sophisticated fusion techniques that can adaptively weight different modalities based on their reliability and relevance in specific deception contexts. The exploration of advanced attention mechanisms and novel architectures that can capture subtle cross-modal interactions will be pivotal in enhancing the accuracy and robustness of deception detection models. For example, the attention-aware multimodal RNN proposed by Hsiao and Sun [47] achieved an accuracy of 96% by effectively combining visual, audio, and transcription features.

The temporal dynamics of deceptive behavior present another significant challenge, particularly in scenarios involving extended interactions. Bai et al. [22] addressed this issue in their study on automatic long-term deception detection in group interaction videos, achieving an accuracy of 70.5% using an ensemble of classifiers. Their work highlights the need for sophisticated temporal modelling techniques. Future research should explore hierarchical attention networks, transformer-based architectures, and other advanced methods capable of discerning long-term dependencies in video data. For instance, the Face-Focused Cross-Stream Network (FFCSN) proposed by [49] achieved an accuracy of 97% by effectively addressing temporal inconsistencies between facial expressions and body motions.

C. ETHICAL CONSIDERATIONS

While the technical challenges in video-based deception detection are significant, the ethical implications of these technologies are equally crucial. Researchers and practitioners must grapple with a complex web of issues including privacy, consent, psychological impact, bias, fairness, accountability, and transparency. The future trajectory of research in this field must prioritize the development of ethically sound methodologies and technologies that address these multifaceted concerns.

Moving forward, it is imperative that researchers focus on enhancing privacy-preserving techniques for data collection and analysis. This effort should be coupled with the development of more robust informed consent procedures that fully apprise participants of the potential risks and uses of their data, ensuring transparency and maintaining trust in the research process. For example, future studies could explore the use of federated learning techniques, which allow models to be trained on decentralized data without compromising individual privacy.

Simultaneously, the advancement of fairness-aware machine learning algorithms is crucial to mitigate biases in deception detection models, promoting equitable outcomes across diverse populations. This is particularly important given the findings of studies like [23], which revealed significant gender differences in non-verbal cues to deception, with gender-specific models outperforming general models in deception detection.

The creation of interpretable AI systems capable of providing clear explanations for their decisions is particularly vital, especially in high-stakes applications where the consequences of misclassification can be severe. This push towards explainable AI not only enhances the trustworthiness of these systems but also facilitates their responsible deployment in real-world scenarios. For instance, the FacialCueNet model proposed by Nam et al. [34] incorporates a spatial-temporal attention mechanism to improve interpretability, achieving an accuracy of 70.79% while providing insights into the most salient facial cues for deception detection.

D. MODEL INTERPRETABILITY AND ROBUSTNESS

As deep learning models continue to demonstrate impressive capabilities in deception detection, their "black box" nature poses significant challenges for adoption in high-stakes scenarios, particularly in legal and ethical contexts. The development of interpretable deep learning models that provide clear, understandable explanations for their detection decisions is paramount. This research direction not only fosters trust in these systems but also enables their responsible application in sensitive domains.

Future research in video-based deception detection should focus on developing hybrid approaches that synergize the robust predictive capabilities of deep learning with the interpretability of traditional machine learning methods. This integration is crucial for enhancing the transparency and trustworthiness of deception detection systems, particularly in high-stakes environments where the rationale behind a model's decision is as important as the decision itself. For instance, the attention mechanism approach proposed by Hsiao and Sun [47] could be extended to provide visual or textual explanations of the most salient features influencing the model's decision, thereby offering insights into the specific cues that contribute to the detection of deceptive behavior.

Recent advancements in explainable AI techniques have shown promising potential for enhancing the interpretability of video-based deception detection models, particularly through LIME (Local Interpretable Model-agnostic Explanations) and SHAP (SHapley Additive exPlanations). LIME operates by creating interpretable representations of complex models through local approximation, generating simplified explanations by sampling perturbed instances around the prediction of interest and fitting a simple model (typically linear regression) to this local region [63]. This technique enables researchers to understand which specific frames or behavioral sequences influence the model's decisions. SHAP, grounded in cooperative game theory, provides a more mathematically rigorous framework by computing Shapley values that fairly distribute the "contribution" of each feature to the prediction [64]. In deception detection contexts, SHAP values quantify how each feature (such as facial expressions, voice modulation, or body movements) contributes to the final prediction by calculating their marginal contributions across all possible feature combinations. For instance, in a study on intrusion detection using CNNs, SHAP values revealed the hierarchical importance of features, with some behavioral cues contributing more significantly to the detection of deceptive patterns [64]. The medical field has further demonstrated the effectiveness of these techniques, where SHAP and LIME have been used to explain complex diagnostic decisions by identifying key biomarkers and their relative importance [65]. This success in medical applications suggests that these techniques could be particularly valuable in deception detection, where understanding the relationship between behavioral cues and deceptive intent is crucial.

Recent advancements in explainable AI for multimodal analysis further underscore the potential of these techniques for video-based deception detection. A study by Talaat demonstrated the effectiveness of an Explainable Enhanced Recurrent Neural Network (ERNN) for lie detection using both audio and video signals [66]. This approach, which incorporates explainable AI capabilities, represents a significant step towards more transparent and interpretable models in the field of deception detection.

The incorporation of explainable AI techniques enables researchers to enhance the precision of deception detection models while simultaneously furnishing stakeholders with lucid and comprehensible insights into the underlying decision-making process. This transparency is of paramount importance for fostering trust in AI-driven deception detection systems and facilitating their ethical deployment in real-world scenarios. Moreover, the capacity to elucidate the rationale behind model decisions could pave the way for novel insights into the nuanced behavioural and physiological cues associated with deception, potentially advancing our theoretical understanding of human deceptive behaviour.

Addressing the challenge of adversarial robustness is critical. Researchers must investigate the limits of current detection methods and develop robust countermeasures against deliberate attempts at deception. This could involve exploring adversarial training techniques to enhance model resilience, as well as developing methods to detect and mitigate potential adversarial attacks on deception detection systems. For example, future studies could investigate the effectiveness of techniques like adversarial training or defensive distillation in improving the robustness of deception detection models against potential manipulation attempts

E. CROSS-CONTEXT GENERALIZATION AND REAL-WORLD APPLICABILITY

Improving the generalizability of deception detection models across different contexts presents another significant challenge. Models developed and tested in controlled, low-stakes experimental settings may not perform equally well in high-stakes real-world scenarios. This discrepancy is evident in the performance variations observed across different datasets, as highlighted in our analysis of classification methods.

For instance, the study by [67] comparing low-stakes and high-stakes deception video datasets revealed that networks trained on high-stakes lies (Real-Life Trial dataset) performed better than those trained on low-stakes lies (Box of Lies dataset), with accuracies of 68.64% and 55.92% respectively. This underscores the importance of considering the context and stakes of deceptive behavior when developing and evaluating detection models.

Future research should explore transfer learning techniques to adapt models trained in controlled settings to real-world scenarios. Domain adaptation methods could be employed to bridge the gap between different cultural and linguistic contexts, addressing the limitations observed in culturally specific datasets like ATSFace [35]. Meta-learning approaches should also be investigated to develop more flexible and adaptable deception detection models capable of quickly adjusting to new contexts or types of deceptive behavior.

The development of robust video-based deception detection datasets faces multifaceted challenges that significantly impact the field's advancement. These challenges encompass issues of limited size and diversity, class imbalance, cultural homogeneity, ethical considerations, ground truth verification, technical quality, and temporal dynamics. A critical limitation in current datasets is the lack of comprehensive personality profiles for participants, neglecting the potential influence of individual personality traits on deceptive behaviours. This oversight hinders the development of more nuanced and personalized deception detection models. Incorporating standardized personality assessments, such as the Big Five Inventory, into dataset collection protocols could provide valuable insights into how personality traits modulate deceptive behaviours across different contexts and cultures [68], [69]

To address these challenges and enable widespread adoption, future research should focus on developing larger, more diverse, and ethically sound datasets that incorporate personality assessments and span different cultures and contexts. This includes creating standardized protocols for data collection, labelling, and model evaluation, as well as developing innovative methods for eliciting and verifying deceptive behaviours ethically. Additionally, researchers should explore computationally efficient models capable of real-time deception detection on resource-constrained devices, such as the LieToMe approach proposed by Avola et al., which achieved 90.96% accuracy using only hand gestures [70]. These efforts should be complemented by investigations into model compression, quantization techniques, and online learning methods for continuous model improvement in real-world applications.

The path forward in video-based deception detection research is both challenging and exciting, requiring collaborative efforts between computer scientists, psychologists, ethicists, and domain experts. By systematically addressing the identified challenges and focusing on key research directions – dataset quality and diversity, advanced methodologies for multimodal and temporal analysis, ethical considerations and privacy preservation, model interpretability and robustness, and cross-context generalization – the field can move toward realizing its full potential. These advancements offer valuable tools for enhancing security, promoting fairness, and advancing our understanding of human behaviour, while upholding the highest standards of scientific rigor and ethical responsibility. As we navigate this complex landscape, the promise of more accurate, robust, and ethically sound deception detection systems comes into view, holding the potential to transform various sectors of society.

VII. DISCUSSION AND CONCLUSION

This systematic literature review provides a comprehensive and critical analysis of the current state of video-based deception detection, offering unique insights into the interplay between datasets, modalities, and computational methods in this rapidly evolving field. The literature search encompassed studies from 2019 to 2024, resulting in the inclusion of 42 papers from various journals and conferences, including valuable contributions from conference papers that drive innovation in this research area.

The analysis of **RQ1** reveals significant trends in dataset utilization for deception detection. A notable finding is that only 52.4% of the identified datasets are publicly accessible, highlighting the pressing need for more open data sharing to foster reproducibility and collaborative research. The review identifies a shift towards more diverse and naturalistic data collection methods, as exemplified by datasets like the Real-Life Trial Dataset [16] and the Box of Lies Dataset [42]. These datasets offer valuable insights into authentic deceptive behaviours in high-stakes and naturalistic scenarios, respectively. The duration of recordings emerges as a crucial factor, with datasets ranging from short clips suitable for micro-expression analysis, such as DOLOS [33] with an average of 3.99 seconds, to extended sessions allowing for the observation of complex behavioural patterns, as seen in the Automatic Long-Term Deception Detection dataset [22] with durations of 30-65 minutes. This diversity reflects the multifaceted nature of deceptive behaviour and suggests that an optimal approach may involve combining both short and long-duration recordings to capture a comprehensive range of deceptive cues.

The quality of video recordings plays a pivotal role in the effectiveness of deception detection models. While some studies have employed professional-grade cameras, such as the FLIR SC6700 used in gender-based deception research [23], many high-quality datasets have been successfully created using standard webcams or smartphone cameras. For instance, the DDPM dataset [29] achieved satisfactory results using a Logitech C920 webcam. This suggests that while high-end equipment can provide additional detail, consumer-grade cameras capable of 720p resolution at 30 fps are often sufficient for capturing the nuances of deceptive behavior.

An often overlooked but crucial aspect in dataset development is the consideration of participants' personalities. The Miami University Deception Detection Database (MU3D) [19] stands out in this regard, providing a detailed codebook that includes personality assessments of participants. Similarly, datasets like Bag-of-Lies [18] and DDPM [29] have incorporated demographic factors and allowed participants flexibility in choosing to lie or tell the truth, potentially reflecting personality aspects in deceptive behavior. This approach aligns with psychological research suggesting that personality traits can significantly influence deceptive behavior [71].

The analysis of these datasets also highlights the importance of ecological validity in experimental design. Datasets that employ realistic scenarios, such as the Box of Lies [20] which uses a game show format, or the Real-Life Trial Dataset [16] which utilizes actual courtroom footage, offer valuable insights into natural deceptive behaviors. However, these approaches must be balanced against ethical considerations and the need for controlled experimental conditions.

Furthermore, the trend towards multimodal datasets, incorporating visual, audio, and sometimes textual data, reflects the complex nature of deception and the need for comprehensive analysis. Studies like those by Gupta et al. [18] and Guo et al. [33] demonstrate the potential of integrating multiple modalities to enhance detection accuracy.

As the field advances, future dataset development should focus on addressing current limitations. Efforts should also be made to create more balanced datasets in terms of truthful and deceptive samples, as class imbalance remains a persistent issue [34]. Additionally, cross-cultural datasets are needed to develop models that can account for cultural variations in deceptive behavior, an aspect currently underrepresented in existing datasets [35].

The practical applications of video-based deception detection systems extend across multiple domains, demonstrating significant potential for real-world implementation. In law enforcement, these systems can enhance interrogation processes by providing objective analysis of behavioral cues, as demonstrated by studies using the Real-Life Trial Dataset [16] and DDCIT Dataset [34], which achieved accuracies of 75.20% and 70.79% respectively in analyzing courtroom and criminal interrogation scenarios. In security applications, particularly at border control and high-security facilities, multimodal systems incorporating facial, vocal, and physiological indicators have shown promise, with studies reporting accuracy improvements of 10-15% compared to traditional methods [29]. The psychological domain has also benefited from these advances, with datasets like the Gender-Based Deception Dataset [23] providing insights into gender-specific deceptive behaviors, achieving 68.90% accuracy in identifying psychological markers of deception. These real-world applications underscore the practical value of continued research in this field, while simultaneously highlighting the need for careful consideration of ethical implications and cultural variations in deceptive behavior [60], [61].

In conclusion, the development of robust, diverse, and ethically sound datasets remains a critical challenge in advancing video-based deception detection. Future research should prioritize the creation of large-scale, multimodal datasets that balance ecological validity with experimental control, while also considering the ethical implications of data collection and use in this sensitive domain.

RQ2 underscores the importance of multimodal approaches in deception detection, integrating visual, audio, and textual cues. The review highlights the potential of combining these diverse data streams to improve detection accuracy, with recent studies exploring sophisticated fusion techniques and attention mechanisms to capture subtle cross-modal interactions. Notably, the analysis of micro-expressions and other subtle non-verbal cues emerges as a promising avenue for enhancing detection capabilities, although challenges remain in capturing and interpreting these fleeting behavioural indicators.

The evaluation of computational methods in **RQ3** reveals a significant shift from traditional machine learning techniques to advanced deep learning architectures. The review identifies Long Short-Term Memory (LSTM) networks and their variants as particularly effective in capturing the temporal aspects of deceptive behaviour. Recent innovations, such as the integration of attention mechanisms and the development of hybrid architectures combining convolutional and recurrent elements, demonstrate the field's rapid evolution towards more sophisticated and nuanced analysis techniques.

Frame-level analysis, exemplified by studies such as Nam m et al. [34] and Yang et al. [42], capitalizes on the potential of micro-expressions and subtle facial cues as indicators of deception. This approach aligns with psychological theories suggesting that deception may leak through brief, involuntary facial expressions [1]. However, the frame-by-frame method poses challenges in terms of computational intensity and the risk of overlooking context-dependent cues.

Conversely, clip-level analysis, as employed by Mathur and Matarić [51] and Ding et al. [49], takes a more holistic view, potentially capturing longer-term behavioral patterns and contextual information. This approach may be more robust against momentary fluctuations but risks missing fleeting yet significant cues.

Interestingly, a trend towards hybrid approaches is emerging, as seen in the work of Khan et al. [46] and Stathopoulos et al. [54]. These methods attempt to bridge the gap between micro and macro analysis, potentially offering a more comprehensive understanding of deceptive behavior. This evolution reflects a growing recognition of the temporal complexity of deception, which may unfold differently across various time scales.

The diversity in approaches also highlights a critical challenge in the field: the lack of standardization in temporal analysis methods. This variability makes cross-study comparisons difficult and may contribute to the inconsistent performance of deception detection models across different datasets, as noted by Mambreyan et al. [62].

Furthermore, the choice between frame-level and clip-level analysis often depends on the specific modalities being examined. For instance, facial micro-expressions naturally lend themselves to frame-level analysis, while vocal patterns or body language may be more appropriately analyzed at the clip level. This suggests that optimal deception detection may require a multimodal approach that intelligently combines different temporal resolutions of analysis.

As the field progresses, there is a clear need for research that directly compares the efficacy of frame-level, clip-level, and hybrid approaches across diverse datasets and deception scenarios. Such comparative studies could provide valuable insights into the temporal dynamics of deceptive behavior and guide the development of more robust and generalizable detection models.

RQ4 addresses the key ethical considerations in video-based deception detection research and application. The review emphasizes the critical importance of privacy protection and informed consent in data collection and analysis. As datasets become increasingly comprehensive and invasive, such as the DOLOS [33] and DDPM [29] datasets which incorporate a wide array of physiological and behavioural cues, the ethical implications of data collection and use become more pronounced. The potential for bias and discrimination in deception detection systems is a pressing concern, as highlighted by Mambreyan et al. [62] in their study on dataset bias. This underscores the need for diverse and representative datasets to ensure fairness and equity in the development and application of deception detection technologies.

Furthermore, the psychological impact on participants in deception detection studies represents another critical ethical consideration. Many experimental designs involve placing participants in stressful or ethically ambiguous situations, potentially leading to lasting psychological consequences. Researchers must carefully weigh the induction of stress or anxiety, often necessary to elicit genuine deceptive behaviours, against potential harm to participants. This ethical dilemma is further compounded by the increasing likelihood of these technologies being deployed in covert surveillance scenarios, raising profound questions about individual privacy rights and the potential erosion of trust in social interactions and institutions [60], [61].

RQ5 outlines the current challenges and future directions in video-based deception detection, revealing a complex landscape of technical, methodological, and ethical considerations. A primary challenge is the scarcity of large-scale, diverse datasets that accurately represent the complexity of human deceptive behavior across different cultures and contexts. Our review found that only 52.4% of identified datasets are publicly accessible, highlighting the urgent need for more open data sharing to foster reproducibility and collaborative research [31], [32], [33], [34], [35].

Future research should prioritize the development of robust, diverse, and ethically sound datasets that balance ecological validity with experimental control. This may involve innovative data collection methods, such as gamified scenarios [20] or carefully designed naturalistic experiments [16], while addressing issues of class imbalance and cultural homogeneity. Incorporating standardized personality assessments, like the Big Five Inventory, into dataset collection protocols could provide valuable insights into how personality traits modulate deceptive behaviors [68], [71].

The integration of multiple modalities – visual, audio, and textual – presents complex analytical challenges. Future work should focus on developing sophisticated fusion techniques that can adaptively weight different modalities based on their reliability and relevance in specific deception contexts. The exploration of advanced attention mechanisms and novel architectures, such as the attention-aware multimodal RNN proposed by Hsiao and Sun [47], will be pivotal in enhancing the accuracy and robustness of deception detection models.

Addressing the temporal dynamics of deceptive behavior, particularly in extended interactions, remains a significant challenge. Future research should explore hierarchical attention networks, transformer-based architectures, and other advanced methods capable of discerning long-term dependencies in video data, building on work like the Face-Focused Cross-Stream Network (FFCSN) [49].

The development of interpretable AI systems is crucial, especially for high-stakes applications. Adapting advanced explainable AI techniques such as LIME and SHAP to video-based deception detection could enhance transparency and trustworthiness [63], [64], [65]. Recent advancements, like the Explainable Enhanced Recurrent Neural Network (ERNN) for multimodal lie detection [66], represent significant steps towards more transparent and interpretable models.

Improving cross-context generalization is another key challenge. Future research should explore transfer learning techniques to adapt models trained in controlled settings to real-world scenarios. Domain adaptation methods could bridge the gap between different cultural and linguistic contexts, addressing limitations observed in culturally specific datasets [35]. Meta-learning approaches should be investigated to develop more flexible and adaptable deception detection models.

Ethical considerations remain paramount. Future research must prioritize privacy-preserving techniques for data collection and analysis, develop robust informed consent procedures, and advance fairness-aware machine learning algorithms to mitigate biases in deception detection models [23]. The potential psychological impact on participants and the broader societal implications of widespread deception detection technologies must be carefully considered [60], [61].

In conclusion, addressing these multifaceted challenges requires collaborative efforts between computer scientists, psychologists, ethicists, and domain experts. By focusing on key research directions – dataset quality and diversity, advanced methodologies for multimodal and temporal analysis, ethical considerations, model interpretability, and cross-context generalization – the field can move toward realizing its full potential. These advancements offer valuable tools for enhancing security, promoting fairness, and advancing our understanding of human behavior, while upholding the highest standards of scientific rigor and ethical responsibility.

REFERENCES

[1] S. Leal, A. Vrij, C. Hudson, P. Capuozzo, and H. Deeb, "The effectiveness of different model statement variants for eliciting information and cues to deceit," *Legal Criminological Psychol.*, vol. 27, no. 2, pp. 247–264, Sep. 2022, doi: 10.1111/lcrp.12200.

[2] L. Ten Brinke, J. J. Lee, and D. R. Carney, "Different physiological reactions when observing lies versus truths: Initial evidence and an intervention to enhance accuracy," *J. Personality Social Psychol.*, vol. 117, no. 3, pp. 560–578, Sep. 2019, doi: 10.1037/pspi0000175.

[3] A. Vrij, H. Deeb, S. Leal, P.-A. Granhag, and R. P. Fisher, "Plausibility: A verbal cue to veracity worth examining?" *Eur. J. Psychol. Appl. Legal Context*, vol. 13, no. 2, pp. 47–53, Jul. 2021, doi: 10.5093/ejpalc2021a4.

[4] Y. Y. Kao, P. H. Chen, C. C. Tzeng, Z. Y. Chen, B. Shmueli, and L. W. Ku, "Detecting deceptive language in crime interrogation," in *Proc. Int. Conf. Hum.-Comput. Interact.*, 2020, pp. 80–90, doi: 10.1007/978-3-030-50341-3_7.

[5] T. R. Levine, "Truth-default theory (TDT)," *J. Lang. Social Psychol.*, vol. 33, no. 4, pp. 378–392, Sep. 2014, doi: 10.1177/0261927x14535916.

[6] C. T. A. Gillen, A. J. Card, J. W. Smith, G. Ambroziak, J. C. Mundt, and C. Servais, "The influence of polygraphs on evaluators' decisions regarding sexually violent persons," *Law Hum. Behav.*, vol. 47, no. 3, pp. 448–461, Jun. 2023, doi: 10.1037/lhb0000530.

[7] D. T. Wilcox, "Application of the clinical polygraph examination to the assessment, treatment and monitoring of sex offenders," *J. Sexual Aggression*, vol. 5, no. 2, pp. 134–152, Jan. 2000, doi: 10.1080/13552600008413304.

[8] J. A. Lewis and M. Cuppari, "The polygraph: The truth lies within," *J. Psychiatry Law*, vol. 37, no. 1, pp. 85–92, 2009.

[9] M. Burzo and M. Abouelenien, "Multimodal deception detection," in *The Handbook of Multimodal-Multisensor Interfaces: Foundations, User Modeling, and Common Modality Combinations*, vol. 2. New York, NY, USA: Association for Computing Machinery and Morgan & Claypool, 2018, pp. 419–453, doi: 10.1145/3107990.3108005.

[10] A. D'Ulizia, A. D'Andrea, P. Grifoni, and F. Ferri, "Detecting deceptive behaviours through facial cues from videos: A systematic review," *Appl. Sci.*, vol. 13, no. 16, p. 9188, 2023, doi: 10.3390/app13169188.

[11] S. A. Prome, N. A. Ragavan, M. R. Islam, D. Asirvatham, and A. J. Jegathesan, "Deception detection using machine learning (ML) and deep learning (DL) techniques: A systematic review," *Natural Lang. Process. J.*, vol. 6, Mar. 2024, Art. no. 100057, doi: 10.1016/j.nlp.2024.100057.

[12] A. D'Ulizia, A. D'Andrea, P. Grifoni, and F. Ferri, "Analysis, evaluation, and future directions on multimodal deception detection," *Technologies*, vol. 12, no. 5, p. 71, May 2024, doi: 10.3390/technologies12050071.

[13] F. Tomas, O. Dodier, and S. Demarchi, "Computational measures of deceptive language: Prospects and issues," *Frontiers Commun.*, vol. 7, Feb. 2022, Art. no. 792378, doi: 10.3389/fcomm.2022.792378.

[14] 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.

[15] M. J. Page et al., "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," *BMJ*, vol. 372, pp. 1–14, Mar. 2021, doi: 10.1136/bmj.n71.

[16] V. Pérez-Rosas, M. Abouelenien, R. Mihalcea, and M. Burzo, "Deception detection using real-life trial data," in *Proc. ACM Int. Conf. Multimodal Interact.*, Nov. 2015, pp. 59–66, doi: 10.1145/2818346.2820758.

[17] M. Ouzzani, H. Hammady, Z. Fedorowicz, and A. Elmagarmid, "Rayyan—A web and mobile app for systematic reviews," *Systematic Rev.*, vol. 5, no. 1, pp. 1–10, Dec. 2016, doi: 10.1186/s13643-016-0384-4.

[18] V. Gupta, M. Agarwal, M. Arora, T. Chakraborty, R. Singh, and M. Vatsa, "Bag-of-lies: A multimodal dataset for deception detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 83–90, doi: 10.1109/CVPRW.2019.00016.

[19] E. P. Lloyd, J. C. Deska, K. Hugenberg, A. R. McConnell, B. T. Humphrey, and J. W. Kunstman, "Miami university deception detection database," *Behav. Res. Methods*, vol. 51, no. 1, pp. 429–439, Feb. 2019, doi: 10.3758/s13428-018-1061-4.

[20] F. Soldner, V. Pérez-Rosas, and R. Mihalcea. *Box of Lies R: Multimodal Deception Detection in Dialogues*. Association for Computational Linguistics. Accessed: Jun. 2019. [Online]. Available: https://www.youtube.com/

[21] K. Radlak, M. Bozek, and B. Smolka, "Silesian deception database: Presentation and analysis," in *Proc. ACM Workshop Multimodal Deception Detection*, Nov. 2015, pp. 29–35, doi: 10.1145/2823465.2823469.

[22] C. Bai, M. Bolonkin, J. Burgoon, C. Chen, N. Dunbar, B. Singh, V. S. Subrahmanian, and Z. Wu, "Automatic long-term deception detection in group interaction videos," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2019, pp. 1600–1605, doi: 10.1109/ICME.2019.00276.

[23] M. Abouelenien, M. Burzo, V. Pérez-Rosas, R. Mihalcea, H. Sun, and B. Zhao, "Gender differences in multimodal contact-free deception detection," *IEEE MultimediaMag.*, vol. 26, no. 3, pp. 19–30, Jul. 2019, doi: 10.1109/MMUL.2018.2883128.

[24] K. Crockett, J. O'Shea, and W. Khan, "Automated deception detection of males and females from non-verbal facial micro-gestures," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2020, pp. 1–7, doi: 10.1109/IJCNN48605.2020.9207684.

[25] 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," *Exp. Syst. Appl.*, vol. 169, May 2021, Art. no. 114341, doi: 10.1016/j.eswa.2020.114341.

[26] R. Rill-García, H. J. Escalante, L. Villasenor-Pineda, and V. Reyes-Meza, "High-level features for multimodal deception detection in videos," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2019, pp. 2424–2433.

[27] L. M. Ngô, W. Wang, B. Mandira, S. Karaoglu, H. Bouma, H. Dibeklioglu, and T. Gevers, "Identity unbiased deception detection by 2D-to-3D face reconstruction," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021, pp. 145–154, doi: 10.1109/WACV48630.2021.00019.

[28] M. Kamboj, C. Hessler, P. Asnani, K. Riani, and M. Abouelenien, "Multimodal political deception detection," *IEEE MultimediaMag.*, vol. 28, no. 1, pp. 94–102, Jan. 2021, doi: 10.1109/MMUL.2020.3048044.

[29] N. Vance, J. Speth, S. Khan, A. Czajka, K. W. Bowyer, D. Wright, and P. Flynn, "Deception detection and remote physiological monitoring: A dataset and baseline experimental results," *IEEE Trans. Biometrics, Behav., Identity Sci.*, vol. 4, no. 4, pp. 522–532, Oct. 2022, doi: 10.1109/TBIOM.2022.3218956.

[30] M. Monaro, S. Maldera, C. Scarpazza, G. Sartori, and N. Navarin, "Detecting deception through facial expressions in a dataset of videotaped interviews: A comparison between human judges and machine learning models," *Comput. Hum. Behav.*, vol. 127, Feb. 2022, Art. no. 107063, doi: 10.1016/j.chb.2021.107063.

[31] H. Zhang, Y. Ding, L. Cao, X. Wang, and L. Feng, "Fine-grained question-level deception detection via graph-based learning and cross-modal fusion," *IEEE Trans. Inf. Forensics Security*, vol. 17, pp. 2452–2467, 2022, doi: 10.1109/TIFS.2022.3186799.

[32] M. A. Younessi Heravi, M. Pishghadam, E. Khoshdel, and S. Zibaei, "A new approach for lie detection using non-linear and dynamic analysis of video-based eye movement," *Frontiers Biomed. Technol.*, vol. 10, no. 1, pp. 88–95, Dec. 2023, doi: 10.18502/fbt.v10i1.11516.

[33] X. Guo, N. M. Selvaraj, Z. Yu, A. W.-K. Kong, B. Shen, and A. Kot, "Audio-visual deception detection: DOLOS dataset and parameter-efficient crossmodal learning," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 22078–22088, doi: 10.1109/iccv51070.2023.02023.

[34] B. Nam, J. Y. Kim, B. Bark, Y. Kim, J. Kim, S. W. So, H. Y. Choi, and I. Y. Kim, "FacialCueNet: Unmasking deception—An interpretable model for criminal interrogation using facial expressions," *Int. J. Speech Technol.*, vol. 53, no. 22, pp. 27413–27427, Nov. 2023, doi: 10.1007/s10489-023-04968-9.

[35] S.-W. Hsiao and C.-Y. Sun, "LoRA-like calibration for multimodal deception detection using ATSFace data," in *Proc. IEEE Int. Conf. Big Data*, Dec. 2023, pp. 2163–2172, doi: 10.1109/bigdata59044.2023.10386630.

[36] L. Dinges, M.-A. Fiedler, A. Al-Hamadi, T. Hempel, A. Abdelrahman, J. Weimann, D. Bershadskyy, and J. Steiner, "Exploring facial cues: Automated deception detection using artificial intelligence," *Neural Comput. Appl.*, vol. 36, no. 24, pp. 14857–14883, Aug. 2024, doi: 10.1007/s00521-024-09811-x.

[37] V. Karpova, P. Popenova, N. Glebko, V. Lyashenko, and O. Perepelkina, "'Was it you who stole 500 rubles—The multimodal deception detection," in *Proc. Companion Publication Int. Conf. Multimodal Interact.*, Oct. 2020, pp. 112–119, doi: 10.1145/3395035.3425638.

[38] T. Sen, M. K. Hasan, Z. Teicher, and M. E. Hoque, "Automated dyadic data recorder (ADDR) framework and analysis of facial cues in deceptive communication," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 1, no. 4, pp. 1–22, Jan. 2018, doi: 10.1145/3161178.

[39] A. S. Constâncio, D. F. Tsunoda, H. D. F. N. Silva, J. M. D. Silveira, and D. R. Carvalho, "Deception detection with machine learning: A systematic review and statistical analysis," *PLoS ONE*, vol. 18, no. 2, Feb. 2023, Art. no. e0281323, doi: 10.1371/journal.pone.0281323.

[40] V. Pérez-Rosas and R. Mihalcea. *Cross-Cultural Deception Detection*. Accessed: 2015. [Online]. Available: http://www.liwc.net/descriptiontable1.php

[41] M. U. Sen, 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.

[42] J. T. Yang, G. M. Liu, and S. C. H. Huang, "Emotion transformation feature: Novel feature for deception detection in videos," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2020, pp. 1726–1730.

[43] L. Mathur and M. J. Matarić, "Introducing representations of facial affect in automated multimodal deception detection," in *Proc. Int. Conf. Multimodal Interact.*, Oct. 2020, pp. 305–314, doi: 10.1145/3382507.3418864.

[44] S. Islam, P. Saha, T. Chowdhury, A. Sorowar, and R. Rab, "Non-invasive deception detection in videos using machine learning techniques," in *Proc. 5th Int. Conf. Electr. Eng. Inf. Commun. Technol. (ICEEICT)*, Nov. 2021, pp. 1–6, doi: 10.1109/ICEEICT53905.2021.9667928.

[45] Y. Li, J. Bian, and R. Song, "Video-based deception detection using wrapper-based feature selection," in *Proc. IEEE Int. Conf. Comput. Intell. Virtual Environments Meas. Syst. Appl.*, Jul. 2024, pp. 1–5, doi: 10.1109/civemsa58715.2024.10586610.

[46] H. U. D. A. Khan et al., "Deception detection in videos using the facial action coding system," *Multimedia Tools Appl.*, Apr. 2024, doi: 10.1007/s11042-024-19153-4.

[47] S.-W. Hsiao and C.-Y. Sun, "Attention-aware multi-modal RNN for deception detection," in *Proc. IEEE Int. Conf. Big Data*, Dec. 2022, pp. 3593–3596, doi: 10.1109/BigData55660.2022.10020331.

[48] P. K. Sehrawat, R. Kumar, N. Kumar, and D. K. Vishwakarma, "Deception detection using a multimodal stacked bi-LSTM model," in *Proc. Int. Conf. Innov. Data Commun. Technol. Appl. (ICIDCA)*, Mar. 2023, pp. 318–326, doi: 10.1109/ICIDCA56705.2023.10099779.

[49] M. Ding, A. Zhao, Z. Lu, T. Xiang, and J. R. Wen, "Face-focused cross-stream network for deception detection in videos," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 7794–7803, doi: 10.1109/CVPR.2019.00799.

[50] Y. Li, J. Bian, Y. Gao, and R. Song, "Video-based deception detection with non-contact heart rate monitoring and multi-modal feature selection," *J. Beijing Inst. Technol.*, vol. 33, no. 3, pp. 175–185, Jul. 2024, doi: 10.15918/j.jbit1004-0579.2024.051.

[51] L. Mathur and M. J. Mataric, "Affect-aware deep belief network representations for multimodal unsupervised deception detection," in *Proc. 16th IEEE Int. Conf. Automatic Face Gesture Recognit.*, Dec. 2021, pp. 1–8, doi: 10.1109/FG52635.2021.9667050.

[52] S. Shilaskar, S. Patukale, P. Oak, and S. Bhatlawande, "An expert system for facial micro-expressions based lie detection," in *Proc. 14th Int. Conf. Comput. Commun. Netw. Technol. (ICCCNT)*, Jul. 2023, pp. 1–10, doi: 10.1109/icccnt56998.2023.10307764.

[53] B. Biçer and H. Dibeklioğlu, "Automatic deceit detection through multimodal analysis of high-stake court-trials," *IEEE Trans. Affect. Comput.*, vol. 15, no. 1, pp. 342–356, Jan. 2024, doi: 10.1109/TAFFC.2023.3322331.

[54] A. Stathopoulos, L. Han, N. E. Dunbar, J. K. Burgoon, and D. Metaxas, "Deception detection in videos using robust facial features," in *Advances in Intelligent Systems and Computing, Springer Science and Business Media Deutschland GmbH*. Cham, Switzerland: Springer, Oct. 2020, pp. 668–682, doi: 10.1007/978-3-030-63092-8_45.

[55] H. Javaid, A. Dilawari, U. G. Khan, and B. Wajid, "EEG guided multimodal lie detection with audio-visual cues," in *Proc. 2nd Int. Conf. Artif. Intell. (ICAI)*, Mar. 2022, pp. 71–78, doi: 10.1109/ICAI55435.2022.9773469.

[56] M. Karnati, A. Seal, A. Yazidi, and O. Krejcar, "LieNet: A deep convolution neural network framework for detecting deception," *IEEE Trans. Cognit. Develop. Syst.*, vol. 14, no. 3, pp. 971–984, Sep. 2022, doi: 10.1109/TCDS.2021.3086011.

[57] S. Venkatesh, R. Ramachandra, and P. Bours, "Robust algorithm for multimodal deception detection," in *Proc. IEEE Conf. Multimedia Inf. Process. Retr. (MIPR)*, Mar. 2019, pp. 534–537, doi: 10.1109/MIPR.2019.00108.

[58] J. Kang, W. Qu, S. Cui, and X. Feng, "Deception detection algorithm based on global and local feature fusion with multi-head attention," in *Proc. 3rd Int. Conf. Image Process. Media Comput.*, Jul. 2024, pp. 162–168, doi: 10.1109/icipmc62364.2024.10586666.

[59] S. Chebbi and S. B. Jebara, "Deception detection using multimodal fusion approaches," *Multimedia Tools Appl.*, vol. 82, no. 9, pp. 13073–13102, Apr. 2023, doi: 10.1007/s11042-021-11148-9.

[60] A. Balmer, *Lie Detection and The Law: Torture, Technology and Truth*. London, U.K.: Routledge, 2018, doi: 10.4324/9781315720258.

[61] K. Suchotzki and M. Gamer, "Detecting deception with artificial intelligence: Promises and perils," *Trends Cognit. Sci.*, vol. 28, no. 6, pp. 481–483, Jun. 2024, doi: 10.1016/j.tics.2024.04.002.

[62] A. Mambreyan, E. Punskaya, and H. Gunes, "Dataset bias in deception detection," in *Proc. 26th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2022, pp. 1083–1089, doi: 10.1109/ICPR56361.2022.9956151.

[63] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 4768–4777.

[64] R. Younisse, A. Ahmad, and Q. Abu Al-Haija, "Explaining intrusion detection-based convolutional neural networks using Shapley additive explanations (SHAP)," *Big Data Cognit. Comput.*, vol. 6, no. 4, p. 126, Oct. 2022, doi: 10.3390/bdcc6040126.

[65] S. Ali, F. Akhlaq, A. S. Imran, Z. Kastrati, S. M. Daudpota, and M. Moosa, "The enlightening role of explainable artificial intelligence in medical & healthcare domains: A systematic literature review," *Comput. Biol. Med.*, vol. 166, Nov. 2023, Art. no. 107555, doi: 10.1016/j.compbiomed.2023.107555.

[66] F. M. Talaat, "Explainable enhanced recurrent neural network for lie detection using voice stress analysis," *Multimedia Tools Appl.*, vol. 83, no. 11, pp. 32277–32299, Sep. 2023, doi: 10.1007/s11042-023-16769-w.

[67] M. K. Camara, A. Postal, T. H. Maul, and G. H. Paetzold, "Can lies be faked? Comparing low-stakes and high-stakes deception video datasets from a machine learning perspective," *Exp. Syst. Appl.*, vol. 249, Sep. 2024, Art. no. 123684, doi: 10.1016/j.eswa.2024.123684.

[68] E. Elaad, "Personality, demographic, and psychophysiological correlates of people's self-assessed lying abilities," in *Detecting Concealed Information and Deception: Recent Developments*. Cambridge, MA, USA: Academic, 2018, doi: 10.1016/B978-0-12-812729-2.00015-X.

[69] G. An, S. I. Levitan, J. Hirschberg, and R. Levitan, "Deep personality recognition for deception detection," in *Proc. Interspeech*, Sep. 2018, pp. 421–425, doi: 10.21437/interspeech.2018-2269.

[70] 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.

[71] H. Li, H. Song, M. Li, and H. Li, "Nonverbal cues to deception: Insights from a mock crime scenario in a Chinese sample," *Frontiers Psychol.*, vol. 15, Feb. 2024, Art. no. 1331653, doi: 10.3389/fpsyg.2024.1331653.

YENI DWI RAHAYU received the master's degree in computer science from the Universitas Islam Indonesia, Indonesia, in 2014. She is currently pursuing the Ph.D. degree with the Institut Teknologi Sepuluh Nopember. She is also a Lecturer of informatics engineering with the Universitas Muhammadiyah Jember. She has published over 40 journal articles and conference papers. Her research interests include computer vision, deep learning, and machine learning.

CHASTINE FATICHAH (Member, IEEE) received the Ph.D. degree from Tokyo Institute of Technology, Japan, in 2012. She is currently a Professor with the Institut Teknologi Sepuluh Nopember, Surabaya. She has published more than 110 computer science-related journal articles and conference papers. Her research interests include artificial intelligence, image processing, and data mining.

ANNY YUNIARTI (Member, IEEE) received the master's degree from The University of Western Australia, Australia, in 2008, and the Ph.D. degree from the Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia, in 2022. She is currently an Associate Professor with the Institut Teknologi Sepuluh Nopember. She has published more than 70 computer science-related journal articles and conference papers. Her research interests include 3-D reconstruction from 2-D images, deep learning, and computer vision.

YUSTI PROBOWATI RAHAYU received the bachelor's and Ph.D. degrees in psychology from Gadjah Mada University, Yogyakarta. She is currently a Professor with Universitas Surabaya. Her expertise lies in the field of legal psychology, also known as forensic psychology. She is a highly regarded figure in the field of forensic psychology in Indonesia, has made significant contributions to the field.