Computers, Materials & Continua

Tech Science Press

DOI: 10.32604/cmc.2024.055311

ARTICLE

LDNet: A Robust Hybrid Approach for Lie Detection Using Deep Learning Techniques

Shanjita Akter Prome[1], Md Rafiqul Islam[2,*], Md. Kowsar Hossain Sakib[1], David Asirvatham[1], Neethiahnanthan Ari Ragavan[3], Cesar Sanin[2] and Edward Szczerbicki[4]

[1]School of Computer Science, Taylor's University, Subang Jaya, 47500, Malaysia

[2]Information Systems, Australian Institute of Higher Education, Sydney, NSW 2000, Australia

[3]Faculty of Social Sciences & Leisure Management, Taylor's University, Subang Jaya, 47500, Malaysia

[4]Faculty of Management and Economics, Gdansk University of Technology, Gdansk, 80-233, Poland

*Corresponding Author: Md Rafiqul Islam. Email: r.islam@aih.edu.au

Received: 23 June 2024 Accepted: 30 September 2024 Published: 18 November 2024

ABSTRACT

Deception detection is regarded as a concern for everyone in their daily lives and affects social interactions. The human face is a rich source of data that offers trustworthy markers of deception. The deception or lie detection systems are non-intrusive, cost-effective, and mobile by identifying facial expressions. Over the last decade, numerous studies have been conducted on deception detection using several advanced techniques. Researchers have focused their attention on inventing more effective and efficient solutions for the detection of deception. So, it could be challenging to spot trends, practical approaches, gaps, and chances for contribution. However, there are still a lot of opportunities for innovative deception detection methods. Therefore, we used a variety of machine learning (ML) and deep learning (DL) approaches to experiment with this work. This research aims to do the following: (i) review and analyze the current lie detection (LD) systems; (ii) create a dataset; (iii) use several ML and DL techniques to identify lying; and (iv) create a hybrid model known as LDNet. By combining layers from Vgg16 and DeneseNet121, LDNet was developed and offered the best accuracy (99.50%) of all the models. Our developed hybrid model is a great addition that significantly advances the study of LD. The findings from this research endeavor are expected to advance our understanding of the effectiveness of ML and DL techniques in LD. Furthermore, it has significant practical applications in diverse domains such as security, law enforcement, border control, organizations, and investigation cases where accurate lie detection is paramount.

KEYWORDS

Artificial intelligence; deception/lie detection; deep learning; facial expression; machine learning

1 Introduction

Lying or deception is one of the most typical and developed human activities, implying the deliberate act of offering false or misleading information. These include fabrications, omissions, and false statements [1]. Deceitful relationships present moral and security risks that are somewhat hazardous, from small lies to severe threats to society. Recently, numerous organizations have been

using private forensic research organizations and requesting their workers take LD tests to combat fraud, spying, and false resume cases. A significant number of research communities have aimed to identify such behaviors. Deceit and cognitive process remain two interconnected strands of research in psychology and relevant sciences. As Fernandes et al. [2] indicated, deceit is characterized by conscious or unconscious behaviors such as reduced speech rate, reddened face, changes in invoice frequency, absence of concern, differently sized pupils, and more congealed posture. As suggested above, there are subtle lies that may help several people interact better, and there are lies that can cause great harm. Over the years, deceitfulness has emerged as a modern social problem, the consequences of which are felt in many fields and have led to considerable losses in millions of US dollars. To minimize the damaging ripper, it is imperative to identify potential liars in real time as trends surge.

LD is of the utmost significance everywhere because it assists in assessing an individual's integrity in statements or claims where needed. Some of the preceding circumstances include the police force, education, healthcare, governmental offices, border checkpoints, military and employment security checks, telecommunications, and informants at embassies and consulates around the globe. This could be significant for both personal and occupational reasons. Some of them include exposing fake or corrupt practices in organizations, maintaining the sanctity of legal practices and the justice system, and protecting the security of a country. Moreover, in business, using lies when recruiting results in bad hires that damage corporate productivity and reputation [3]. Thus, it is much more critical to identify lying and studying this subject can improve interpersonal relationships because it will allow people to understand the goals and needs of others. However, LD remains a challenging task because there is no sign which always indicates lies. It has been found that most people can identify lies at an average level of about 54% without employing specialized instruments [4]. This is contrary to what many people would wish to believe: it is possible to tell whether a person is lying by observing their body language, particularly their face. However, recent innovations and advances in DL and ML provide hope for improving the accuracy of identifying lies. These methods present a more nuanced understanding of lies, yielding higher reliability and accuracy for judging integrity.

Researchers have worked hard in the past few years to explore artificial intelligence (AI) for LD [5]. Existing techniques employed either mechanical indicators or behavioral signs, thereby attaining a moderate level of accuracy in detecting deceit. Traditional procedures such as polygraphs were primarily applied, except they had their demerits, such as intrusiveness, slowness, and malleability to deception. Several innovative methods are utilized daily to energize a thorough framework to reach the best precision. The growing incorporation of technology in society presents new ways of lying that make it crucial to establish reliable techniques for identifying when a person is telling a lie. How AI systems function and produce results can be compared to a person's cognitive system [6]. Cognitively capable machines and robots can identify a person's deceitful state by analyzing their facial, vocal, and psychological cues. However, some potential remains, which can be deemed an opportunity to leverage multiple modalities to learn from and apply sophisticated learning techniques. Newer studies on LD involving face and facial expressions and pulse rate [7], 3D mapping [8], multimodal fusion analysis [9], and computer vision [10] have most recently brought DL into play to solve the deception detection problem. Despite the many LD systems that have been developed, this field remains open to improvement. In a previous study, the authors employed a range of ML, DL, and visual analytics techniques to examine text data in this domain [11,12]. Therefore, more research would help explore this issue more extensively. Additionally, it is crucial to have a substantial amount of high-quality data that includes samples of deceptive behaviors, but there is a constraint in this regard. However, due to the limited number of video clips (121) and the fact that only half of them include deceit, this dataset is

inadequate for effectively training a deep neural network-based model, which has been the prevailing technique in recent advancements in automatic deceit detection (ADD) [13].

Recent research on LD can be broadly divided into two groups: While contact-facilitating techniques are contact-based, there are also non-contact-facilitating techniques. The first group involves methods based on extracting silico and bio signal measurements, while the second group involves approaches based on verbal, thermal, visible, and audible information [14,15]. Visual data techniques can present a substantial amount of information that can be used to determine whether a person is telling the truth or lying. Nonverbal cues such as eye contact, nodding, and even tone of voice is familiar to us as they express various emotions, while some micro expressions may suggest dishonesty. In addition, the data obtained from visual images, including face images, involves brow movements and lip contour. Whether a person makes eye contact or not, it is less invasive than other interrogative approaches. It is worth noting that, in LD, multiple modalities are currently more popular than single modalities used in the relatively recent LD systems, and they offer promising results in terms of the increase in the accuracy rate. The proposed method has strengths in acquiring various cues of deception and incorporating multimodal data, but specific challenges exist with the data collection aspect of many current multimodal datasets.

Nevertheless, accuracy can be enhanced by focusing on single-modal approaches and video analysis. Despite the challenges associated with using video data for LD, such as complex data processing, difficulty in identifying signals from video, and a need for deceptively labeled video data, there is still room to improve the LD process. Therefore, this study aims to explore these opportunities, apply advanced methods, and address data scarcity issues. The limited availability of datasets may hamper the performance of existing models, so we created our dataset to improve the data quality.

This research presents the creation of a novel dataset that exhibits enhanced data quality and increased quantity. To implement the ML model, we initially collected facial action points. Then, we applied several ML and DL models to perform the classification. Using layer ensembling techniques, we have developed a hybrid model that significantly enhances the findings of this research. This research makes the following pivotal contributions: First, we explore the ML and DL techniques for LD, analyze previous studies, and highlight features. Secondly, we created our dataset by conducting interview sessions, alleviating the difficulties involved with the limited availability of video datasets for LD. We employ various ML and DL methods such as random forest (RF), k-nearest neighbors (KNN), support vector machine (SVM), decision tree (DT), multi-layer perceptron (MCP), convolutional neural network (CNN), long short-term memory (LSTM), DenseNet121, ResNet50, Inception, and Vgg16 to detect lies. Finally, we develop a hybrid model by ensembling two distinct layers, resulting in improved accuracy for lie classification.

The following is just how the rest of the paper is structured: Section 2 discusses closely related work. Section 3 explains the work's methodology. Section 4 describes the experimental result, and Section 5 provides the key insights and limitations. Finally, Section 6 concludes with future directions.

2 Background Study

The field of LD has recently seen a rise in innovative methods driven by developments in DL and ML. Notably, to enhance this field, researchers are now exploring more and more advanced techniques to get deeper insights that improve the outcome for LD. Hence, this section discusses the methods and approaches utilized in conducting LD research and acknowledges state-of-the-art and new developments.

2.1 Lie Detection

Lying has always been a severe social and legal issue as well as a moral one. It involves presenting false information while simultaneously displaying facial expressions and physical gestures that may not align with the truth. Therefore, LD, on the other hand, includes the aspect of being able to determine whether a person is telling the truth or lying. These use different methods, such as verbal and non-verbal communication, physical responses, and behavioral patterns. While it is sometimes impossible to do so, there are psychological techniques like open-ended questioning, asking questions for which the respondent is least likely to have prepared an answer, and putting the respondent through a cognitive process that will make it both tiresome and difficult to lie.

Day by day, the rate is increasing, and lying is a significant reason for the growing rate, so LD is an essential concern to be addressed. However, some people are skilled liars, which adds to the difficulty of detecting lies. The characteristics of a competent liar have been the focus of surprisingly minimal research, although six characteristics stand out as particularly significant. The most skilled liars are those who (i) exude confidence through their natural behavior, (ii) refrain from finding it cognitively challenging to lie, (iii) do not feel fear, guilt, or joy when they are lying, (iv) good actors who convey an honest image, (v) whose good looks may imply virtue and honesty, (vi) are good psychologists.

Not only are people demotivated to expose liars, nor is the task of detecting lies hard, but they also make systematic mistakes in the assessment process. In that situation, machines are helpful and give us more profound and valuable insights.

2.2 Machine Learning and Deep Learning Classification

Deception detection has been mainly focused on in recent years due to AI and ML improvements. Multiple research papers have investigated different techniques and approaches to enhance the precision and dependability of automated systems that detect deception. This field is characterized by integrating disciplines such as philosophy, control theory, neuroscience, psychology, and neuroscience, all contributing to its development [16]. Because of its complexity, LD is still a developing study area, with researchers constantly experimenting with different approaches. However, ML models have progressed in this field and revealed many remarkable discoveries. ML can be used to complete various tasks, such as anomaly detection, association rules, grouping, regression, and classification. Although ML, with the integration of multiple modalities, dramatically improves the categorization field by utilizing several data sources to enhance the accuracy and reliability of predictive models and achieve a more comprehensive understanding of the data.

For example, Sehrawat et al. [17] proposed a multimodal technique that integrates visual, aural, and textual data to detect fraudulent behavior. Their model utilized advanced DL techniques to extract and combine data from many modalities, resulting in a substantial enhancement in detection performance compared to models that only employ a single modality. Wu et al. [18] developed a multimodal method using information from audio streams, video, and transcriptions to extract a new modality. Facial movements, which are considered micro-expressions, are once again utilized and retrieved using a classifier that has been trained. In addition, video sequences are analyzed using IDT (improved dense trajectories). The audio modality extracts and encodes mel-frequency cepstral coefficients (MFCC). Ultimately, glove (global vectors for word representation) analyzes the transcriptions. The researchers studied individual features and their combinations using a straightforward late fusion method, achieving optimal outcomes when integrating all the modalities.

An SVM classifier with an AUC of almost 70% was obtained by Rill-Garcia et al. [19] after they gathered a new data set of 42 videos in which interviewees discussed their truthful and deceptive

opinions about abortion. They investigated high-level features for the deception detection challenge extracted from videos utilizing open tools at the "view" and "modality" levels. Their approach included a unique set of boosting-based techniques that were competitive with LSTM, a DL approach. The LSTM layer is provided with frame analysis sequences and 200 hidden units. The output of the LSTM layer is then passed to a fully connected layer with 100 hidden units. Finally, the output of this fully connected layer is fed into another fully connected layer that has a single output. Only the visual and audio modalities were used to test this design, and the results were 0.560 and 0.730 AUC, respectively (the SVM values were 0.574 and 0.638, respectively). Khalil et al. [20] introduced brain computer-based lie detection using bio signals and processed with various DL techniques. The accuracies of their models ranged from 44.0% to 86.0%.

Kawulok et al. [21] explored using rapid smile intensity detectors in conjunction with SVM classifiers to extract temporal traits. They analyzed the time series of smile intensity without localizing or tracking facial landmarks using only a face detector. An SVM classifier is then used to enhance training using poorly labeled datasets. The grin detectors are then trained using uniform local binary pattern features. This makes it possible to distinguish between posed and spontaneous reactions in real-time. Su et al. [22] aimed was to evaluate the applicability of computer vision techniques for face cues to LD in high-stakes scenarios. They conducted facial analyses of eye blinks, eyebrow motions, wrinkle occurrences, and mouth motions. They integrated them with a facial behavior pattern vector employing invariant 2D features from nine distinct facial regions. The RF algorithm classifies the patterns as either truthful or deceptive.

Karnati et al. [1] conducted several experiments, extracted features from distinct modalities, and introduced a deep convolution neural network framework for detecting deception called LieNet. It evaluates each modality individually to find discriminative clues for deceit detection. LieNet includes numerous scale kernels, which aid in extracting more resilient and noise-invariant features from different image patches and shorter connections between layers closer to the input and output, improving the efficiency and dependability of training. Ding et al. [13] extracted R-CNN network detection features to enhance face-focused cross-stream network (FFCSN) through adversarial learning, meta-learning, and cross-stream integration of the over-correlation class into the sample. 93.16% accuracy has been obtained. In addition, the model achieved 84.33% accuracy in face recognition. FFCSN integrates meta-learning with adversarial learning. Meta-learning, which involves learning to learn, enhances the model's generalization capability and prevents overfitting to limited training data. Meanwhile, the data augmentation technique involves adopting adversarial learning-based feature synthesis. By efficiently merging these two FFCSNs, even the existing real-life deception detection benchmarks with very sparse data can be trained.

Even though the research used a range of characteristics extracted from various modalities, nearly all used the same datasets due to the limited amount of publicly available video datasets. Additionally, single DL models were used in the studies focused on multimodal datasets; hybrid models are somewhat uncommon. This research project focuses on this by developing a hybrid model and a new dataset.

2.3 Facial Expressions

Facial expressions have a significant role in deception detection. However, more than a decade of empirical research has yet to show that this assertion is true. Though facial expressions are the best way to receive hints and aid in determining whether a person is telling the truth or trying to hide information with their expressions, many psychologists have shifted their focus from studying this area

toward more cognitive techniques for detecting lying [23]. These expressions aid in the understanding of what drives everyone. Similarly, machines are beneficial in lie detection as they can easily collect and recognize expressions. Machines use facial action points to interpret facial expressions based on multiple facial movements. According to Constâncio et al., emotional traits are essential for detecting deceit since lying is prompted by and dependent on emotional states, and facial expressions are the primary means of expressing emotions [24]. These expressions are divided into two categories: macro expressions and micro expressions, which are discussed in this section.

Micro expressions are quick, unconscious facial gestures individuals generate when attempting to conceal their feelings or intentions [25,26]. They are considered the most potent and persistent cues for deception. Research has demonstrated that micro expressions, which are emotions that individuals attempt to conceal, may be effectively recognized manually. These micro-expressions are a dependable indicator of deception [27,28]. However, due to the excessive rarity, several studies have expressed concerns about micro expressions [29]. There is no association between accuracy and causation when assessing deception detection ability utilizing micro expression recognition and training [30,31]. Although micro expressions can be a sign of genuine emotions, the diagnostic value as markers of deception is diminished because these are erratic, unpredictable, and do not always occur during fraudulent occurrences. Macro facial expressions are regular facial expressions coded with the facial action-unit (AU) system [32]. While micro expressions in natural settings may sometimes generate AU with very low intensity, they have a longer duration than micro expressions. Prior research on facial expression identification has employed DL with other methodologies. This involved following a conventional sequence of steps, which included extracting landmarks, recognizing faces, and doing AU regression [33,34]. As an illustration, Chang et al. [35] employed a convolutional neural network (CNN) to calculate Action Units (AUs) and subsequently utilized these outcomes to forecast degrees of valence and arousal. They presented FATAUVA-Net, an integrated DL architecture, to accomplish the tasks of valence-arousal estimation, action unit detection, and facial attribute recognition. They used AUs as a mid-level representation in a DL system for V-A estimation. In contrast, Khorrami et al. [36] employed a comprehensive methodology to accurately forecast valence and arousal levels based on normalized facial images. CNN was used to extract AUs implicitly. They demonstrated that CNNs trained to recognize emotions could model high-level properties closely matching FAUs qualitatively and statistically.

2.4 Exploring Lie Detection in Contemporary Computer Vision Studies

Rapid advancements in ML, computer vision, and computer graphics have made it feasible to produce incredible solutions for almost every sector [37]. In the same way, the field of deception detection for these advancements has been expanding constantly. Several studies have been identified that have focused on the automation of computer vision systems for LD. Three cues have been highlighted: facial expressions, gaze aversion, and body language. Some of the three signals are mixed in specific situations. Facial micro-gestures prove valuable in LD, while facial expressions provide essential insights into deceptive attitudes [38]. Computer vision is a branch of AI and computer science that strives to equip machines to interpret and understand digital images and videos. With complex algorithms, automatic analysis and interpretation of visual information recorded or produced through video are possible. It presents an efficient way toward LD.

For example, in the field of lie detection, Serras Pereira et al. [39] compared a perceptual and an automated vision-based system with a primary focus on children. They have investigated the cue validity of body motions for this kind of classification and have employed a thorough methodology to analyze the ease with which it can be ascertained whether a youngster is being truthful or not

in a gaming environment. They examined kids between the ages of six and seven, which naturally raises the question of how their conduct relates to people in other age groups. The findings indicated that judges could accurately differentiate between authentic and false images in a series of perception studies and that the more movement the kids displayed in a clip, the more likely it was to be fraudulent. The eye-tracking study also showed that judges focus primarily on the face region despite movement occurring in other body parts. Singh et al. [40] developed a novel method for using image processing techniques to detect lies. Their method effectively monitors and counts a person's eye blinks from an input video image. The inter-eye blink interval and the measured blink rate can quickly determine whether someone is telling the truth or lying. Su et al. [22] explored whether facial cues could be used in high-stakes scenarios as markers of deceit. Three stages of computer video analysis comprise their method: preprocessing, dynamic feature analysis, and classification. Face detection and facial landmark localization are initially used in the preprocessing step to register the face. Subsequently, an anthropocentric model divides the face into multiple facial regions.

Satpathi et al. [41] developed an approach for detecting deceit using facial thermal imaging in human subjects. They diligently created a database based on nearly real-life theft cases by observing several individuals over time at a government hospital in exchange for a free health examination. Feng [42] developed a comprehensive end-to-end method for identifying lies in videos primarily based on facial expressions. They first used video clips to create time-stacked images of people's facial expressions. Next, they used a computer vision technique to turn the images of people's faces taken on camera into encoding vectors, which they further classified to predict whether it's true or false. Moreover, Khan et al. [10] applied advanced computer vision and ML techniques, and a real-time deception detection method has been developed to mimic non-verbal fraudulent behavior. Wu et al. [18] used vision, text, and audio data videos to detect deception. The novel method combined high and low-level micro expressions to attain a high level of effectiveness.

2.5 Existing Deception Datasets

A significant constraint for deception research is the need for more meticulously annotated, precise, and well-constructed data sets. Since real-world deceptive data with annotation is typically hard to get by, many researchers create their deceptive datasets by paying people to behave deceitfully. The four publicly accessible deception detection datasets, three with low-stake contexts and one with high-stake contexts are explored as shown in Fig. 1.

(a) Box-of-lies dataset (Reprinted from Reference [44])

(b) Real-life trial database (Reprinted from Reference [45])

Figure 1: (Continued)

(c) MU3D dataset (Reprinted from Reference [46])

Figure 1: Publicly available deception video datasets

2.5.1 Bag-of-Lies

The multimodal dataset Bag-of-lies (BgL) is intended for deception detection across many modalities, such as audio and video [43]. The dataset provides a realistic environment for data collection while attempting to investigate the cognitive side of deception and combining it with vision. Thirty-five distinct subjects comprise BgL, which offers 325 annotated samples with a balanced distribution of truth and deception. Researchers use this dataset to improve deception detection algorithms and make them more applicable to practical situations. The samples are obtained this way: A participant is asked to describe a photo shown on a screen before them. The participant can decide whether to tell the photo honestly or dishonestly.

2.5.2 Box-of-Lies

The Box-of-Lies (BxL) dataset is derived from the game of the same name featured on a late-night television program [44]. In this program, the host and guest alternately describe an object as genuine or fake or, depending on the opponent's move, determine whether the description appears accurate. In addition to linguistic and conversational elements, various nonverbal cues, such as those from the gaze, eye, lips, eyebrows, face, and head, have been manually ground-truthed throughout time.

2.5.3 Real-Life-Trials

Pérez-Rosas et al. emphasized the importance of identifying deception in court trial data because of the significant stakes involved [45]. To achieve this, they unveiled a multimodal deception detection system that uses verbal and nonverbal modalities to discern between defendants' and witnesses' truthful and false statements. They also introduced Real-Life trial (RL), a novel dataset comprising clips from real court proceedings.

2.5.4 Miami University Deception Detection Database

A free resource, the Miami University Deception Detection Database (MU3D), has 320 videos of people stating the truth and lying [46]. Eighty targets, including twenty black females, twenty black males, twenty white females, and twenty white males, were recorded discussing their social interactions truthfully and deceptively. Three hundred twenty videos were produced, fully spanning target race, target gender, statement valence, and statement truthfulness. Four movies encompassing both positive and negative truths and lies were made by each target, creating a completely crossed dataset that can

be used to study research areas. Naive raters transcribed and rated the videos and produced subjective assessments and descriptive analyses of the video features.

3 Methodology

This section explains how various ML and DL approaches are used, as well as the design and implementation, as illustrated in Fig. 2. It also provides comprehensive details on the collection and analysis of data, feature extraction, and the distribution of nonverbal characteristics across truth-tellers and liars.

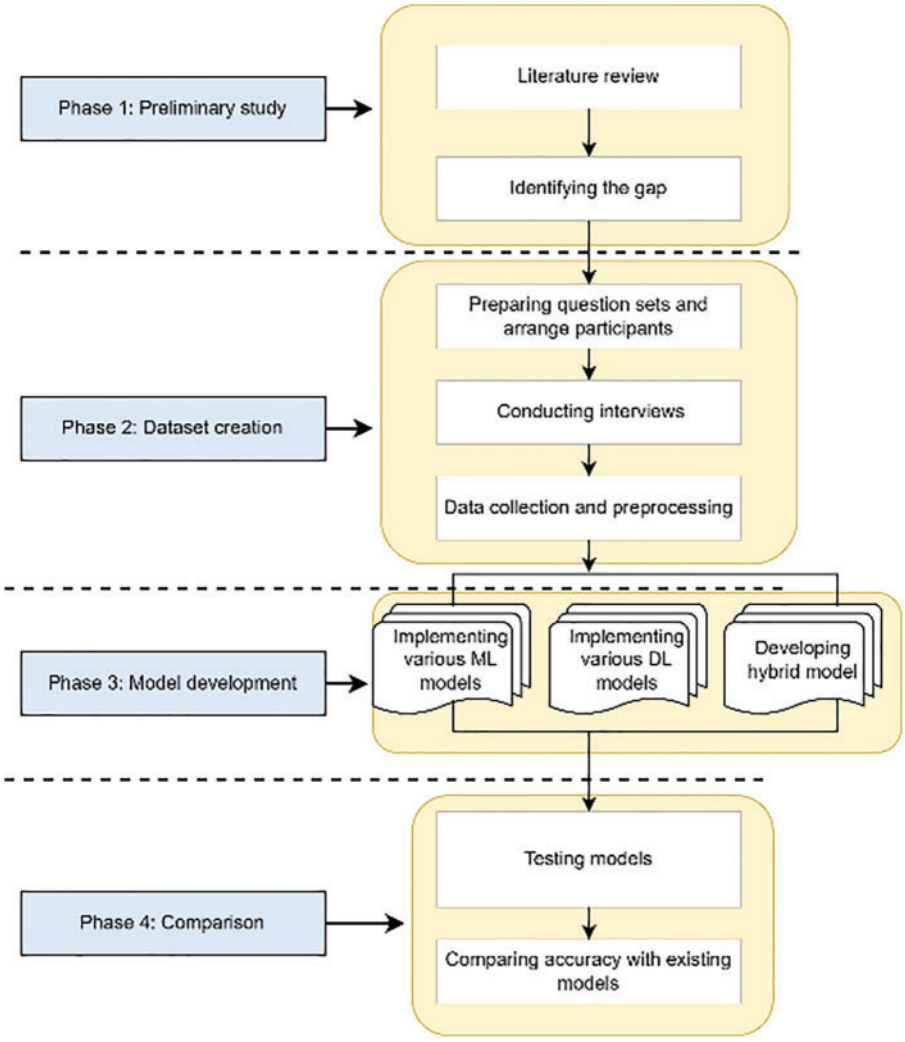


Figure 2: Overview of the comprehensive project architecture

3.1 Data Collection

We have experimented with our own created video data in this study. We used a straightforward strategy to assist the students in conducting interviews using a set of fundamental questions, as shown in Table 1. We asked the participants to answer all the questions and let them know the video sessions

would be recorded. The interviewer asked the questions and recorded their responses. The participants were asked to provide truth and lie answers to the 11–12 questions. We have selected 80 participants for the interview sessions, which ensured a diverse range of individuals. Before starting the interview session, participants were told to choose one set and respond to every question in that set while having their facial expressions recorded for the sessions. We received two types of responses, truthful and lying, and based on this, we have categorized our video data. We were careful to ensure the detailed capture of the participants' facial expressions during the entire duration of the interview session. The participants were male and female, 24 and 56 in number, aged 18 to 40, and coming from different areas of study. In the final stage of work, all this information has been stored and is ready for comprehensive analysis and utilization in our ongoing research endeavors.

Table 1: List of questions used during the interview session for video data capture

Question set 1	Question set 2
1. Do you have your breakfast today?	1. What is your current citizenship?
2. What is your area of study?	2. What are you studying now? Do you like this?
3. Which subject did you enjoy studying and why?	3. Which is your favorite subject and why?
4. Which subject did you dislike studying and why?	4. Do you enjoy your classes? why/why not?
5. Have you ever cheated on an exam? if yes tell us about it.	5. Do you like assignments? why/why not?
6. What leadership position did you hold?	6. Do you consider yourself a kind person? Give an example of your kindness.
7. What is the most memorable event you have been involved in? What was your role?	7. What was the most challenging situation you have ever faced?
8. What has been the most complex challenge you have faced in your life?	8. Has anyone cheated on you before, and how?
9. Whom have you fought with the most? what caused the fight, and what is the state of the relationship now?	9. Did you reveal any secret of someone who trusted you? why?
10. Do you consider yourself a patient person? give an example of your patience.	10. Did you ever deceive a family member/friend? How?
11. Have you visited other countries? What do you like or dislike about traveling? give an example of your patience.	11. Do you have an iPhone? which model?
12. How do you manage stress and self-care?	12. How do you manage stress and self-care?

Following the data collection phase, CapCut split the videos into video clips. Every question round was continually recorded; therefore, each interview had to be divided using CapCut into many separate video clips. Containing the single respondents' responses, the only data points that comprise the corpus. After that, each response was saved as a video clip. The files are ultimately assigned a true or false response. We collected a total of 900 video clips, and after completing the cleaning process, we

employed 700 video clips for our experiment. We took frames from the video clips and preprocessed them for further analysis. The frames were extracted using the formula duration per video clip/50. A total of 6000 images (frames) were captured for further experiments.

3.2 Data Processing

Face cropping is an approach to automatic face detection. First, the original image is converted to grey-scale, and then grey-scale images are converted to binary by applying the threshold technique, as shown in Fig. 3. We have calculated the contours from which the bounding rectangle was calculated using these binary images. After that, the region of interest (ROI) is defined based on the bounding rectangle, and the image cropped according to ROI. To get a better result, the cropped images zoomed in to view the face with a better view.

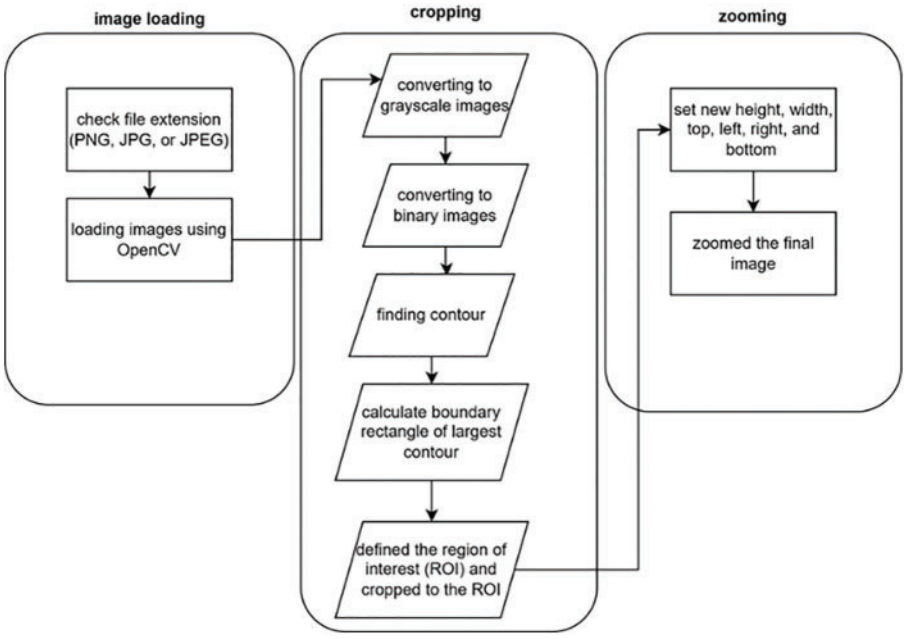


Figure 3: Flowchart illustrating the comprehensive process of image data processing

Image filtering removes images that do not satisfy specific selection criteria. These images can significantly affect the model's accuracy. To overcome this issue, we manually checked all zoomed-in images we received. In some images, the upper part of the face is cut away, while the lower part is missing in others. As a result, we filter these images to improve the model's accuracy and quality of our dataset, ensuring that they are ready to fit each model

3.3 Feature Extraction

The preprocessed datasets have gone through the feature extraction stage to extract the necessary features. The feature extraction phase followed several steps, such as (i) loading the dataset, (ii) initializing any landmark extraction library, (iii) extracting facial landmarks, and (iv) saving the extracted features in a suitable format (e.g., CSV, JSON) for subsequent analysis and model training. In this study, dlib (a facial landmarks extraction library) library has been utilized, which extracted 68 facial landmarks as shown in Fig. 4.

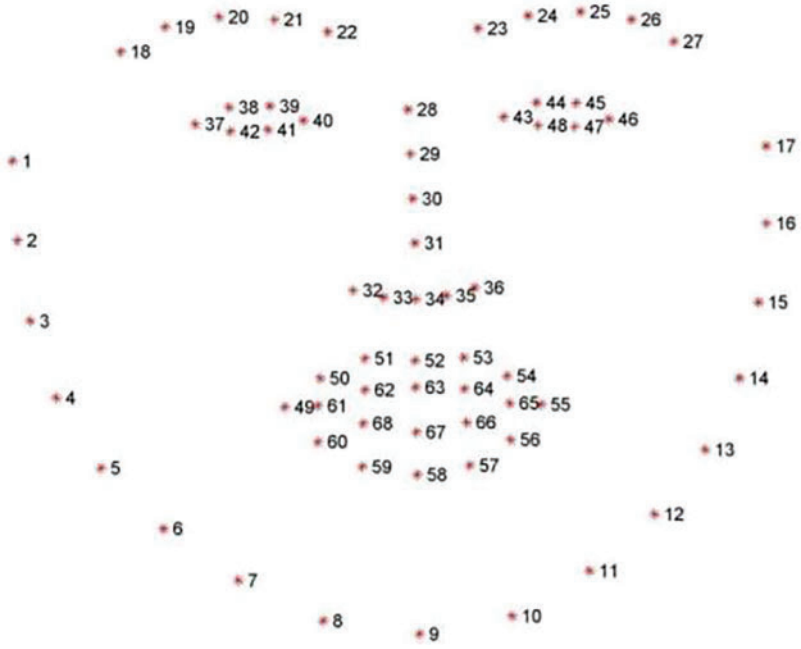


Figure 4: Example of facial landmarks detected using dlib (68 landmarks)

3.3.1 Landmark Acquisition Using Dlib

Face and facial landmark detection can be useful in developing an LD model. Facial landmarks are critical points on a face that can be used to analyze facial expressions and emotions. These landmarks include the corners of the eyes, the tip of the nose, the corners of the mouth, and the eyebrows. By detecting and tracking these landmarks, it is possible to analyze changes in facial expressions and detect when someone is lying. Several techniques can be used for facial landmark detection, including ML-based and DL-based methods. These methods involve training a model on a large dataset of images with annotated facial

Landmarks and then using this model to detect landmarks in new images. Facial landmark detection is essential in computer vision applications, particularly facial recognition, analysis, and expression detection. We collected data for each frame of the recorded videos to extract facial traits, including mouth position, gaze, and a set of coordinates known as landmarks, from the subject's facial image. To obtain such information, we utilised a software package known as dlib. The facial landmark extractor utilizes Python and the dlib library to process the cropped facial photos. Dlib can analyze real-time facial behavior, including landmark detection, head position estimation, facial emotion recognition, and gaze estimation. The suggested methodology extracts facial landmarks from specific regions, including the mouth, right and left eyebrow, right and left eye, nose, and jaw. The facial areas above consist of a total of 68 points, each of which is represented by its two-dimensional coordinates. The facial landmark detection can be described as:

$$Fv = \{(xi, yi) \mid i = 1, \ldots, 68\} \tag{1}$$

Fv represents the collection of visual features, while (xi, yi) indicates the position of the i-th face landmark.

3.3.2 Overview of Extracted Features

From the collected facial landmarks, we were able to extract a few facial traits. Furthermore, we delineated and computed the subsequent attributes constituting an individual's facial expressions: brow movement and tilt, eye movement and area, mouth area, and blink rate. The amount of time variation in the iris (the center of the eye) coordinates is used to indicate eye movement. The framework of the computation for each feature is explained in this section. Points P0 through P67 stand for the coordinates from 0 to 67

Eyebrow tilt (left and right) The right and left eyebrow's slope was determined using the following points: P17, P18, P19, P20, P21 (P22, P23, P24, P25, P26).

Region of the eyes (left and right) This is the hexagon region created by joining the six points that encircle the right (or left) eye. The collection of points P36, P37, P38, P39, P40, P41 for the right eye (P42, P43, P44, P45, P46, P47 for the left eye) is used to calculate this area.

The mouth (internal) This is the portion of the mouth's inner perimeter made by joining the points P60, P61, P62, P63, P64, P65, P66, and P67.

The mouth (external) This is the region of the dodecagon created by joining the points on the mouth's outer perimeter, P48, P49, P50, P51, P52, P53, P54, P55, P56, P57, P58, and P59.

3.4 Proposed Model

Various stages followed to introduce our architecture. Fig. 5 depicts the general architecture of our hybrid model. Preprocessing the data is the initial stage. To begin with, the dataset is cleaned to remove any corrupt or incorrectly labeled images. Any incorrect images are eliminated to maintain data integrity and stop the model from learning from inaccurate or noisy samples. The amount and diversity of the dataset are then increased using data augmentation techniques. Random crops, scaling, and flipping are examples of common augmentations. This stage makes the model more resilient and improves its ability to generalize on data that hasn't been seen before. After that, every image is downsized so that the deep learning model may use them all at the same size. Nowadays, DL models, particularly CNNs, have demonstrated remarkable performance in image detection tasks. Applications for CNN include image classification, video and image recognition, and image segmentation. While CNN gathers facial traits from real-time streams, it keeps them in windows as follows: $R^{h\times w\times c}$ where h is the height, w is the width, and c is the number of channels in the input image. This research study modified the conventional model design to enhance its appropriateness for a specific analytical task. The Vgg16 and DenesNet121 models were chosen as the foundational models and were subsequently merged to create the hybrid model. Through the dynamic modification of the importance of different spatial regions in the feature maps, these attention modules enable the model to focus on relevant facial attributes and fraud artifacts. Our hybrid model named "LDNet" combines the best aspects of Vgg16's straightforward architecture with the feature reuse of DenseNet121. The DenseNet121 and Vgg16 architectures are represented by the two distinct streams via which the suggested hybrid model processes visual inputs. Several changes were made to the model, including adding layers explicitly designed to help with feature extraction. Advanced regularization methods, like dropout and batch normalization, were also added to lower the chance of overfitting and improve the model's generalization ability. We thoroughly validated these modifications through methodical experimentation, proving the enhanced performance of the model.

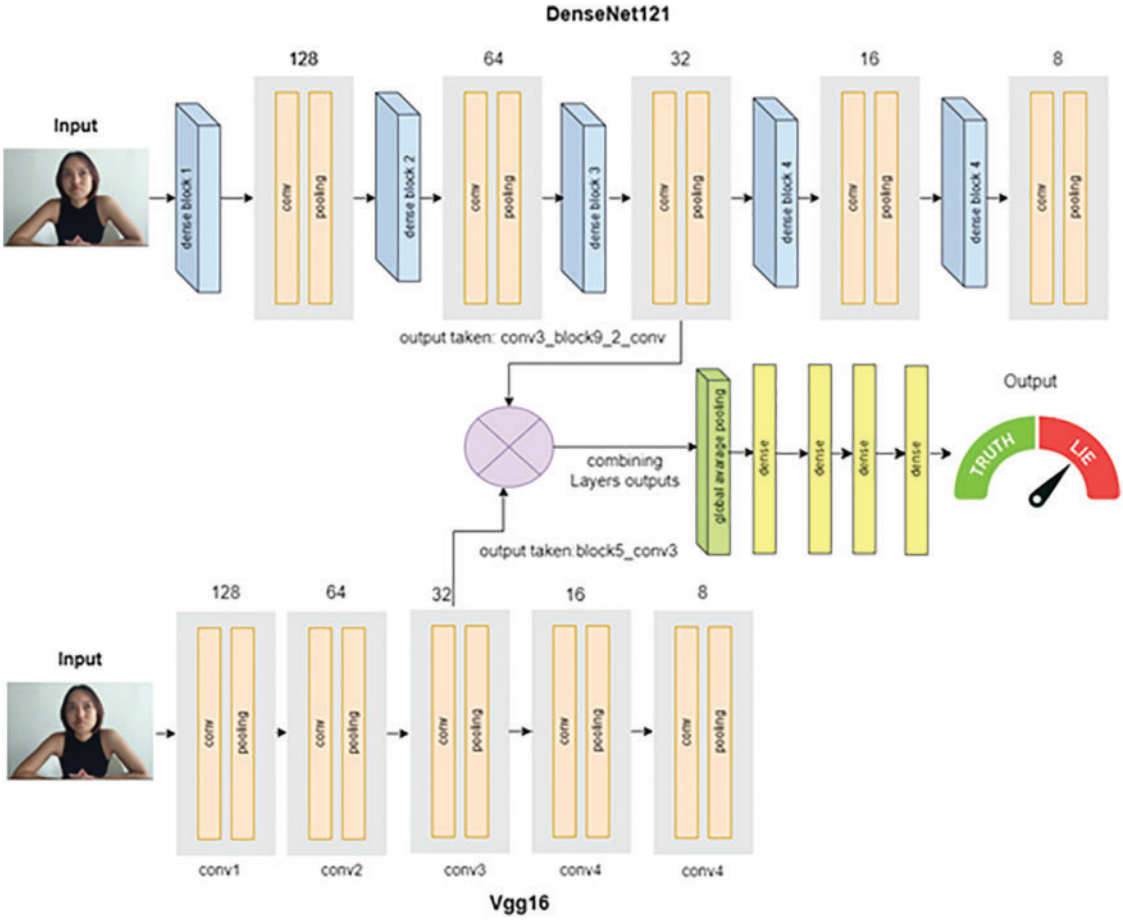


Figure 5: An overview of the LDNet framework, which integrates layers from both Vgg16 and DenseNet121 architectures

The layers of the CNN models used in this architecture are the convolutional layer: The CNN structure's initial layer, responsible for extracting features from the input image. The mathematical process is applied to the input image by sliding the size M × M filter. This results in a feature map that provides general input information, such as edges and corners. Pooling layer: Usually placed after the convolutional layer, the pooling layer's main objective is to minimize the size of the feature map produced by the preceding convolutional layer. To achieve this, the pooling layer weakens the link between layers. Fully connected layer: This layer makes the last layers and is typically positioned ahead of the dense layers to create connections between neurons in two separate levels. Dense layer: Located at the end of the CNN design, the thick layer's primary functions are to execute vector multiplications and receive inputs from all earlier layers. Weights: Each neuron in the neural network applies a specific function to input values it receives from the layer above to produce an output value. We refer to the weight and bias as filters since they represent certain qualities, like the input's form. Dropout: Dropout is a regularization technique that is extensively used to decrease over-fitting in neural networks by preventing the adaptation of the input data. The dropout function randomly omits visible and hidden units during the training phase. Softmax: This function converts a vector of real values into another vector of fundamental values that sums to one because the input values can be zero, harmful, or positive. Converts the data into a range of 1 to 0 so that the probabilities can be understood. To provide

a more precise understanding of our developed model, a high-level algorithm is being outlined by the following step-by-step process as detailed below:

Algorithm 1: Hybrid model for lie detection

1: Input shapes and dimensions
Original image: $X_{\text{original}} \in \mathbb{R}^{128 \times 128 \times 3}$
Segmented image: $X_{\text{segmented}} \in \mathbb{R}^{128 \times 128 \times 3}$
2: Feature extraction using Vgg16
Load Vgg16 pre-trained model
Extract features from layer "block5_conv3"
Output tensor shape: $X_{\text{Vgg16}} \in \mathbb{R}^{4 \times 4 \times 512}$
Apply Global Average Pooling to obtain $X_{\text{Vgg16\_pooled}} \in \mathbb{R}^{1 \times 512}$
3: Feature extraction using DenseNet121
Load DenseNet121 pre-trained model
Extract features from layer "conv3_block9_2_conv"
Output tensor shape: $X_{\text{DenseNet}} \in \mathbb{R}^{16 \times 16 \times 512}$
Apply global average pooling to obtain $X_{\text{DenseNet\_pooled}} \in \mathbb{R}^{1 \times 512}$
4: Combine features
Concatenate pooled features:
$F\text{combined} = \text{concat}(X_{\text{Vgg16\_pooled}}, X_{\text{DenseNet\_pooled}}) \in \mathbb{R}^{1 \times 1024}$
5: Apply batch normalization and dense layers
First dense layer:
$F_{\text{dense1}} = \text{ReLU}(\text{BN}(W_1 \cdot F_{\text{combined}} + b_1))$
where $W_1 \in \mathbb{R}^{1024 \times 256}$, $b_1 \in \mathbb{R}^{256}$, BN denotes Batch Normalization, and the output $F_{\text{dense1}} \in \mathbb{R}^{256}$
Second dense layer:
$F_{\text{dense2}} = \text{ReLU}(\text{BN}(W_2 \cdot F_{\text{dense1}} + b_2))$
where $W_2 \in \mathbb{R}^{256 \times 128}$, $b_2 \in \mathbb{R}^{128}$, and output $F_{\text{dense2}} \in \mathbb{R}^{128}$
Third dense layer:
$F_{\text{dense3}} = \text{ReLU}(\text{BN}(W_3 \cdot F_{\text{dense2}} + b_3))$
where $W_3 \in \mathbb{R}^{128 \times 64}$, $b_3 \in \mathbb{R}^{64}$ and output $F_{\text{dense3}} \in \mathbb{R}^{64}$
6: Output layer
The final output layer computes the logits for each class:
$\text{logits} = W_4 \cdot F_{\text{dense3}} + b_4$
where $W_4 \in \mathbb{R}^{64 \times 2}$, $b_4 \in \mathbb{R}^2$
Softmax activation:
$P(\text{truth} \mid \text{input}) = \text{softmax}(\text{logits}) \in \mathbb{R}^2$
The softmax function converts logits into probabilities for each class (truth or lie).

DenseNet121 Stream: Dense blocks and transition layers make up the DenseNet121 stream. Several convolutional layers, each coupled to all other layers in a forward way, comprise the dense blocks. The feature maps between dense blocks are made less dimensional by transition layers, which comprise pooling and convolution procedures. The convolutional output is retrieved from the ninth dense block's third convolutional layer for additional processing.

Vgg16 Stream: To gradually lower the feature map dimensions, the Vgg16 stream is organized with successive convolutional procedures divided into blocks and then max-pooling layers. The output of

the fifth block, following the third convolutional layer, is chosen for fusion with the DenseNet121 stream.

Feature Fusion: Using the output maps from the DenseNet121 and Vgg16 blocks, this method combines the best features of both models. To ensure that the resulting feature maps are beneficial and effective to the classification objective, the fusion process is improved.

Classification Pipeline: Global average pooling is used for the fused feature maps to minimize overfitting and lower the computational cost. The pooled features are then sent through multiple dense layers that operate as fully connected layers to learn non-linear feature combinations. Finally, using a softmax activation function, the model generates probabilities for the binary classes "truth" and "lie". LDNet aims to increase classification accuracy by carefully combining two robust architectures. For the hybrid model, the algorithm starts by using the Vgg16 network with the "block5_conv3" layers to extract the features, as shown in Algorithm 1. And then reduced to a feature vector of size 1×512 through global average pooling. At the same time, DenseNet121 extracts features from the "conv3_block9_2_conv" layer, while the latter is narrowed down to 1×512 through global average pooling layers. These two feature vectors are combined to form a single vector of size 1×1024. The resulting combined feature vector is passed through a series of dense layers where batch normalization is applied before each layer, passing through three layers of dense layers of size 256, 128, and 64 respectively with ReLU activation function. Finally, the logits are computed in the output layer, which is a dense layer with 2 nodes. The softmax function is used to classify the input images to check whether they are true or false.

4 Experiments and Results

This section discusses the experiments performed to detect lying and the outcome. In the first section, the author explains the details of the experiments mentioned in this paper in terms of the setup as well as the data preparation procedures. The next part is the effects of assessing the training model's effectiveness and reviewing the model's outcome. Then, the author considers how effective and superior our proposed model for identifying lying behavior is by comparing it with other deception-detection models.

4.1 Experimental Setup

Following the preprocessed datasets, the frames were extracted at frequency duration/30, and the video's resolution was 640×360. Our experimental programs were implemented using Python 3 and Anaconda 3, an environment specially tailored for data analysis and ML with the help of tools such as sci-kit-learn, matplotlib, and Dlib. We had to use Kaggle's GPU to run DL models for LD.

4.2 Result Analysis

This section introduces the results of several experiments and presents the proposed model. However, it begins by providing a concise overview of the various measurements used to evaluate the performance of these models. Following usual practice, performance measures like accuracy, precision, recall, F1 score, and AUC ROC scores were used to assess the model's ability to detect deception. This approach also enables us to evaluate not only the general accuracy of the model but also helps to understand whether the model accurately predicts deceptive instances and minimizes false positive or false negative results. Thus, performance metrics in this case indicate our goal of fine-tuning the deception detection model to promote better deception detection practices in real-world scenarios.

A careful data splitting technique is employed to ensure the stability and effectiveness of the proposed hybrid deep learning model, allocating 80% of the dataset for training and 20% for testing. The splitting technique facilitated a comprehensive assessment of our model's performance, enabling it to learn from most of the data while undergoing an extensive study on distinct, unseen sections. Several experiments were conducted on the dataset to compare their outcomes with those of existing studies. Several ML models, such as SVM, RF, KNN, DT, and MLP, have been utilized to demonstrate superior performance with our dataset and provide a comparison with existing studies. We trained our dataset by implementing the ML models, with feature extraction performed using Dlib. The findings indicated that RF and KNN outperformed other models in accuracy. Specifically, RF attained an accuracy of 84.00%, while KNN reached an accuracy of 80.00%, as presented in Table 2. This comparison highlights the significance of the study.

Table 2: Comparison of ML model accuracy: existing studies *vs*. our study

	Authors	Datasets	Visual features extraction	Algorithm	Accuracy
Existing study	Pérez-Rosas et al. [45]	121 videos	MUMIN scheme	DT	68.59%
	Jaiswal et al. [47]	121 videos	OpenFace	SVM	67.20%
	Su et al. [22]	324 videos	Pittpatt	RF	76.92%
	Rill-Garcia et al. [19]	121 videos	OpenFace	Linear SVM	57.40%
		42 videos			60.00%
	Wu et al. [18]	121 videos	Micro expressions	DT	72.69%
				RF	80.64%
				LR	73.98%
				K-SVM	75.40%
				L-SVM	75.02%
	Gupta et al. [43]	201 videos		LBP+MLP	54.22%
	Şen et al. [48]	121 videos	FACS	SVM	53.67%
				RF	61.58%
				NN	57.63%
	Karnati et al. [1]	201 videos	LBP	SVM	60.00%
				RF	61.00%
				MLP	60.00%
		325 videos		SVM	55.00%
				RF	73.00%
				MLP	55.00%
Our study		700 videos	Dlib	SVM	65.00%
				RF	84.00%
				KNN	80.00%
				DT	76.00%
				MLP	59.00%

As RF showed better performance among all other ML models, we have verified the significant difference in RF between our study *vs*. existing studies, as shown in Fig. 6. The mean accuracy determined from the existing studies was 68.26%, with a standard deviation of 8.17%. A one-sample *t*-test is conducted to identify the significant difference between existing studies (mean accuracy) and ours. The *t*-test revealed the statistic *t*-statistic of −3.85, and the corresponding *p*-value equals 0.03. Since the *p*-value is less than the commonly used significance level of 0.05, the accuracy estimates of 84% achieved in our study are statistically significantly different from the mean accuracy of the other studies at the 5% level of statistical significance.

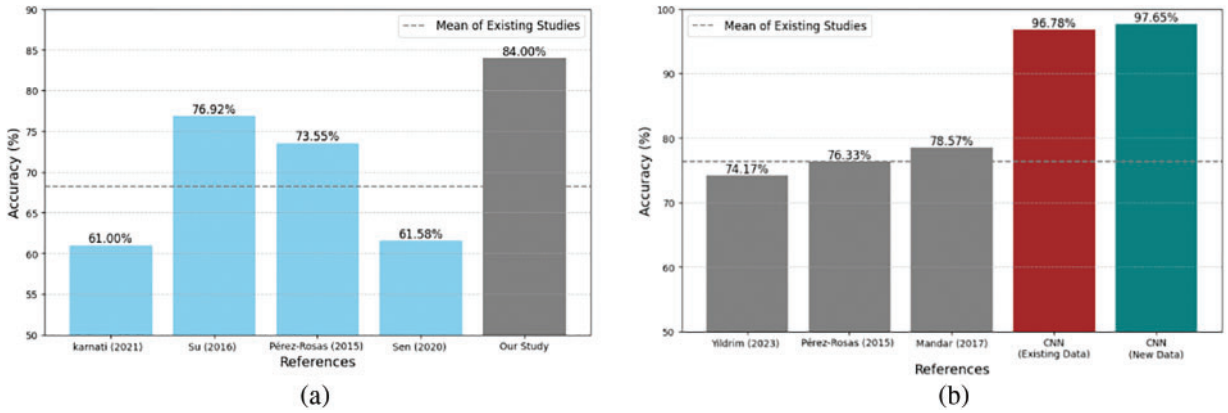


Figure 6: Comparison of accuracies for both RF and CNN models. (a) Comparison of RF accuracies-existing studies (Karnati et al. [1]; Su et al. [22]; Pérez-Rosas et al. [45]; Sen et al. [48]) *vs*. our study. (b) Comparison of CNN accuracies-existing studies (Yildrim et al. [28]; Pérez-Rosas et al. [45]; Gogate et al. [52]) *vs*. our study

This study also employed numerous advanced DL models, mainly showcasing the advantages of utilizing deep networks for image classification. The findings of different DL models used in prior research are presented in Table 3. Additionally, the outcome of our modified DL models includes 121 videos from existing datasets and 700 videos from our dataset. By analyzing the current research, it is noticeable that most studies utilized CNN and LSTM techniques. Only a small number of studies introduced customized models, and very few studies proposed any hybrid models. However, the performance of previous studies can be improved further. This drives us to implement a hybrid model by utilizing advanced optimal methods for the image classification task. It focuses on extracting facial action points and introduces a hybrid model that combines the Vgg16 and DenseNet121 architectures well-suited for image datasets. The findings outperform previous models and provide a new data set that fills current research gaps.

Table 3: Comparison of DL model accuracy: existing studies *vs*. our study

	Reference	Datasets	Model	Accuracy
Existing study	Pérez-Rosas et al. [45]	121 videos	CNN	76.33%
	Dhabarde et al. [49]	2538 facial images	CNN+SVM	58.28%
	Hasan et al. [50]	121 videos	WFM	70.00%

(Continued)

Table 3 (continued)

	Reference	Datasets	Model	Accuracy
	Karimi et al. [51]	121 videos	Dev-visual	75.00%
	Gogate et al. [52]	121 videos	CNN	78.57%
	Ajibade et al. [53]	121 videos	CNN+BiLSTM	61.00%
	Ahmed et al. [54]	121 videos	FACS with LSTM	89.49%
	Yildrim et al. [28]	201 videos	CNN	74.17%
	Karnati et al. [1]	Set A (201 videos)	LieNet	90.70%
		Set B (325 videos)		96.00%
	Rill-	121 videos	LSTM	56.00%
	Gracia et al. [19]	42 videos	LSTM	38.40%
Our study		Existing dataset (121 videos) [48]	ResNet-50	82.16%
			DenseNet121	95.08%
			Vgg16	89.47%
			CNN	96.78%
			Inception	95.91%
			LDNet	95.66%
		Our dataset (700 videos)	ResNet-50	93.04%
			DenseNet121	96.48%
			Vgg16	96.43%
			CNN	97.65%
			Inception	94.13%
			LDNet	99.50%

According to Table 3, ResNet50 achieved 82.16% accuracy, DenseNet121 achieved 97.08% accuracy, Vgg16 achieved 89.47% accuracy, CNN achieved 96.78% accuracy, and Inception achieved 95.91% accuracy by utilizing existing datasets. Therefore, our dataset and model provide better results than existing ones. In contrast, we utilized the same models with the dataset we generated. By utilizing our dataset instead of other available datasets, we have observed that models such as ResNet50, DenseNet121, Vgg16, and CNN provide superior outcomes, with RestNet50 achieved an accuracy of 93.04%, DenseNet121 achieved 96.48%, Vgg16 achieved 96.43%, and CNN achieved 97.65%. These findings showcase the high quality and efficacy of our datasets. However, the CNN model demonstrated superior performance for both datasets. The CNN model attained an accuracy of 96.78% with the available data and 97.65% while utilizing our dataset. To assess the statistical significance of the differences in accuracy between our CNN model and other CNN models, we employ Welch's t-test. This is because there are variations in variance and sample size. The Welch's t-test yielded a result of $t = -15.54$ and $p = 0.0017$, indicating the presence of a statistically significant difference. Given that the p-value is less than 0.05, this outcome suggests that our CNN model exhibits a statistically significant improvement over the existing models, as illustrated in Fig. 6.

Therefore, this study introduced the hybrid model using Vgg16 and DenseNet121, which were chosen for their simplicity and effectiveness. This combination exhibited superior performance compared to other model combinations. The outcomes derived from the Vgg16 network are presented in Table 4. The table additionally displays the results of a separate experiment in which DenseNet121

was employed to classify truth and lie. In addition, the outcomes obtained from our hybrid model, which combines Vgg16 and DenseNet121, have been included in Table 4, showcasing the noteworthy effectiveness of this methodology. The confusion matrices of these three models are also shown in Fig. 7. Our LDNet model achieved an accuracy of 95.66% (with the existing dataset) and 99.50% (with our datasets). Additionally, we evaluate LDNet with the existing WFM [53] and Dev visual models [51]. According to the *t*-test, the LDNet outperforms the WFM (70%) and Dev visual models. The *t*-statistic analysis reveals that these values are remarkably high, indicating the improved validity and accuracy of the mentioned LDNet model compared to the other models.

Table 4: Performance metrics of Vgg16, DenseNet121 and LDNet models

Vgg16				
Precision	**Recall**	**F1 score**	**Accuracy**	**ROC AUC score**
0.9532	0.9863	0.9707	0.9643	0.9912
DenseNet121				
Precision	**Recall**	**F1 score**	**Accuracy**	**ROC AUC score**
0.9477	0.9931	0.0.9698	0.9648	0.9963
LDNet				
Precision	**Recall**	**F1 score**	**Accuracy**	**ROC AUC score**
0.9923	0.9923	0.9885	0.9950	0.9892

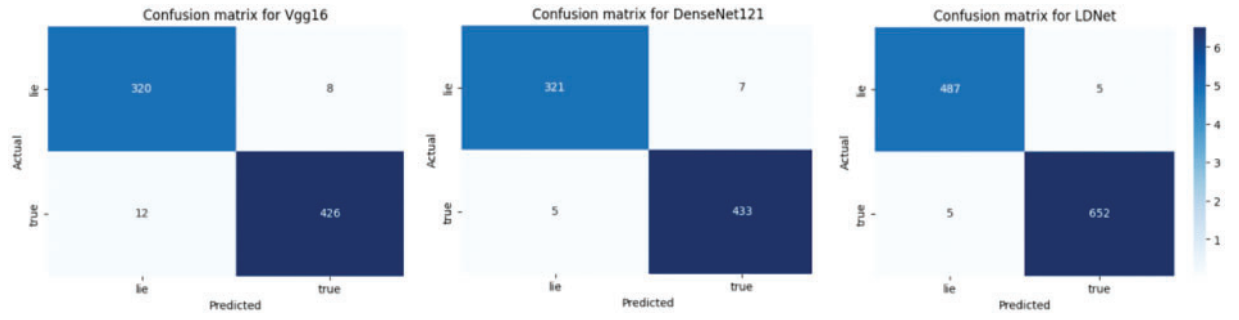


Figure 7: Confusion matrices for different models showcase their effectiveness in binary classification tasks. Each matrix visualizes the number of true positives, true negatives, false positives, and false negatives predicted by the respective model

4.3 Model Testing

After implementing the model, we also conducted the testing phase. We have tested our model in real-time to ensure its accuracy. We performed our model, uploaded a new video, and ran the model. Our model calculates the percentage of truth and lies during the testing phase, as shown in Fig. 8. After adding the input, frames from the video are retrieved and used to detect whether the person is lying. Dlib and FaceMesh are utilized to recognize face landmarks, which are employed as features in the trained model. The prediction, i.e., whether a specific subject in the video is lying or telling the truth, is based on the action with the highest probability. It also shows the prediction result in an

overlay window over the video with the headline "Lie Detection." The prob_viz function displays the likelihood of prediction on the frames of the video being analyzed in real-time. The loop to update and show frames continues until the "q" key is pushed, at this point, the video capture stops, and all windows are closed. The result can be true or false depending on the frame being evaluated. Based on the Facial Landmarks collected from the video stream, the suggested model differentiates participants' true and fraudulent behavior.

Figure 8: Real-time model testing using video input for lie detection

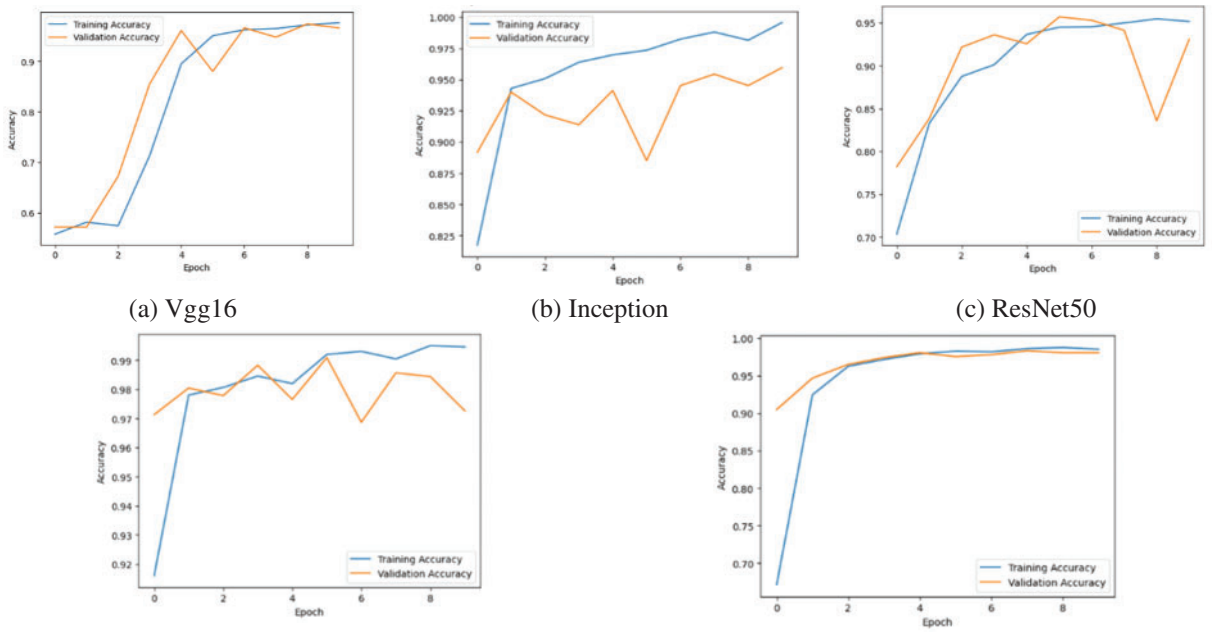
5 Discussions and Limitations

Lying has become apparent in all branches of human activity, there are great opportunities to study how to find lies and create a complex system of countermeasures against this behavior. The areas where our model can be used include Benchmark security, which is beneficial for interrogation analysis; security and law enforcement, where the model is handy for forensic psychology for criminal investigation and applicable for any recruitment process where users are prone to tell lies. Also, many organizations nowadays hire private forensic teams, and the LD model is employed to investigate fraudulent or harassment activities.

We introduced a new dataset consisting of 900 video clips and including 80 participants, with 24 being male and 56 being female. The participants were selected explicitly as young individuals aged between 18 and 40. The project's objective was to develop a versatile and accessible resource for academics interested in conducting studies that involve carefully regulated claims of truth and lies. We do not expect our dataset to be a complete replacement for all current LD stimuli. However, the qualities of this dataset provide researchers with a more comprehensive and substantial dataset of higher quality. We are confident this will generate new study prospects for individuals studying LD, intergroup connections, and social perception on a broader scale.

The wide range of ML and DL models, including CNN, ResNet50, DenseNet121, Vgg16, and Inception for DL, as well as SVM, RF, KNN, DT, and MLP for ML, allows for a thorough exploration of different methodologies and approaches, allowing for a more comprehensive evaluation and comparison of their effectiveness in various applications. We only compared the accuracy levels of different bagging models on several datasets. Through vigorous testing of the other datasets, we could identify the strengths and weaknesses of the various models. Encouraged by the results derived from ML and DL models, we introduce LDNet, a new ML architecture derived from both paradigms. LDNet uses feature extraction from ML algorithms, combining Vgg16 and DenseNet121 models in

its DL design. By applying a combination of DL approaches as utilized in the LDNet framework, the model attains higher accuracy than the individual models. Fig. 9 compares the testing and training accuracy based on the number of epochs.



(a) Vgg16 (b) Inception (c) ResNet50

(d) DenseNet121 (e) CNN

Figure 9: Training *vs*. testing accuracy over epochs

The current study has several limitations. Initially, the author focused on single modalities, such as facial expressions, to investigate deception detection. However, using several modalities, such as audio and text, will enhance effectiveness and provide further insights into identifying fraudulent behaviors. Audio exclusion in this study was primarily due to the focus on facial expressions as a primary sign of deception. Including audio in this study would have presented numerous challenges and required complex computational processing beyond the study's scope. A concentrated focus on the specific nonverbal indicator, such as facial expressions, in this case, yielded a more profound understanding of the effectiveness of this method for detecting deception.

Additionally, body posture, hand gestures, and head motions will improve the video features. By just concentrating on facial expressions, we developed a more controllable model within the limitations of our study. Furthermore, while implementing our model in real-world scenarios, we will encounter problems, such as dealing with various facial reactions. Facial expressions vary from person to person when it comes to both truthful and deceptive answers. This variability will impose limitations.

Moreover, the model might need help with cases where facial expressions are subtle or ambiguous, making distinguishing between truth and lies difficult. If the input images are of low quality, blurred, or obstructed (e.g., partially covered faces), the feature extraction may be compromised, leading to inaccurate predictions. By incorporating more diverse datasets with other modalities, the situation can be improved to make the model work in various settings.

6 Conclusion and Future Works

With the rising prevalence of lying in all fields, there are many opportunities to study LD and develop comprehensive systems to deal with this problem. Artificial intelligence and other advanced techniques present an outstanding possibility to solve a variety of challenging scenarios, including LD. When compared to people, machines produce less biased and more accurate results. Another drawback of the current research is the scarcity of publicly accessible datasets, particularly video datasets. To address the challenge of the inadequate quantity of existing datasets, we have first taken the necessary steps and created the datasets in this work. Then, utilizing our dataset and existing datasets, we focused on the facial expressions and experimented using various advanced techniques, such as ML and DL. A hybrid model named LDNet was developed by combining layers from two models, such as Vgg16 and DenseNet121. Compared to other models, LDNet provides the best performance, which is a significant improvement for identifying deceptive behaviors.

However, it is crucial to acknowledge the variability of facial expressions among individuals, which makes it challenging to achieve the utmost accuracy solely based on facial expressions. Therefore, our future research will focus on enhancing accuracy by including an increasing range of gestures and facial expressions, such as head position, hand posture, pulse rate, etc., to uncover even more information. These enhanced factors will support the current framework and facilitate the effective exploration of various concepts. In addition, future studies will incorporate the usage of multimodalities. The present study exclusively assessed individual modalities, such as videos. In addition, this region will include audio and text, offering supplementary information to enhance the efficacy of deception detection models. Utilizing several modalities, rather than relying on a single modality, improves understanding of the situation and adds value. We also plan to extend our study by employing diverse datasets and real-time testing in various environments. Additionally, we intend to incorporate explainable techniques such as LIME and SHAP in future iterations of our work. These methods will allow us to offer more precious insights into the model's decision-making process, improving transparency and user confidence.

Acknowledgement: We thank our families and friends who provided us with moral support and Taylor's University, who funded this project.

Funding Statement: This publication is funded by the Ministry of Higher Education (MOHE), Malaysia under the Fundamental Research Grant Project (FRGS/1/2021/SS0/TAYLOR/02/6). This grant was received by the Faculty of Hospitality, Food, and Leisure Management at Taylor's University, Malaysia.

Author Contributions: The authors confirm their contribution to the paper as follows: study conception and design: Shanjita Akter Prome, Md Rafiqul Islam; data collection: Shanjita Akter Prome, Md. Kowsar Hossain Sakib; implementation and analysis of results: Shanjita Akter Prome, Md. Kowsar Hossain Sakib; draft manuscript preparation: Shanjita Akter Prome; review: Md Rafiqul Islam, David Asirvatham, Neethiahnanthan Ari Ragavan, Cesar Sanin, Edward Szczerbicki. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are available from the corresponding author, Md Rafiqul Islam, upon reasonable request.

Ethics Approval: This study involved human subjects and was conducted by ethical standards. Taylor's University, Malaysia's Human Ethics Committee (HEC) reviewed and approved the study protocol. The ethics approval reference number is HEC 2023/021.

Conflicts of Interest: The authors declare that there are no conflicts of interest to report regarding the present study.

References

[1] M. Karnati, A. Seal, A. Yazidi, and O. Krejcar, "LieNet: A deep convolution neural network framework for detecting deception," *IEEE Trans. Cogn. Dev. Syst.*, vol. 14, no. 3, pp. 971–984, 2021. doi: 10.1109/TCDS.2021.3086011.

[2] S. V. Fernandes and M. S. Ullah, "A comprehensive review on features extraction and features matching techniques for deception detection," *IEEE Access*, vol. 10, no. 17, pp. 28233–28246, 2022. doi: 10.1109/ACCESS.2022.3157821.

[3] Z. Liu, S. Shabani, N. G. Balet, and M. Sokhn, "Detection of satiric news on social media: Analysis of the phenomenon with a French dataset," presented at the 2019 28th Int. Conf. Comput. Commun. Netw. (ICCCN), Valencia, Spain, 2019, pp. 1–6. doi: 10.1109/ICCCN.2019.8847041.

[4] G. Krishnamurthy, N. Majumder, S. Poria, and E. Cambria, "A deep learning approach for multimodal deception detection," presented at the Int. Conf. Comput. Linguist. Intell. Text Process., Hanoi, Vietnam, Springer, Mar. 18–24, 2018, pp. 87–96. doi: 10.1007/978-3-031-23793-5_8.

[5] S. A. Prome, N. A. Ragavan, M. R. Islam, D. Asirvatham, and A. J. Jegathesan, "Deception detection using ml and dl techniques: A systematic review," *Nat. Lang. Process. J.*, vol. 6, 2024, Art. no. 100057. doi: 10.1016/j.nlp.2024.100057.

[6] A. Jan, H. Meng, Y. F. B. A. Gaus, and F. Zhang, "Artificial intelligent system for automatic depression level analysis through visual and vocal expressions," *IEEE Trans. Cogn. Dev. Syst.*, vol. 10, no. 3, pp. 668–680, 2017. doi: 10.1109/TCDS.2017.2721552.

[7] K. Tsuchiya, R. Hatano, and H. Nishiyama, "Detecting deception using machine learning with facial expressions and pulse rate," *Artif. Life Robot.*, pp. 1–11, 2023. doi: 10.1007/s10015-023-00869-9.

[8] W. Castillo *et al.*, "Techniques for enhanced image capture using a computer-vision network," *US Patent App.*, vol. 17/163, p. 043, Aug. 5, 2021.

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

[10] W. Khan, K. Crockett, J. O'Shea, A. Hussain, and B. M. Khan, "Deception in the eyes of deceiver: A computer vision and machine learning based automated deception detection," *Expert. Syst. Appl.*, vol. 169, 2021, Art. no. 114341. doi: 10.1016/j.eswa.2020.114341.

[11] S. A. Prome, M. R. Islam, D. Asirvatham, M. K. H. Sakib, and N. A. Ragavan, "LieVis: A visual interactive dashboard for lie detection using machine learning and deep learning techniques," presented at the 2023 26th Int. Conf. Comput. Inf. Technol. (ICCIT), Bangladesh, 2023, pp. 1–6. doi: 10.1109/ICCIT60459.2023.10441173.

[12] M. K. Hossain Sakib *et al.*, "*MVis4LD*: Multimodal visual interactive system for lie detection," presented at the Asian Conf. Intell. Inf. Database Syst., United Arab Emirates, 2024, pp. 28–43. doi: 10.1007/978-981-97-4985-0_3.

[13] M. Ding, A. Zhao, Z. Lu, T. Xiang, and J. -R. Wen, "Face-focused cross-stream network for deception detection in videos," presented at the IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Long Beach, AK, USA, 2019, pp. 7802–7811.

[14] Y. Xie, R. Liang, H. Tao, Y. Zhu, and L. Zhao, "Convolutional bidirectional long short-term memory for deception detection with acoustic features," *IEEE Access*, vol. 6, pp. 76527–76534, 2018. doi: 10.1109/ACCESS.2018.2882917.

[15] M. Kanmani and V. Narasimhan, "Optimal fusion aided face recognition from visible and thermal face images," *Multimed. Tools Appl.*, vol. 79, pp. 17859–17883, 2020. doi: 10.1007/s11042-020-08628-9.
[16] J. Alzubi, A. Nayyar, and A. Kumar, "Machine learning from theory to algorithms: An overview," *J. Phys.: Conf. Series*, vol. 1142, 2018, Art. no. 012012. doi: 10.1088/1742-6596/1142/1/012012.
[17] P. K. Sehrawat, R. Kumar, N. Kumar, and D. K. Vishwakarma, "Deception detection using a multimodal stacked Bi-LSTM model," presented at the 2023 Int. Conf. Innov. Data Commun. Technol. App. (ICIDCA), Uttarakhand, India, 2023, pp. 318–326. doi: 10.1109/ICIDCA56705.2023.10099779.
[18] Z. Wu, B. Singh, L. Davis, and V. Subrahmanian, "Deception detection in videos," presented at the AAAI Conf. Artif. Intell., vol. 32, no. 1, 2018. doi: 10.1609/aaai.v32i1.11502.
[19] R. Rill-García, H. Jair Escalante, L. Villasenor-Pineda, and V. Reyes-Meza, "High-level features for multimodal deception detection in videos," presented at the IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Long Beach, AK, USA, 2019. doi: 10.1109/CVPRW.2019.00198.
[20] M. A. Khalil, J. Can, and K. George, "Deep learning applications in brain computer interface based lie detection," presented at the 2023 IEEE 13th Annu. Comput. Commun. Workshop Conf. (CCWC), Las Vegas, NV, USA, 2023, pp. 189–192.
[21] M. Kawulok, J. Nalepa, K. Nurzynska, and B. Smolka, "In search of truth: Analysis of smile intensity dynamics to detect deception," presented at the 15th Ibero-Am. Conf. AI, San José, Costa Rica, Nov. 23–25, 2016, 325–337. doi: 10.1007/978-3-319-47955-2_27.
[22] L. Su and M. Levine, "Does "Lie to me lie to you? an evaluation of facial clues to high-stakes deception," *Comput. Vis. Image Underst.*, vol. 147, no. 2, pp. 52–68, 2016. doi: 10.1016/j.cviu.2016.01.009.
[23] M. Zloteanu, "The role of emotions in detecting deception," in *Deception: An Interdisciplinary Exploration*, Oxford, Inter-Disciplinary Press, 2015, pp. 203–217.
[24] 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, 2023, Art. no. e0281323. doi: 10.1371/journal.pone.0281323.
[25] M. G. Frank and E. Svetieva, "Microexpressions and deception," in *Understand. Facial Express. Commun.*, Springer: New Delhi, 2015, pp. 227–242.
[26] X. Ben *et al.*, "Video-based facial micro-expression analysis: A survey of datasets, features and algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 9, pp. 5826–5846, 2021. doi: 10.1109/TPAMI.2021.3067464.
[27] D. Y. Choi, D. H. Kim, and B. C. Song, "Recognizing fine facial micro-expressions using two-dimensional landmark feature," presented at the 2018 25th IEEE Int. Conf. Image Process. (ICIP), Athens, Greece, 2018, pp. 1962–1966. doi: 10.1109/ICIP.2018.8451359.
[28] S. Yildirim, M. S. Chimeumanu, and Z. A. Rana, "The influence of micro-expressions on deception detection," *Multimed. Tools Appl.*, vol. 82, no. 19, pp. 29115–29133, 2023. doi: 10.1007/s11042-023-14551-6.
[29] M. Zloteanu, "Reconsidering facial expressions and deception detection," in *Handbook of Facial Expression of Emotion*, FEELab Science Books & Leya, 2020, vol. 3, pp. 238–284.
[30] S. Jordan, L. Brimbal, D. B. Wallace, S. M. Kassin, M. Hartwig and C. N. Street, "A test of the micro-expressions training tool: Does it improve lie detection?" *J. Investig. Psychol. Offender Profiling*, vol. 16, no. 3, pp. 222–235, 2019. doi: 10.1002/jip.1532.
[31] M. Zloteanu, "Emotion recognition and deception detection," 2019. doi: 10.31234/osf.io/crzne.
[32] R. Zhi, M. Liu, and D. Zhang, "A comprehensive survey on automatic facial action unit analysis," *Vis. Comput.*, vol. 36, no. 5, pp. 1067–1093, 2020. doi: 10.1007/s00371-019-01707-5.
[33] P. Werner, S. Handrich, and A. Al-Hamadi, "Facial action unit intensity estimation and feature relevance visualization with random regression forests," presented at the 2017 Seventh Int. Conf. Affect. Comput. Intell. Interact. (ACII), San Antonio, TX, USA, 2017, pp. 401–406. doi: 10.1109/ACII.2017.8273631.
[34] D. Vinkemeier, M. Valstar, and J. Gratch, "Predicting folds in poker using action unit detectors and decision trees," presented at the 2018 13th IEEE Int. Conf. Automatic Face Gesture Recognit. (FG 2018), Xi'an, China, 2018, pp. 504–511. doi: 10.1109/FG.2018.00081.

[35] W. -Y. Chang, S. -H. Hsu, and J. -H. Chien, "FATAUVA-Net: An integrated deep learning framework for facial attribute recognition, action unit detection, and valence-arousal estimation," presented at the IEEE Conf. Comput. Vis. Pattern Recognit. Workshops, Honolulu, HI, USA, 2017, pp. 17–25. doi: 10.1109/CVPRW.2017.246.

[36] P. Khorrami, T. Paine, and T. Huang, "Do deep neural networks learn facial action units when doing expression recognition?" presented at the IEEE Int. Conf. Comput. Vis. Workshops, 2015, pp. 19–27. doi: 10.1109/ICCVW.2015.12.

[37] S. Agarwal, H. Farid, O. Fried, and M. Agrawala, "Detecting deep-fake videos from phoneme-viseme mismatches," presented at the IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops, Seattle, WA, USA, 2020, pp. 660–661. doi: 10.1109/CVPRW50498.2020.00338.

[38] G. Pons and D. Masip, "Supervised committee of convolutional neural networks in automated facial expression analysis," *IEEE Trans. Affect. Comput.*, vol. 9, no. 3, pp. 343–350, 2017. doi: 10.1109/TAFFC.2017.2753235.

[39] M. Serras Pereira, R. Cozijn, E. Postma, S. Shahid, and M. Swerts, "Comparing a perceptual and an automated vision-based method for lie detection in younger children," *Front. Psychol.*, vol. 7, 2016, Art. no. 223592. doi: 10.3389/fpsyg.2016.01936.

[40] B. Singh, P. Rajiv, and M. Chandra, "Lie detection using image processing," presented at the 2015 Int. Conf. Adv. Comput. Commun. Syst., Coimbatore, India, 2015, pp. 1–5. doi: 10.1109/ICACCS.2015.7324092.

[41] S. Satpathi, K. M. I. Y. Arafath, A. Routray, and P. S. Satpathi, "Analysis of thermal videos for detection of lie during interrogation," *EURASIP J. Image Video Process.*, vol. 2024, no. 1, 2024, Art. no. 9. doi: 10.1186/s13640-024-00624-5.

[42] K. J. Feng, "Deeplie: Detect lies with facial expression (computer vision)," 2023. Accessed: Sep. 29, 2024. [Online]. Available: https://cs230.stanford.edu/projects_spring_2021/reports/0.pdf.

[43] 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*, Long Beach, AK, USA, 2019. doi: 10.1109/CVPRW.2019.00016.

[44] F. Soldner, V. Pérez-Rosas, and R. Mihalcea, "Box of lies: Multimodal deception detection in dialogues," presented at the 2019 Conf. North Am. Chap. Assoc. Comput. Linguist.: Human Lang. Technol., Minneapolis, MN, USA, 2019, vol. 1, pp. 1768–1777.

[45] V. Pérez-Rosas, M. Abouelenien, R. Mihalcea, and M. Burzo, "Deception detection using real-life trial data," presented at the 2015 ACM Int. Conf. Multimod. Interact., Seattle, WA, USA, 2015, pp. 59–66. doi: 10.1145/2818346.28207.

[46] F. Abdulridha and B. M. Albaker, "Non-invasive real-time multimodal deception detection using machine learning and parallel computing techniques," *Soc. Netw. Anal. Min.*, vol. 14, no. 1, 2024, Art. no. 97. doi: 10.1007/s13278-024-01255-4.

[47] M. Jaiswal, S. Tabibu, and R. Bajpai, "The truth and nothing but the truth: Multimodal analysis for deception detection," presented at the 2016 IEEE 16th Int. Conf. Data Min. Workshops (ICDMW), Barcelona, Spain, 2016, pp. 938–943. doi: 10.1109/ICDMW.2016.0137.

[48] M. U. Şen, V. Perez-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, 2020. doi: 10.1109/TAFFC.2020.3015684.

[49] R. J. Dhabarde, D. Kodawade, and S. Zalte, "Hybrid machine learning model for lie-detection," presented at the 2023 IEEE 8th Int. Conf. Converg. Technol. (I2CT), Lonavla, India, 2023, pp. 1–5. doi: 10.1109/I2CT57861.2023.10126460.

[50] K. Hasan *et al.*, "Facial expression based imagination index and a transfer learning approach to detect deception," presented at the 2019 8th Int. Conf. Affect. Comput. Intell. Interact. (ACII), Cambridge, UK, 2019, pp. 634–640. doi: 10.1109/ACII.2019.8925473.

[51] H. Karimi, "Interpretable multimodal deception detection in videos," presented at the 20th ACM Int. Conf. Multimod. Interact., USA, 2018, pp. 511–515. doi: 10.1145/3242969.326496.

[52] M. Gogate, A. Adeel, and A. Hussain, "Deep learning driven multimodal fusion for automated deception detection," presented at the 2017 IEEE Symp. Series Comput. Intell. (SSCI), Honolulu, HI, USA, 2017, pp. 1–6. doi: 10.1109/SSCI.2017.8285382.

[53] F. Ajibade and O. Akinola, "A model for identifying deceptive acts from non-verbal cues in visual video using bidirectional long short term memory (BiLSTM) with convolutional neural network (CNN) features," *Univ. Ibadan J. Sci. Logics ICT Res.*, vol. 7, no. 1, pp. 39–47, 2021.

[54] H. U. D. Ahmed, U. I. Bajwa, F. Zhang, and M. W. Anwar, "Deception detection in videos using the facial action coding system," 2021, *arXiv:2105.13659*.