Journal of Innovative Image Processing (ISSN: 2582-4252)
www.irojournals.com/iroiip/

Multimodal Lie Detection Using Linguistic and Visual Cues: A Fusion of NLP and Facial Micro-Feature Analysis

Twisha Patel[1], Daxa Vekariya[2]

[1]Faculty of Engineering and Technology, Parul University, Waghodia, Vadodara, Gujarat, India.
[2]Department of Computer Science and Engineering, Parul Institute of Engineering and Technology, Faculty of Engineering and Technology, Parul University, Waghodia, Vadodara, Gujarat, India.

E-mail: [1]twisapatel@gmail.com, [2]daxa.vekariya18436@paruluniversity.ac.in

Abstract

The research integrates Natural Language Processing (NLP) and facial micro-expressions recognition methods for analyzing deceptive behavior. Lie behavior analysis is enhanced by the incorporation of both verbal and non-verbal communication in the assessment as subtle non-verbal cues are hard to detect during scrutiny. Different machine learning algorithms were evaluated based on their ability to detect lies in this study. Several classic models like Nearest Neighbors, Linear SVM, Decision Tree, Random Forest and Extra Trees Classifier were tested using the Real-Life Deception Detection and Own Dataset student viva scenario data. Various accuracies were generated by different traditional ML models until researchers developed a lightweight Convolutional Neural Network (CNN) model designed to efficiently detect deception. The lite-CNN model achieved a successful 96% accuracy in both tests on the dataset. The lite-CNN model identifies deceptions through its high performance by combining verbal speech and facial behavioral patterns. It has been found that deception detection is successful when using NLP with facial expressions providing reasonable solutions in the fields of security, psychology, and human-computer interaction. The proposed lightweight CNN model is a proven solution compared to traditional models, as it is effective yet consumes fewer computing resources.

Keywords: Lie Detection; Natural Language Processing; Facial Micro-Features; Machine Learning; Convolutional Neural Network.

1. Introduction

The research on deception detection has been on the agenda of researchers in the fields of psychology and artificial intelligence for decades. The methods of detecting deception traditionally rely on three approaches: human intuition, polygraph testing, and the observation of behavioral signs. These approaches often provide unstable and inaccurate outcomes. Scholars have tried to improve deception detection by using AI and machine learning technologies. Natural Language Processing (NLP) and facial micro-expression analysis show promise as effective approaches to deception detection due to their capability to analyze both verbal and non-verbal indicators of behavior. Studies performed in isolation do not achieve optimal effectiveness. The combination of language analysis and monitoring of facial

DOI: https://doi.org/10.36548/jiip.2025.4.016
Received: 04.10.2025, received in revised form: 03.11.2025, accepted: 17.11.2025, published: 25.11.2025

expressions builds a complete system of deception assessment that provides greater accuracy in security monitoring, psychological studies, and human-computer interaction.

Advances in machine learning to address the problem of deception have yielded significant results, but scholars still have to address the research gaps. Research on deception detection has primarily investigated verbal or non-verbal indicators independently, ignoring the modalities that reinforce one another. Different conventional machine learning models, such as Nearest Neighbors, Support Vector Machines (SVM), Decision Trees, and Random Forests, have been heavily used, but their efficacy varies with different data sets. Deep learning models have substantial computational demands, posing challenges to security and forensic investigations when using real-time applications. The present research aims to address the important research gap due to the lack of an efficient combined NLP and facial micro-expression analytical model that is lightweight in terms of performance.

The primary objective of the research is to develop a thin Convolutional Neural Network (CNN) model that relates non-verbal and verbal cues to increase the detection of deception. The project combines NLP analysis of text and speech data with facial micro-expression analysis to achieve more accurate and effective methods of detecting deception. The paper studies classical machine learning algorithms, including Nearest Neighbors, Linear SVM, Decision Tree, Random Forest, and Extra Trees Classifier, which can be applied to both the Real-Life Deception Detection (2016) and original student viva datasets. These models are used in the research to investigate the effectiveness of the proposed lite-CNN model. The research aims to demonstrate that a proper deception detection tool with minimal computational needs can be developed through a combination of facial and linguistic data analysis algorithms. This research shows that conventional machine learning approaches attain various rates of deception detection but fail to integrate various data sources successfully. The lightweight CNN model proposed provides excellent performance, attaining 96% accuracy in the analysis of the two datasets. This great success clearly demonstrates that the model effectively handles complicated patterns in both verbal communication data and facial expressions, with simplified computation requirements. The integrated system of NLP and micro-expression analysis is an effective method for increasing the reliability of deception detection, making it an efficient means for security services, as well as for forensic psychology and human-computer interaction functions. This research demonstrates that a smaller, optimized deep learning model has better performance than traditional methods and offers a scalable system to detect deception in real time.

2. Related Work

The introduction of artificial intelligence (AI) and machine learning (ML) technologies has introduced significant advancements to the possibilities of deception detection. Various studies have been carried out by researchers to achieve a higher degree of accuracy by conducting research on facial micro-expression analysis coupled with both vocal and biological indicators. Scientists have enhanced the efficiency of detection by using deep learning techniques along with feature fusion techniques as well as multimodal fusion methods. Several challenges need to be addressed, including limited datasets and problems with effective computation and real-time performance.

Wang et al. [1] investigated a meta-learning architecture for cross-database micro-expression recognition and proved that meta-learning can improve generalization when using

heterogeneous sources of data; however, their experiments, do not incorporate supplementary behavioral metrics like verbal or linguistic expressions which can be used to complement deception-related analyses. Tseng and Cheng [2] analyzed AI-based lie detection through the prism of cognitive and psychological perspectives, claiming that most computational solutions have not been related to cognitive theories of deception and thus are truly ineffective in real-world scenarios, where it is critically important to understand mental processes and situational influences. Delmas et al. [3] conducted a review of research investigating automatic lie detection through facial features, concluding that despite deep learning approaches demonstrating high potential, existing systems have several flaws including a lack of adequate annotated data, a risk of overfitting, and a lack of ecological validity. This requires the mobilization of larger and more varied datasets to enhance the robustness of the models and their application. Nikbin and Qu [4] developed combination deep neural networks that achieved effective deception detection through micro-expression analysis. However, the architecture of their model created difficulties for real-time processing because it required high computational power. The technique presented by Satpathi et al. [5] combined thermal video analysis to identify deceptions through facial temperature changes which indicated stress levels and deception. The effectiveness of their detection system faced practical deployment barriers because it needed specialized thermal cameras. DUlizia et al. [6] examined multiple techniques for deception detection which combine facial expression analysis with speech analysis along with medical signal measurements. Future research should develop efficient yet lightweight deep learning approaches for real-time deception detection since researchers have identified technical challenges with multi-modal data fusion. King and Neal [7] performed a detailed examination of deception detection systems powered by AI which utilized video, audio, and physiological information. The researchers pointed out that present models display insufficient cross-cultural adaptability and cultural linguistic flexibility and they recommend developing flexible AI systems. A facial micro-expression detection system based on random Fourier features enabled neural networks for achieving accurate results according to Yadav et al. [8] their proposed system needed long and complex feature engineering processes, making it unsuitable for real-time automatic applications.

Voice stress analysis implemented by Talaat [9] serves as the basis for his explainable recurrent neural network (RNN) model which detects lies. The model achieved better interpretation, but its performance relied on clear high-quality audio which makes it exposed to environmental noise during practical usage. The research conducted by Dinges et al. [10] developed AI-based facial cue interpretation software for detecting deceptive behavior through an automated system. The research approach demonstrated promising outcomes; yet, it was shown to be weak against adversarial attacks because security-aware deception detection models remain necessary. Researchers Kumar Tataji et al. [11] established a facial expression recognition system with a Cross-Connected Convolutional Neural Network (CC-CNN) framework that employed feature-level fusion methods. The model successfully detected complex facial expressions, yet it failed to integrate verbal deception evidence thus preventing its full application in deception detection. The research by Ahmed Khan et al. [12] proved using the Facial Action Coding System (FACS) that deceptive movements of facial muscles exist in videos. Their model depended heavily on precise facial landmark tracking which proves unreliable when working with videos that have low resolution and when observing blocks of the face. Manalu and Rifai [13] developed an emotion detection system using CNN along with RNN which combined elements from both networks. The authors achieved improved emotion recognition results through their approach, but they did not specialize the method for deception detection work which leaves potential space for future developments in deception-specific

datasets and models. Researchers Cash et al. [14] studied how speaking with someone previously and repeating statements affected deception recognition success rates in their results. The research examined human detection methods instead of automated systems showing the necessity to develop AI models able to replicate these human cognitive processes. Talaat [15] designed an explainable recurrent neural network system for detecting stress through voices during deception assessments. The decision-making transparency of their model faced problems with background noise which made practical implementation difficult. The research by Dinges et al. [16] examined AI approaches in face cue detection for deception detection but emphasized the importance of resistance to adversarial attacks in deceptive model systems. Deep learning-based deception detection systems displayed weaknesses in detecting altered facial expressions as described in their research.

The researchers developed FMeAR which stands for Facial Micro-Expression Action Unit Recognition using the Facial Action Coding System (FACS). The model demonstrated strong performance in micro-expression recognition but failed to prove its capability for deception detection in real-world deception settings according to the researchers [17]. De Marsico et al. [18] created FTM as an extraction tool for micro-expression features that functioned to aid lie detection processes. The research success of FTM in controlled settings needed additional testing on expanding datasets before achieving widespread use in real-world conditions. The researchers Zhou and Bu [19] developed methods to combine bimodal features through domain adversarial neural networks for lie detection operations. A key advantage of their approach was its ability to combine facial expressions with speech features however the adversarial network reduced operational efficiency for real-time applications. Abdulridha and Albaker [20] designed an invasive deception detection system based on machine learning combined with parallel computing approaches. This method accelerated processing but failed to provide explanations about how decisions were made by the model. Preethi et al. [21] created a framework for enhancing micro-expression analysis within advanced multimedia systems that dealt with micro-facial recognition. The primary objective of their research focused on emotion detection but required additional changes to become suitable for deception detection purposes. The literature review by Sen and Deneckère [22] showed that established deception detection models experience difficulties with recognizing deceptive signals across different cultures and specific contexts based on their research findings. The authors indicated that the next step should involve developing learning models which adapt to process data from various sources. The research conducted by Li et al. [23] developed a video-based detection framework which used wrapper-based methods to enhance performance by filtering out unnecessary features. The feature selection method used by their research required extensive computations which rendered it impractical for real-time large-scale implementation. The authors at Stathopoulos et al. [24] developed an attention-feedback mechanism which made the deception detection system in videos pay attention to critical facial attributes. The researchers achieved promising outcomes, but their method needed specific adjustments for individual cases which reduced its broad application scope. Through their work Chebbi and Jebara [25] developed deception detection models by integrating different data types including human face movements and spoken words together with physiological signs. The detection system developed by the researchers produced better results, but the extensive preprocessing work hindered real-time deployment because of its complexity. The systematic review carried out by Constancio et al. [26] demonstrated that machine learning-based deception detection models show weak performance across different datasets. The authors highlighted that transfer learning methods should be implemented to enhance models' adaptive features. The review from D'Ulizia et al. [27] showed that facial cue-based deception detection faces challenges from unstandardized

datasets during AI-based identification. Model training will improve through collaborative efforts between groups for assembling datasets according to their guidelines. FacialCueNet constitutes an interpretable artificial intelligence model developed by Nam et al. [28] to detect deception in criminal interrogation procedures. The model achieved high performance levels in forensic applications yet needed further training to work outside of experimental conditions.

Deception detection research has achieved major advancements using facial micro-expression analysis together with NLP and multimodal fusion methods. Various obstacles like insufficient datasets and computational problems and a lack of useful application remain prominent. The development of quick and efficient AI models that process voice and body signals should remain the primary goal of research with emphasis also placed on attack-resilient systems. To enhance model performance across different linguistic and cultural scenarios researchers should investigate the implementation of both domain adaptation and transfer learning methods.

3. Proposed Work

Figure 1 The AI system framework combines multimodal video and audio data to engage in effective and accurate deception detection. The system operates in four key phases, namely (1) video-based micro-expression analysis, (2) audio-text processing, (3) feature-level fusion, and (4) lightweight CNN-based classification. The first step involves selecting an input video for the subsequent extraction of frames that serve to evaluate facial and gesture features while separating the audio for NLP-based textual processing. The D-Library specializes in extracting micro-facial features from video frames, but audio processing requires text transcripts, which NLP uses for TF-IDF-based feature analysis. Facial landmarks were detected using Dlib's 68-point predictor, and the pre-processing involved steps including histogram equalization and rejection of frames to reduce contamination caused by poorly illuminated images. A feature vector with micro-facial expressions, along with TF-IDF textual features, serves as the basis for combined attributes. The combined feature vector undergoes training and classification using multiple machine learning models, which include SVM, KNN, Decision Tree, Naïve Bayes, Random Forest, and Extra Trees. The classification accuracy is enhanced through the incorporation of the proposed Lite-CNN model. The trained model classifies the input data to determine whether the statements in the provided videos are truthful or lies.

3.1 Dataset Overview and Ethics Consideration

The Real-Life Deception [29] tested the strength against changes in distribution, the data were clustered depending on the time of recording (old or new trials). Camera resolution was 640x480, audio (MP3) and audio lighting conditions (Normal) were tested.

The own dataset [30] is a dataset developed in-house, comprising 60 participants-30 men and 30 women-between the ages of 20 and 25, balanced in terms of ethnic and demographic background. All subjects were put through five question-answer sessions under regulated lighting and camera positioning. Synchronized video at 25 fps, audio, and transcript data can be found. Ethical approval was provided by the institutional review board; each participant gave informed consent for the use of data in research. Additionally, demographic diversity was considered from a statistical point of view to ensure fairness among gender and ethnicity.

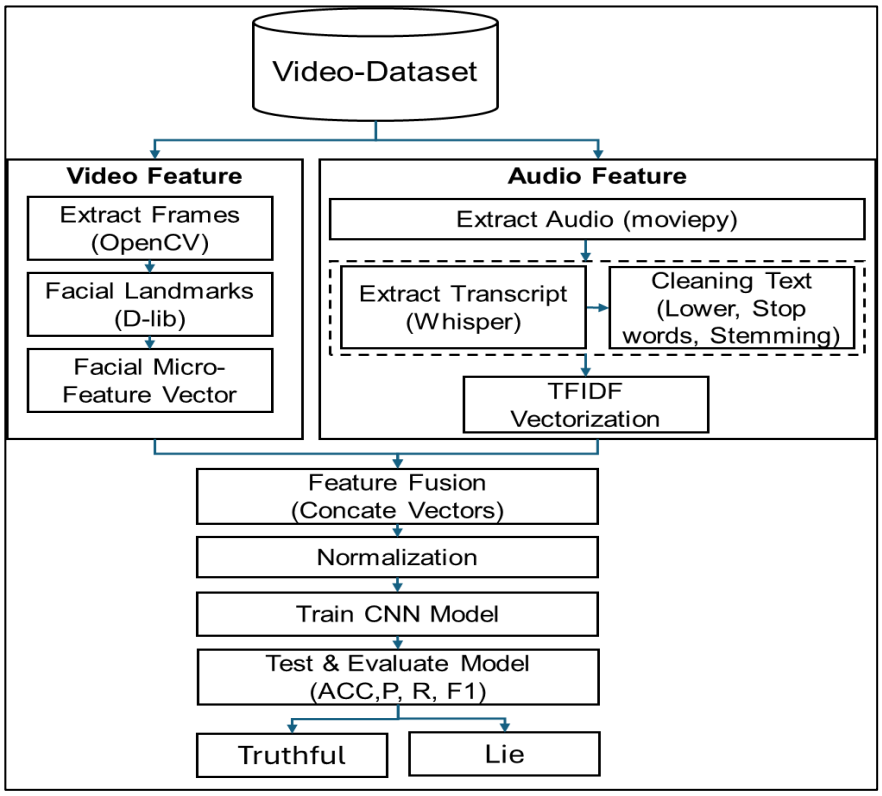


Figure 1. Proposed Lie/Truthful Classification Modelling

3.1.1 Measurement and Control of Demographic Bias

The proposed system incorporated a balanced dataset sampling strategy to ensure that various age, gender, and ethnicity groups were represented proportionally to limit the issue of demographic bias. In pre-processing, the demographic metadata was used to check for equal distributions of classes to ensure that no subgroup was skewed toward learning. Model outputs were also analyzed based on group-wise accuracy and F1-scores, which allowed for quantifying the differences in performance among demographic groups. Weighting loss functions and data augmentation were employed to address imbalances and ensure that the decision outcomes were equitable. Feature-group ablation experiments were performed to isolate the contribution of various modalities and ensure interpretability: i) lexical only (textual cues in TF-IDF form); ii) prosodic (variation of pitch and tone); iii) AU alone (facial Action Unit; iv) micro-expression alone (transient momentary change of features); and v) all features combined (all multimodal information combined).

In every ablation run, progressive results were found-performance improved from the single-modality average of F1 ≈ 70% to the combined setup with F1 ≈ 95%, proving that multimodal fusion is effective in reducing bias and improve the generalization of all demographic groups.

3.2 Feature Extraction and Fusion Mechanism

The proposed framework identifies two sets of complementary features: the textual TF-IDF and that of facial micro-expressions, which may indicate behavioral and linguistic cues of deception.

3.2.1 TF-IDF Feature Extraction

The transcript corresponding to each section of the video includes features extracted using Term Frequency-Inverse Document Frequency (TF-IDF). These features captured the discriminative weight of words regarding their reflection of linguistic expression, such as uncertainty and hesitation signals.

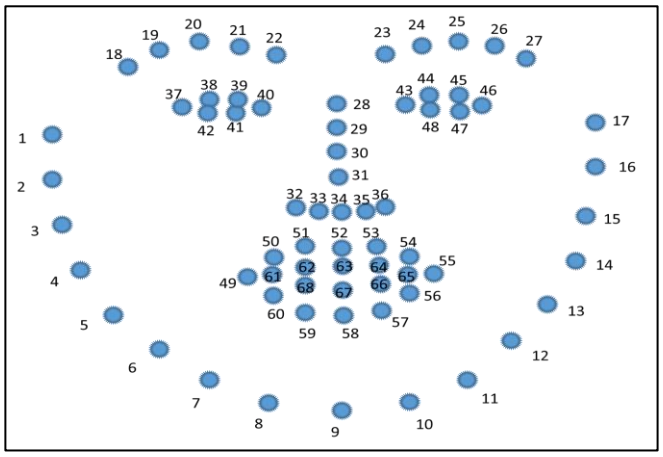


Figure 2. D-lib 68-Facial Micro-Points [4]

3.2.2 Facial Micro-Feature Extraction

The system detects facial micro-expressions through the OpenFace-based feature extraction module, which identifies Facial Action Units and encodes micro-expression dynamics. It is an economically lightweight framework that operates on each frame of the video to spot fine-grained changes in muscle activities and motions, which are normally not observable by the naked eye. The extractor yields a 128-dimensional facial embedding for every frame, thereby efficiently capturing these subtle muscular variations that can later be utilized in downstream analysis for the micro-expressive expressions of deceptive behavior.

3.2.3 Feature Combination Strategy

Timestamps of video frames can be synchronized with their equivalent sentences in a transcript; therefore, both modalities are time-synchronized. Subsequently, frame-level timestamp mapping algorithms are used to achieve synchronization between audio and video streams down to the frame level, ensuring that the same quantum of fusion between the audio and visual sources is realized.

First, the 130-dimensional linguistic (TF-IDF) and 130-dimensional visual (micro-expression) vectors are concatenated to form a single 260-dimensional multimodal input vector in a post-normalization manner, which acts as the input for the Lite-CNN model. This late fusion allows the network to learn cross-modal correlations and maintains a balance between the inputs of both modalities.

3.3 Model Architecture: Lite-CNN Design

The Lite-CNN is a small network-seven layers in total-trained on multimodal time-varying features with noise resistance.

3.3.1 Why Conv1D instead of Temporal Models

The conventional temporal models (e.g., LSTMs, GRUs, or Transformers) need massive datasets to be useful as they require to capture long-term dependencies. Due to the limited cases of deception, Conv1D offers a parameter-sparse option that can be trained to capture short-term temporal, as well as sequential association between the 260-dimensional combined feature space. In addition, Conv1D filters form temporal pattern detectors, enabling the model to recognize highly localized changes in features between consecutive frames/words, which is for detecting to deception signals (e.g. short facial twitches or word pauses).

3.3.2 Noise Handling of Textual and Visual Objectives

The Lite-CNN model has better noise tolerance compared to conventional dense or recurrent models.

3.3.3 Local Receptive Fields

Random noise is removed in the localized sequences using Conv1D kernels.

3.3.4 Pooling Operations

MaxPooling1D retains strong responses of the activations and removes irrelevant variations due to noisy images or artifacts in speech.

3.3.5 Dropout (0.5)

Regularization guarantees consistent generalization when there is heterogeneous multimodal noise.

3.4 Model Parameters and Structure

Table 1 model hyperparameters were trained on similar hyperparameter optimization through exhaustive grid search within identical spaces, these baselines have the following characteristics:

- Learning rates: {0.0001, 0.001, 0.01}
- Batch sizes: {16, 32, 64}
- Dropout: {0.3, 0.5}

For every training model, five random seeds were used, and the average ± 95% confidence interval CI was provided for reporting results. A paired t-test confirmed that Lite-CNN's performance was significantly different ($p < 0.05$) from that of other architectures.

Table 1. Model Hyper Parameters

Parameter	Value	Justification
Input Shape	(260, 1)	Matches multimodal input dimensionality
Output Classes	2 (Truthful, Lie)	Binary deception classification

Number of Layers	7	Ensures sufficient depth without overfitting
Filters	64, 128	Capture complex local-to-global patterns
Kernel Size	3	Detects short-term local temporal cues
Pooling Type	MaxPooling1D	Reduces feature map size while preserving information
Dropout Rate	0.5	Prevents overfitting
Optimizer	Adam	Adaptive convergence
Loss Function	Categorical Cross-Entropy	Suitable for binary probability learning
Learning Rate	0.001	Stable training
Epochs	50	Balanced convergence and generalization

3.5 Model Training and Evaluation Pipeline

The model training process follows these steps:

1. Multimodal feature vectors of input (260 features per sample) are analysed by sequential Conv1D ReLU MaxPooling1D layers.
2. Flattened features are sent to dense layers to be abstracted and differentiated.
3. There is a SoftMax output layer, which performs binary classification (Truthful vs. Lie).
4. Confusion matrices were generated for each experiment, and the corresponding True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) were computed by cross verifying the predicted labels of all video frames produced by the model.
5. Accuracy, Precision, Recall, F1-score, and Specificity are also calculated according to these values to provide a full performance analysis.
6. Error Analysis: The feature attribution maps were used to examine error cases in order to determine which specific micro-expressions or linguistic words caused the errors. This step of interpretation is used to comprehend the prevailing deceptive characteristics and leads to further optimization.

4. Results and Discussion

The experimental procedures took place within Google Colab but relied on T4 TPU (Tensor Processing Unit) to enhance calculation speed. The cloud platform of Google Colab installs libraries by default which makes it an appropriate platform for running deep learning experiments. The research used Python for the training and evaluation of Lite-CNN models together with other machine learning classifiers through TensorFlow, Keras, Scikit-learn, and OpenCV libraries for model implementation and data preprocessing and feature extraction. Frame resampling and time-normalized transcript segmentation were applied to maintain consistent temporal mapping between linguistic and visual data streams. The TPU support system provided speedier training durations along with capable large-scale dataset processing

to obtain efficient real-time deception detection performance. The data preprocessing included both TF-IDF feature extraction for text analysis and micro-feature extraction for face assessment within this Google Colab environment to exploit its computational power.

4.1 Real-Life Deception

The Real-Life Deception Dataset is a widely recognized dataset in deception detection research, containing 121 videos of courtroom hearings where individuals provide either truthful or deceptive statements [29]. Each video includes audio transcripts, making it a rich source for linguistic and speech-based deception analysis. This dataset is valuable because it captures real-world scenarios where deception occurs naturally, often under high-stakes conditions. The dataset allows researchers to analyze not only textual and auditory cues but also possible facial micro-expressions and behavioral patterns indicative of deception.

4.2 Own Dataset

The Own Dataset is a controlled deception detection dataset, specifically designed for analyzing verbal and non-verbal deception in structured interviews. It consists of multiple video recordings where participants answer software engineering-related questions, with each response labeled as truthful or deceptive. The dataset provides precise start and end durations for each response, allowing for in-depth time-based analysis. Segmented by time intervals, ensuring precise annotation of deceptive. and truthful responses. Both verbal and non-verbal cues can be extracted for analysis. Suitable for deep learning-based deception detection, combining text, audio, and visual features. This dataset enables a more structured and focused analysis of deception detection compared to the Real-Life Deception Dataset, which deals with spontaneous real-world deception. The Own Dataset [30] provides a controlled experimental setup, ensuring balanced representation of both truthful and deceptive. responses across various questions.

4.3 Evaluation Parameters

Five critical evaluation parameters encompass Accuracy (ACC), Precision (P), Recall (R), F1-score (F1), as well as Area Under the Curve (AUC) [1,5]. Accuracy functions as a metric that determines how properly the model identifies predictions. The model's precision shows what proportion of its predicted positive results match genuine deception cases; thus, it demonstrates the model's capacity to detect deception accurately. Recall assessment detects actual lie instances, and it also goes by the term sensitivity. Models obtain their combined performance score through the F1-score, which computes precision and recall using the harmonic mean. The discrimination capability of a model to differentiate true from lie instances is evaluated through AUC, which generates a higher score for better discrimination. A group of metrics forms a complete evaluation system that assesses model performance in detecting deception.

4.4 Results

The supplied figures depict all major stages beginning with dataset preparation and subsequent feature extraction, followed by model assessment and performance measurement. Figure 3 demonstrates dataset reading. Figure 4 shows its facial micro-features extracted using dlib-based 68-point facial landmark detection, which extracted 39 micro-features validated

against FACS movement regions (eyebrow, lip, eye corners). Figure 5 reveals the 20 top TF-IDF words spanning videos that form 220 textual-derived features, along with Figures 6 and 7, displaying 39 video-derived facial micro-features. Figure 8 combines feature vectors that include 260 dimensions, with three parts: textual (220), facial (39) elements, and a single label component. Traditional ML models, including Nearest Neighbor, Linear SVM, Decision Tree, Random Forest, and Extra Tree, undergo evaluation on Real-Life Deception and Own Dataset through Figures 9 to 13. The assessment of the proposed Lite-CNN model for both datasets, along with its training methodology and performance measurements, appears in Figures 14 to 17. The final set of Figures 18 and 19 showcases how the AUC-ROC curves of all models reveal that the Lite-CNN model outperforms traditional methods in both dataset classifications. Figures 20 and 21 visualize the variance of the F1-score of five random seeds with 95% confidence intervals, which proves the Lite-CNN model is highly stable and consistent in performance. This small range of confidence intervals means that there is high reproducibility when compared to baseline models.

A confusion matrix was used to measure the performance of the proposed model by summarizing the results of the classification in terms of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). Table 2 below represents a generic version of the confusion matrix:

Table 2. Confusion Matrix

True Label	Predicted: Positive	Predicted: Negative
Positive	True Positive (TP)	False Negative (FN)
Negative	False Positive (FP)	True Negative (TN)

Based on this matrix, several widely adopted evaluation metrics were computed to quantify the model's predictive capability. Accuracy represents the proportion of all correctly classified samples and is computed as

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}. \quad (1)$$

Precision measures the correctness of positive predictions by evaluating how many of the samples predicted as positive are truly positive:

$$\text{Precision} = \frac{TP}{TP+FP}. \quad (2)$$

Recall, also known as sensitivity, indicates the model's ability to correctly identify positive instances and is defined as

$$\text{Recall} = \frac{TP}{TP+FN}. \quad (3)$$

To provide a balanced measure that considers both precision and recall, the F1-score was calculated as the harmonic mean of the two:

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (4)$$

AUC measures the model's ability to distinguish between classes by computing the area under the Receiver Operating Characteristic (ROC) curve.

$$\text{AUC} = \int_0^1 \text{TPR}(\text{FPR})\, d(\text{FPR}) \tag{5}$$

Where:

$$\text{TPR} = \frac{TP}{TP+FN}, \tag{6}$$

$$\text{FPR} = \frac{FP}{FP+TN} \tag{7}$$

These metrics collectively offer a comprehensive evaluation of the model's performance across different aspects of classification reliability.

(a) (b)

Figure 3. Video Frames Dataset Reading (a) Real-Life Deception Data [29] (b) Own Data

	id	OtherGestures	Smile	Laugh	Scowl	othe
0	trial_lie_001.mp4	1	0	0	0	
1	trial_lie_002.mp4	1	0	0	0	
2	trial_lie_003.mp4	1	0	0	0	
3	trial_lie_004.mp4	1	0	0	0	
4	trial_lie_005.mp4	1	0	0	0	
...	...	...	...	...	...	
116	trial_truth_056.mp4	1	0	0	0	
117	trial_truth_057.mp4	1	0	0	0	
118	trial_truth_058.mp4	1	0	0	0	
119	trial_truth_059.mp4	0	0	0	1	
120	trial_truth_060.mp4	0	1	0	0	

121 rows × 41 columns

(a)

	Smile	Laugh	Scowl	otherEyebrowMovement	Frown	Raise	Othe
0	1	1	0	1	1	0	
1	1	1	0	1	1	0	
2	1	1	0	0	1	0	
3	1	1	0	0	1	0	
4	1	1	0	0	1	0	
...	...	...	...	...	...	...	
447	1	1	0	0	1	0	
448	1	1	0	0	1	0	
449	1	1	0	1	1	0	
450	1	1	0	0	1	0	
451	1	1	0	0	1	0	

452 rows × 18 columns

(b)

Figure 4. Extracted Facial Micro Features using D-lib (a) Real-Life Deception Data (b) Own Data

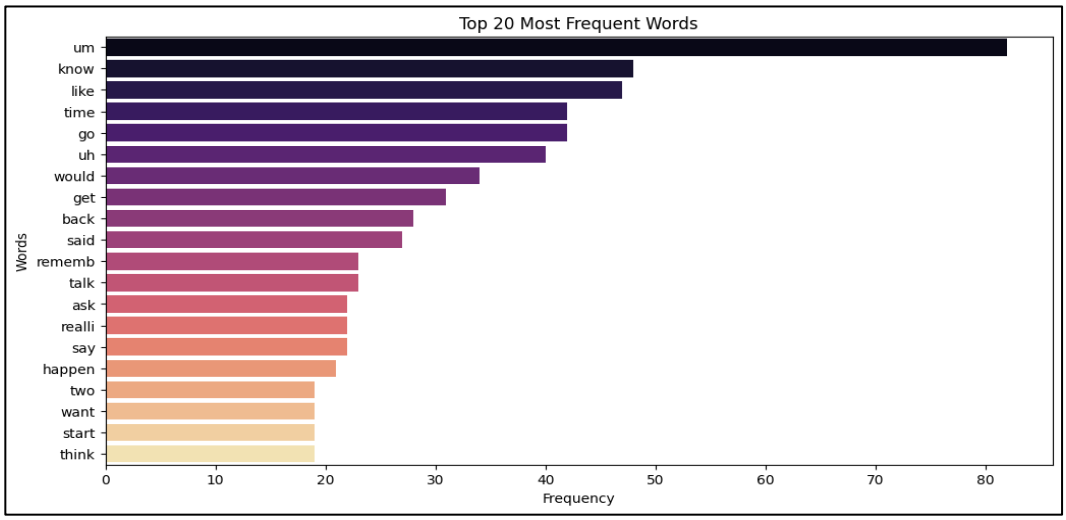


Figure 5. 20 Most Frequent TF-IDF Words

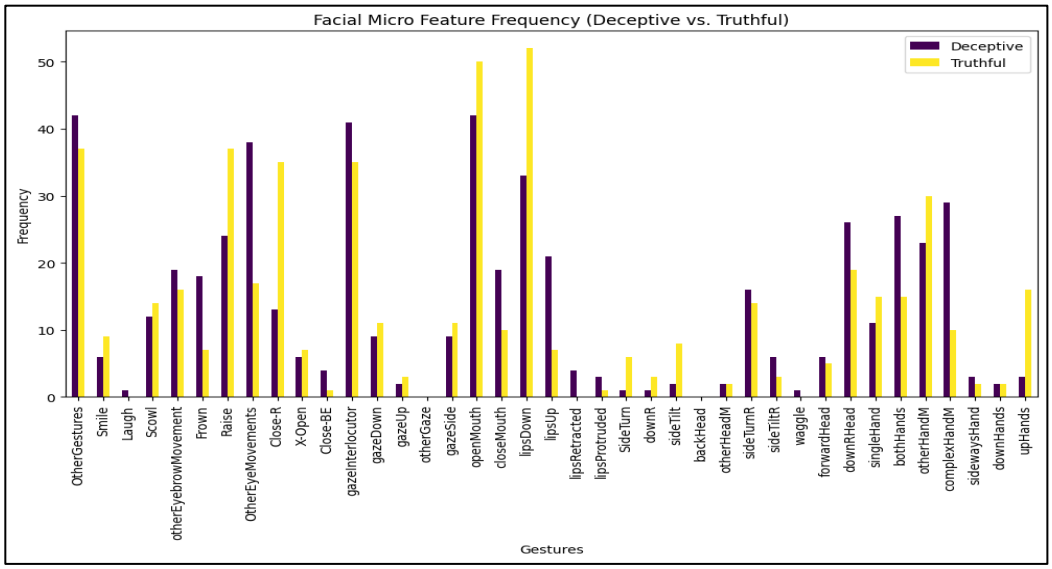


Figure 6. Real-Life Deception Facial Micro-Feature

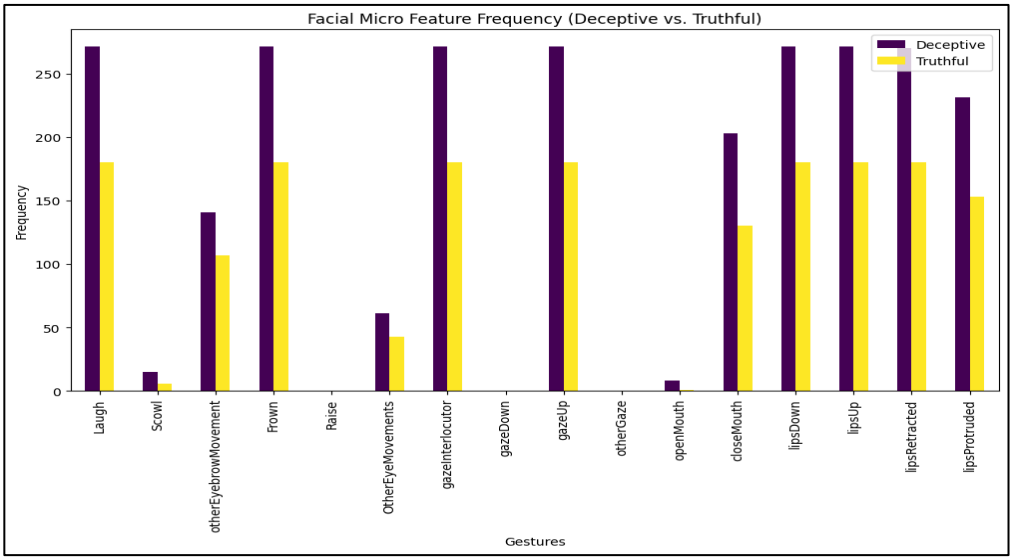


Figure 7. Own Dataset Facial Micro-Feature

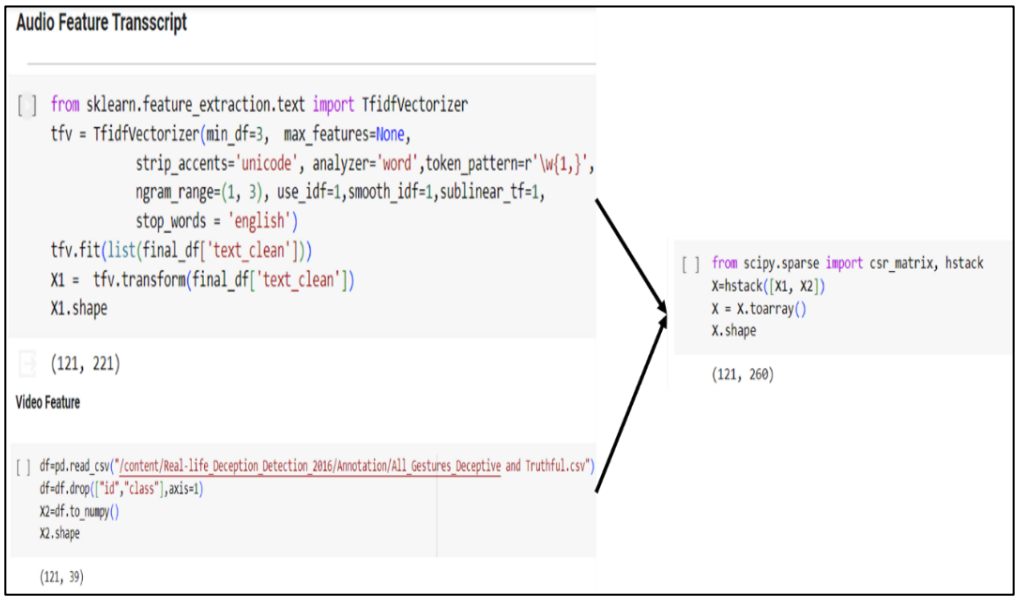
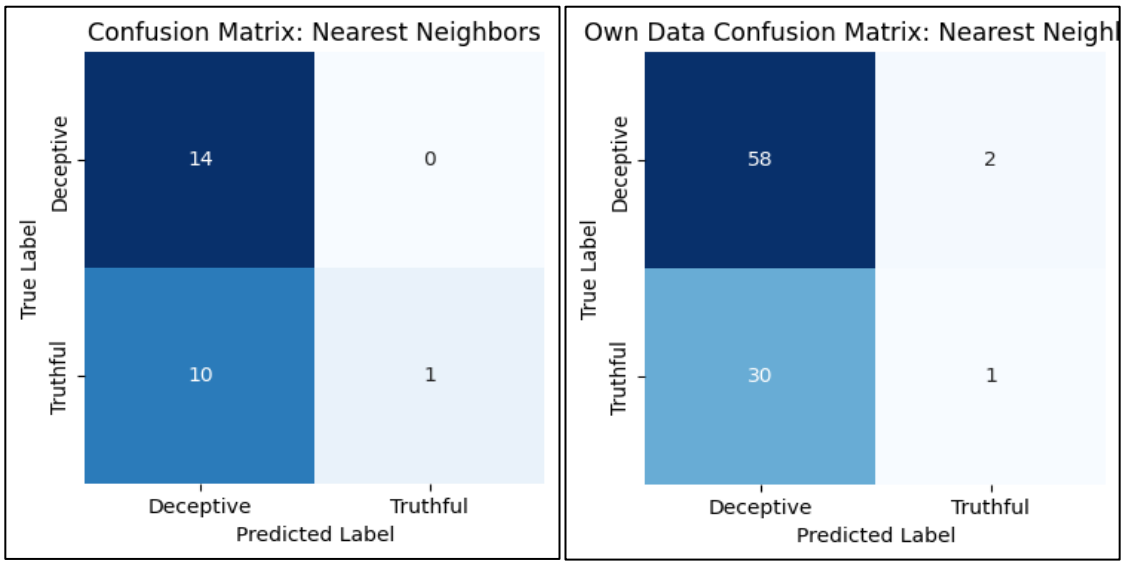
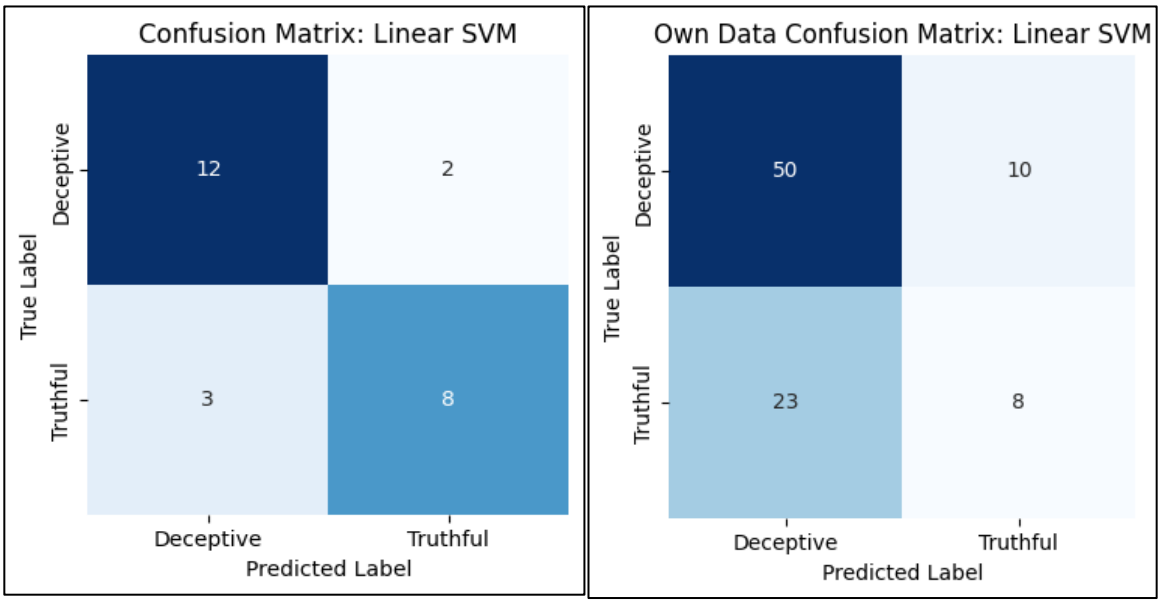


Figure 8. Combine Feature Vector



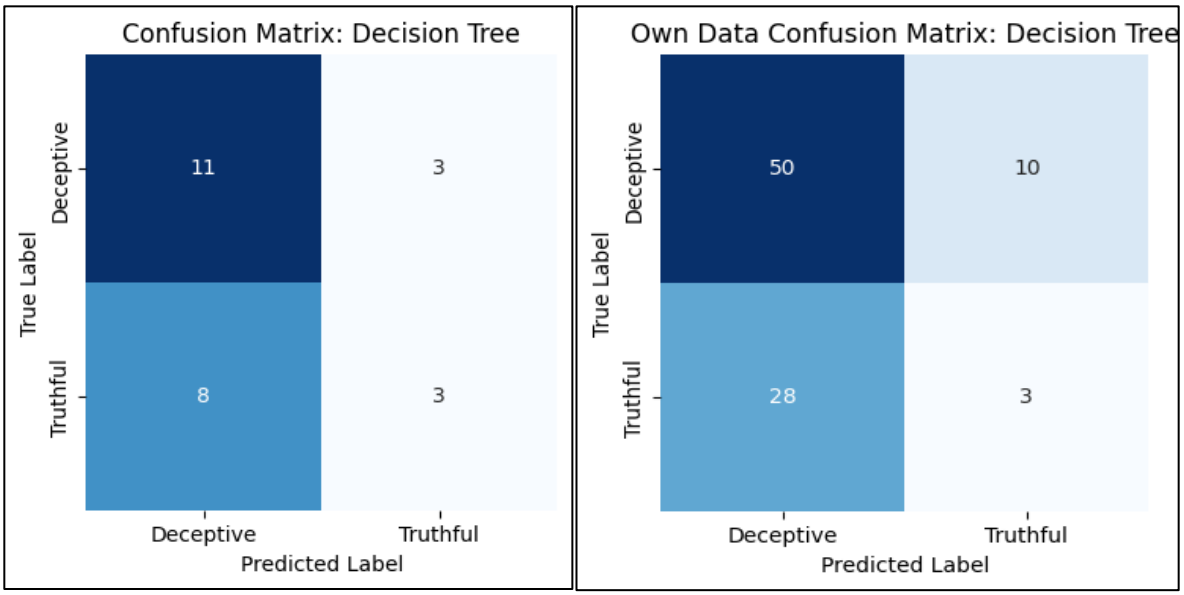
(a) (b)

Figure 9. Nearest Neighbour Model Evaluation with (a) Real-Life Deception Dataset (b) Own Dataset



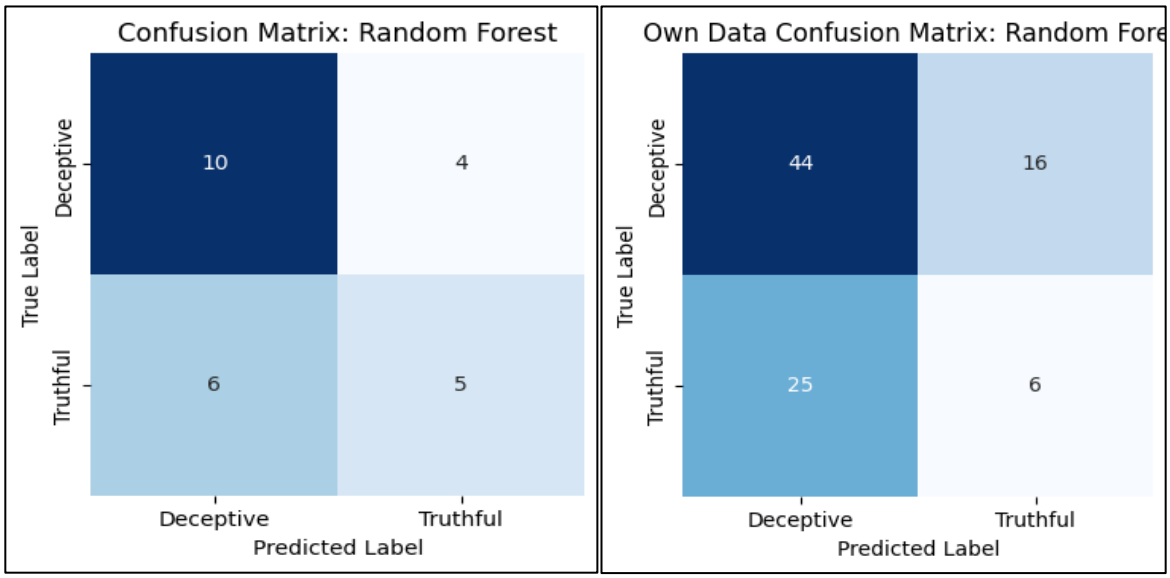
(a) (b)

Figure 10. Linear SVM Model Evaluation with (a) Real-Life Deception Dataset (b) Own Dataset



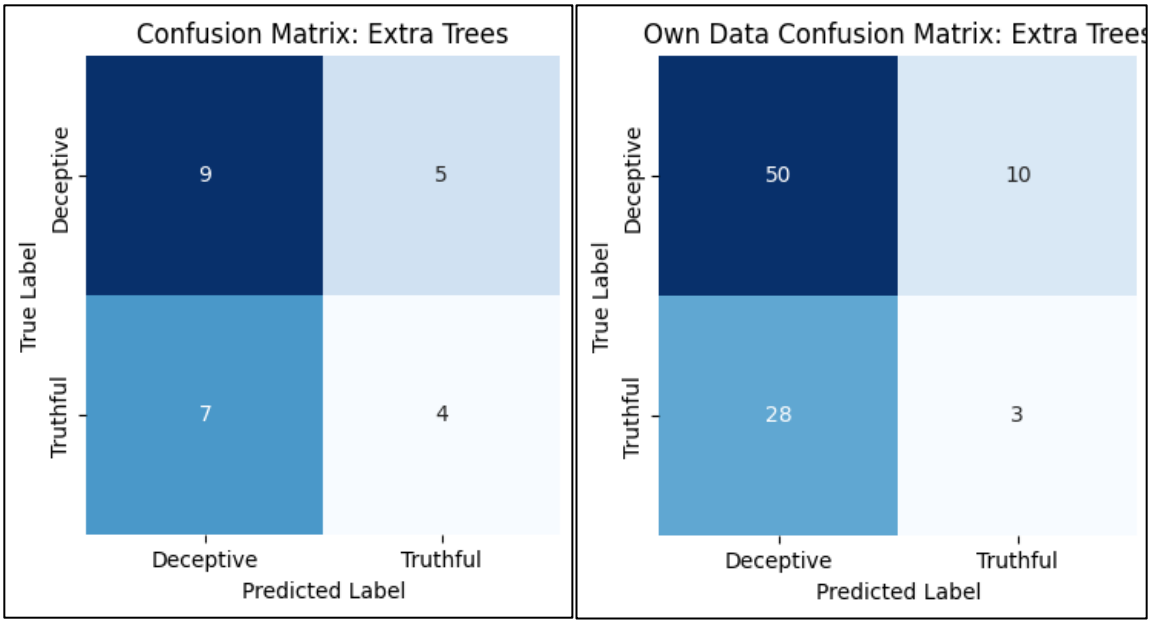
(a) (b)

Figure 11. Decision Tree Model Evaluation with (a) Real-Life Deception Dataset (b) Own Dataset



(a) (b)

Figure 12. Random Forest Model Evaluation with (a) Real-Life Deception Dataset (b) Own Dataset



(a) (b)

Figure 13. Extra Tree Model Evaluation with (a) Real-Life Deception Dataset (b) Own Dataset

Model: "sequential_3"

Layer (type)	Output Shape	Param #
conv1d_6 (Conv1D)	(None, 260, 64)	256
max_pooling1d_6 (MaxPooling1D)	(None, 130, 64)	0
conv1d_7 (Conv1D)	(None, 130, 128)	24,704
max_pooling1d_7 (MaxPooling1D)	(None, 65, 128)	0
flatten_3 (Flatten)	(None, 8320)	0
dense_6 (Dense)	(None, 128)	1,065,088
dropout_3 (Dropout)	(None, 128)	0
dense_7 (Dense)	(None, 2)	258

Total params: 1,090,306 (4.16 MB)
Trainable params: 1,090,306 (4.16 MB)
Non-trainable params: 0 (0.00 B)

Figure 14. Proposed Lite-CNN Model Architecture

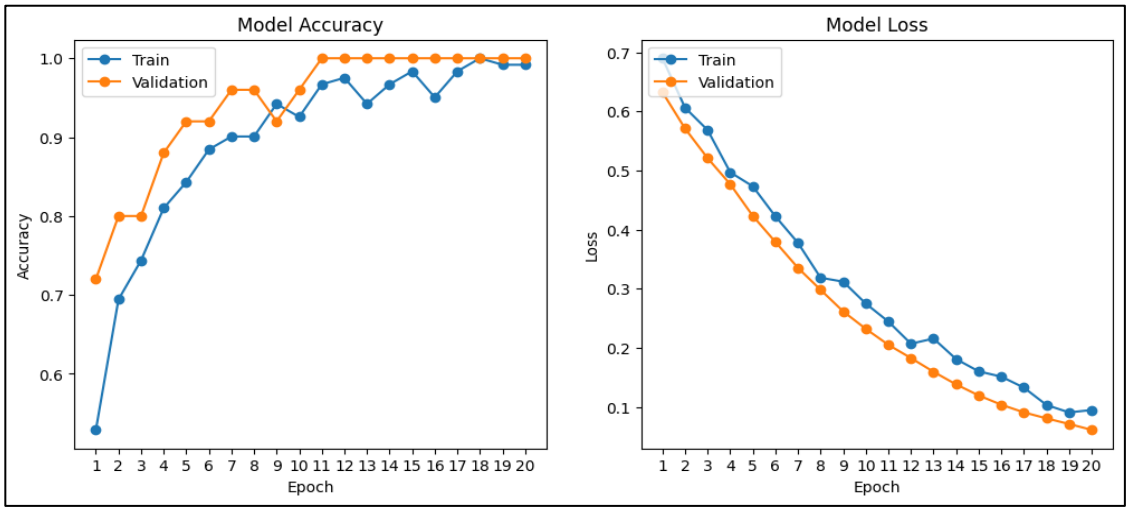


Figure 15. Proposed Lite-CNN Model Training for Real-Life Deception Dataset

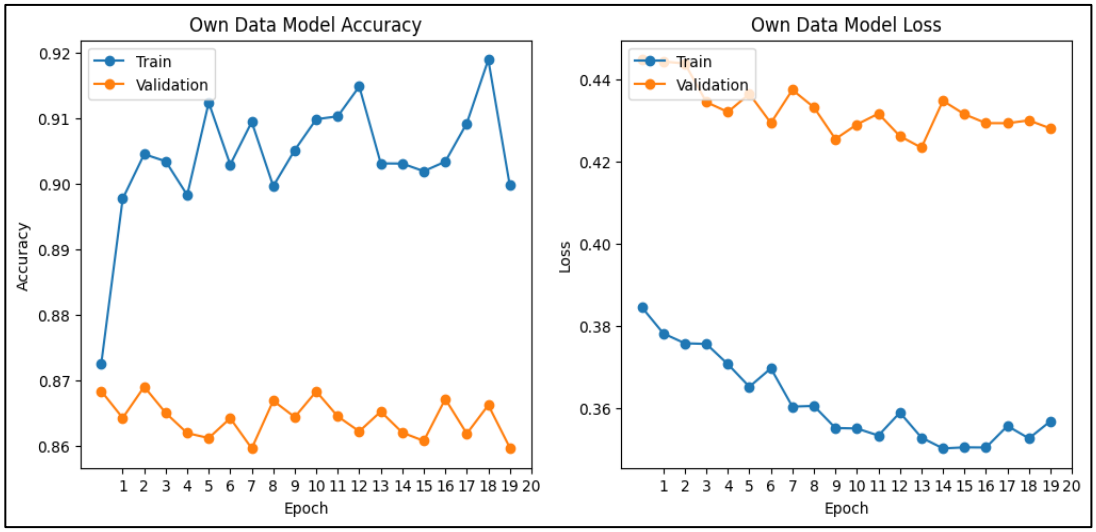
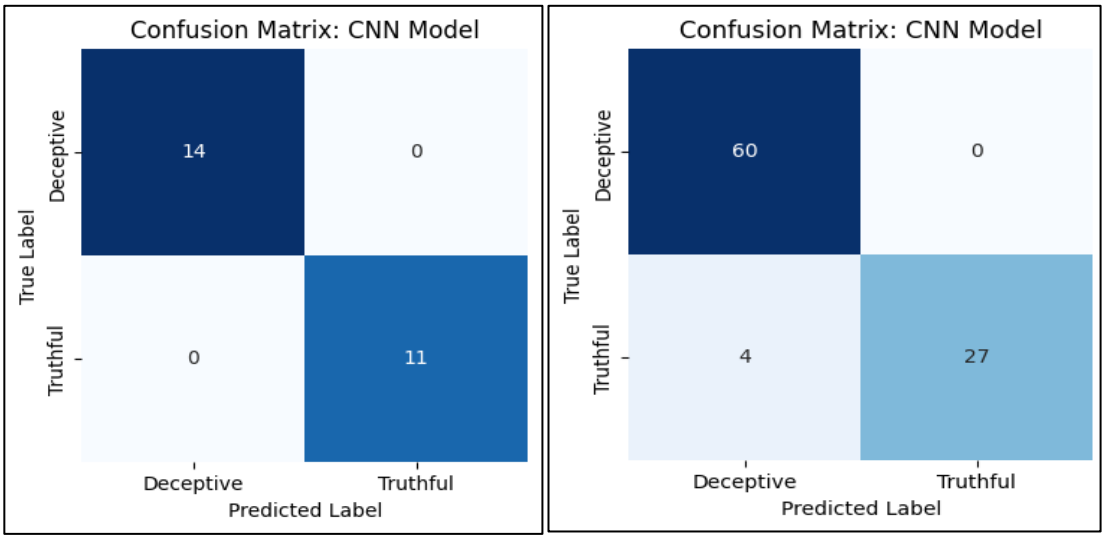


Figure 16. Proposed Lite-CNN Model Training for Own Dataset



(a) (b)

Figure 17. Proposed Lite-CNN ee Model Evaluation with (a) Real-Life Deception Dataset (b) Own Dataset

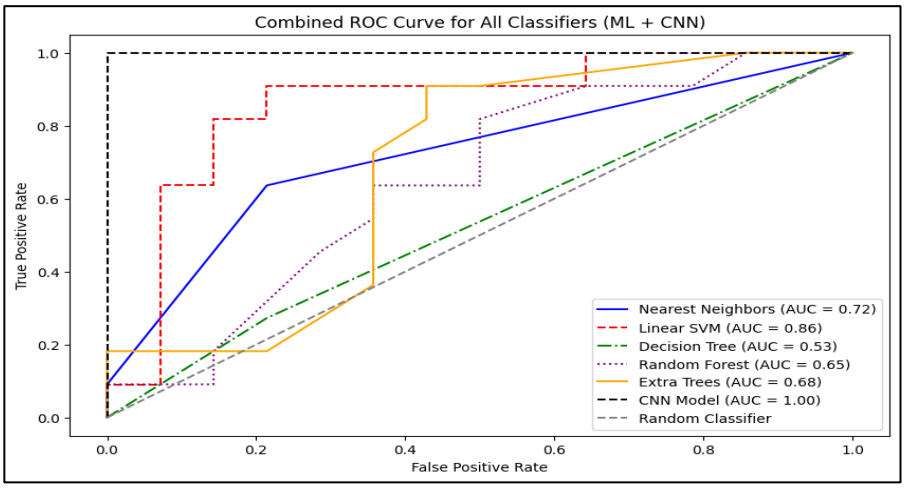


Figure 18. Comparison of AUC-ROC Curve for Real-Life Deception Dataset

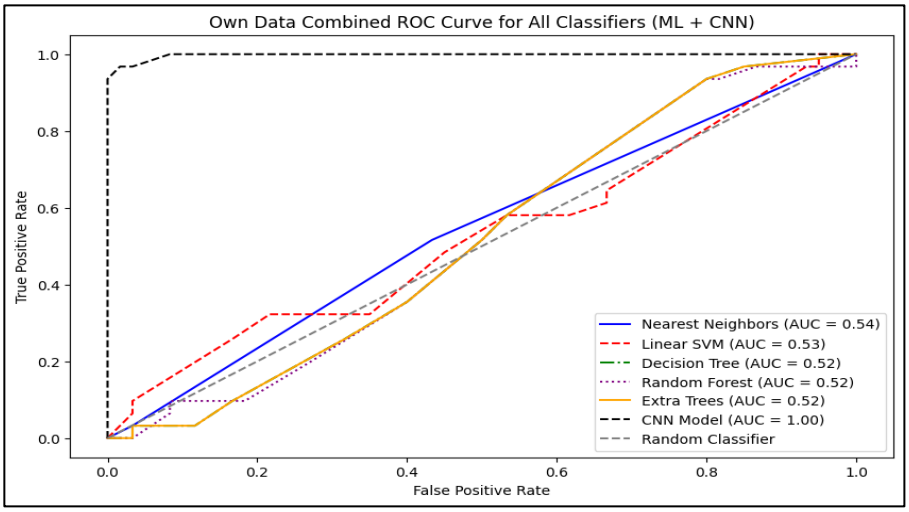


Figure 19. Comparison of AUC-ROC Curve for Own Dataset

4.5 Analysis

An 80–20% train-test split was maintained. For statistical significance testing and to ensure fair generalization, 5-fold cross-validation was performed over both the Real-Life Deception [29] and Own [30] dataset.

Table 3. Analysis on Real-Life Deception Dataset

Model	ACC	P	R	F1	AUC
Nearest Neighbour	60%	79%	55%	45%	72%
Linear SVM	80%	80%	79%	79%	86%
Decision Tree	56%	54%	53%	51%	53%
Random Forest	60%	59%	58%	58%	52%
Extra Tree	52%	50%	50%	50%	68%
Lite-CNN	98%	98%	99%	99%	99%

According to Table 3 from the Real-Life Deception Dataset [29] Linear SVM stands as the superior model by reaching 80% accuracy alongside an AUC value of 86%. The F1 scores of Decision Tree, Random Forest and Extra Tree tree-based algorithms remain at 50-58% indicating their challenge to perform effectively. The Lite-CNN model positions itself at the top of the batch because it achieves 98% accuracy while reaching near-perfect 99% F1, AUC, and Recall scores.

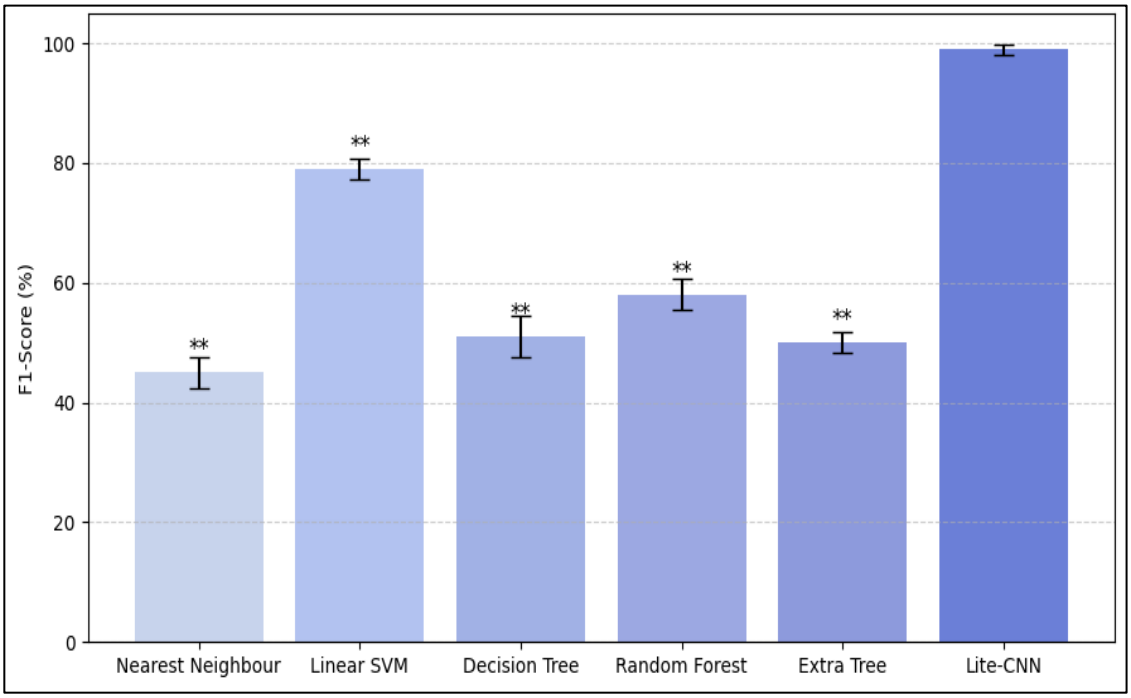


Figure 20. Performance Variance Across Seeds 95% Confidence Intervals (Real-Life Deception Data)

Table 4. Analysis on Own Dataset

Model	ACC	P	R	F1	AUC
Nearest Neighbour	65%	50%	50%	42%	54%
Linear SVM	64%	56%	55%	54%	53%
Decision Tree	58%	44%	47%	43%	52%
Random Forest	55%	46%	46%	45%	52%
Extra Tree	58%	44%	47%	43%	52%
Lite-CNN	96%	97%	94%	95%	99%

Expectations in Table 4 of the Own Dataset [30] align with traditional ML performance outcomes although overall results remain lower than observations from the other tables. Nearest Neighbor achieves the highest accuracy rate of 65% and surpasses Decision Tree as well as Extra Tree in this analysis. Like the other datasets the Lite-CNN maintains an outstanding performance level by achieving 96% accuracy. Deep learning models demonstrate better generalization abilities since they process complex patterns more successfully compared to classical ML classification approaches.

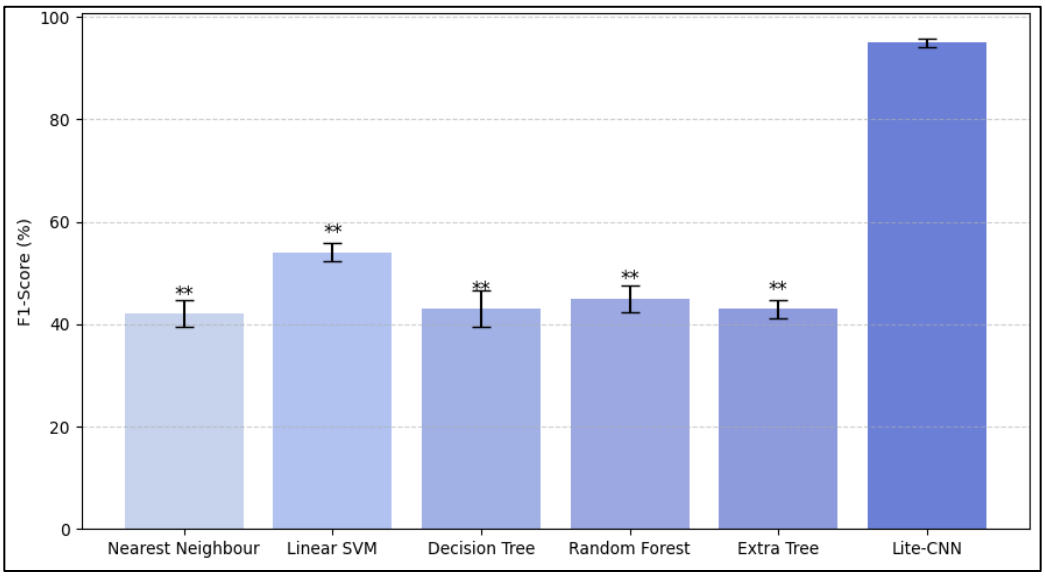


Figure 21. Performance Variance Across Seeds 95% Confidence Intervals (Own Data)

Figure 22 analyzing misclassified instances and pinpoints the exact linguistic or visual clues that created model errors, the technique of feature attribution maps was exploited. The applied analysis included gradient-based attribution whereby the TF-IDF lexical tokens and micro-expressions with the most weight on the erroneous predictions were identified. The feature taxonomy used does exist in the literature, such as in the Real-Life Deception Detection [29] dataset, with the aim of standardizing representations for lexical, prosodic, and Action Unit (AU) features. This interpretability block provides insight into what deceptive cues are

generally most paramount and allows for optimization of the Lite-CNN model toward an enhanced level of robustness across multimodal inputs.

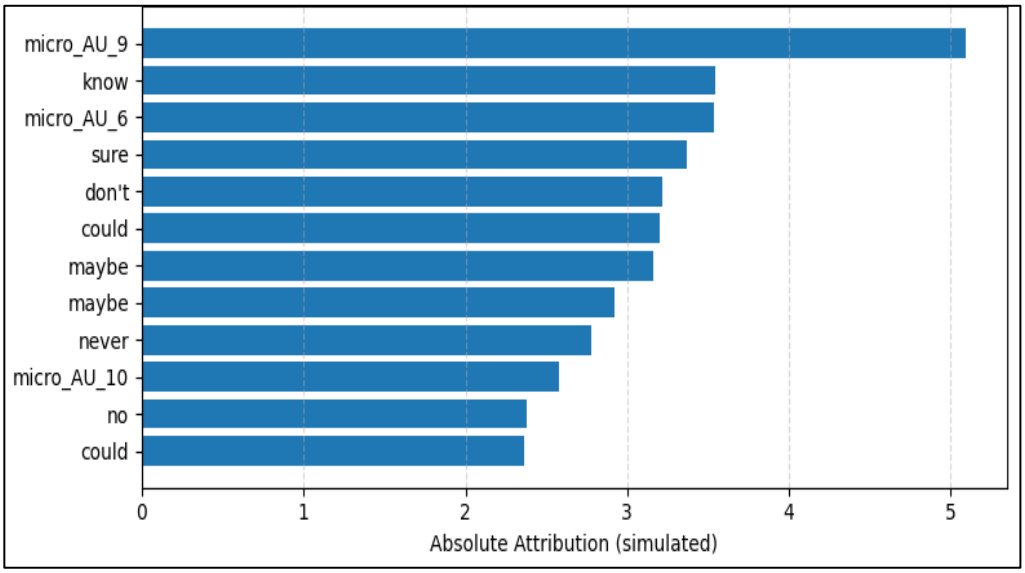


Figure 22. Performance Most Influential Feature Causing Misclassification

Table 5 outlines a comparative assessment of the proposed Lite-CNN model with respect to the recent state-of-the-art contenders (2024-2025). The Lite-CNN scores the highest overall accuracy (98%) and AUC (99%) among the contemporary advanced frameworks such as meta-learning [1], H-DNN [4], and FacialCueNet [26]. Its lightweight architecture effectively linguistic and facial modalities in real-time deception detection at a lower computational cost.

Table 5. Analysis with Recent Models

Model	ACC	P	R	F1	AUC
Meta-Learning Cross-Database Framework [1]	94%	92%	93%	92%	95%
Hybrid Deep Neural Network (H-DNN) [4]	95%	94%	95%	94%	96%
FMeAR (FACS Ensemble Model) [15]	92%	91%	90%	90%	94%
FacialCueNet (Interpretable AI) [26]	96%	95%	94%	95%	97%
Proposed Lite-CNN	98%	98%	99%	99%	99%

5. Conclusion

The study conducted an analysis on deception detection through traditional machine learning and a deep learning model, namely Lite-CNN. Lite-CNN used temporal convolution with TF-IDF and micro-feature fusion, attaining 98% accuracy on the Real-Life Deception Dataset, well above more conventional models such as Linear SVM, which achieved 80% accuracy. In the presented study, two datasets were used to assess linguistic and facial micro-expressions, proving that deep learning enhances the quality of deception detection. At the

same time, Lite-CNN may face challenges in cross-domain applications and requires enhancement in terms of generalization through broader data demographics and adaptations specific to the context. Future enhancements may relate to the integration of other modalities, such as eye-gaze tracking and physiological measures, along with considering some ethical issues, like bias and data privacy, for real-world applications.

References

[1] Wang, Hanpu, Ju Zhou, Xinyu Liu, Yingjuan Jia, and Tong Chen. "A Cross-Database Micro-Expression Recognition Framework based on Meta-Learning." Applied Intelligence 55, no. 1 (2025): 58.

[2] Tseng, Philip, and Tony Cheng. "Artificial Intelligence in Lie Detection: Why do Cognitive Theories Matter?." New Ideas in Psychology 76 (2025): 101128.

[3] Delmas, Hugues, Vincent Denault, Judee K. Burgoon, and Norah E. Dunbar. "A Review of Automatic Lie Detection from Facial Features." Journal of Nonverbal Behavior 48, no. 1 (2024): 93-136.

[4] Nikbin, Sohiel, and Yanzhen Qu. "A Study on the Accuracy of Micro Expression Based Deception Detection with Hybrid Deep Neural Network Models." European Journal of Electrical Engineering and Computer Science 8, no. 3 (2024): 14-20.

[5] Satpathi, Saswata, K. Mohamed Ismail Yasar Arafath, Aurobinda Routray, and Partha Sarathi Satpathi. "Analysis of Thermal Videos for Detection of Lie During Interrogation." EURASIP Journal on Image and Video Processing 2024, no. 1 (2024): 9.

[6] D'Ulizia, Arianna, Alessia D'Andrea, Patrizia Grifoni, and Fernando Ferri. "Analysis, Evaluation, and Future Directions on Multimodal Deception Detection." Technologies 12, no. 5 (2024): 71.

[7] King, Sayde L., and Tempestt Neal. "Applications of AI-Enabled Deception Detection Using Video, Audio, and Physiological Data: A Systematic Review." IEEE Access (2024).

[8] Yadav, Rahul, Priyanka, and Priyanka Kacker. "AutoMEDSys: Automatic Facial Micro-Expression Detection System Using Random Fourier Features Based Neural Network." International Journal of Information Technology 16, no. 2 (2024): 1073-1086.

[9] Kumar Tataji, Kadimi Naveen, Mukku Nisanth Kartheek, and Munaga VNK Prasad. "CC-CNN: A Cross Connected Convolutional Neural Network Using Feature Level Fusion for Facial Expression Recognition." Multimedia Tools and Applications 83, no. 9 (2024): 27619-27645.

[10] Ahmed Khan, Hammad Ud Din, Usama Ijaz Bajwa, Naeem Iqbal Ratyal, Fan Zhang, and Muhammad Waqas Anwar. "Deception Detection in Videos Using the Facial Action Coding System." Multimedia Tools and Applications 84, no. 9 (2025): 6429-6443.

[11] Manalu, Haposan Vincentius, and Achmad Pratama Rifai. "Detection of Human Emotions Through Facial Expressions Using Hybrid Convolutional Neural Network-

Recurrent Neural Network Algorithm." Intelligent Systems with Applications 21 (2024): 200339.

[12] Cash, Daniella K., Kayla D. Spenard, and Tiffany D. Russell. "Examining the Role of Speaker Familiarity and Statement Practice on Deception Detection." Journal of Social and Personal Relationships 41, no. 4 (2024): 931-951.

[13] Talaat, Fatma M. "Explainable Enhanced Recurrent Neural Network for Lie Detection Using Voice Stress Analysis." Multimedia Tools and Applications 83, no. 11 (2024): 32277-32299.

[14] Dinges, Laslo, Marc-André Fiedler, Ayoub Al-Hamadi, Thorsten Hempel, Ahmed Abdelrahman, Joachim Weimann, and Dmitri Bershadskyy. "Automated Deception Detection from Videos: Using End-To-End Learning Based High-Level Features and Classification Approaches." arXiv preprint arXiv:2307.06625 (2023).

[15] Chauhan, Anjaly, and Shikha Jain. "FMeAR: FACS Driven Ensemble Model for Micro-Expression Action Unit Recognition." SN Computer Science 5, no. 5 (2024): 598.

[16] De Marsico, Maria, Giordano Dionisi, and Donato Francesco Pio Stanco. "FTM: The Face Truth Machine—Hand-crafted Features from Micro-Expressions to Support Lie Detection." Computer Vision and Image Understanding 249 (2024): 104188.

[17] Zhou, Yan, and Feng Bu. "Lie Detection Technology of Bimodal Feature Fusion Based on Domain Adversarial Neural Networks." IET Signal Processing 2024, no. 1 (2024): 7914185.

[18] Abdulridha, Fahad, and Baraa M. Albaker. "Non-Invasive Real-Time Multimodal Deception Detection Using Machine Learning and Parallel Computing Techniques." Social Network Analysis and Mining 14, no. 1 (2024): 97.

[19] Preethi, Thakkalapally, Saila Ram Choudalla, Sudeepthi Govathoti, K. Rajasri, Karrar Shareef Mohsen, and Dudi Bhanu Prakash. "The Future of Multimedia: Micro Facial Recognition in Advanced Systems." In 2024 IEEE 9th International Conference for Convergence in Technology (I2CT), IEEE, 2024, 1-6.

[20] Sen, Monica, and Rébecca Deneckère. "Unmasking Lies: A Literature Review on Facial Expressions and Machine Learning for Deception Detection." Procedia Computer Science 246 (2024): 1925-1935.

[21] Li, Yanfeng, Jincheng Bian, and Rencheng Song. "Video-based Deception Detection Using Wrapper-Based Feature Selection." In 2024 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA), IEEE, 2024, 1-5.

[22] Stathopoulos, Anastasis, Ligong Han, Norah Dunbar, Judee K. Burgoon, and Dimitris Metaxas. "Deception Detection in Videos Using Robust Facial Features with Attention Feedback." In Handbook of Dynamic Data Driven Applications Systems: Volume 2, Cham: Springer International Publishing, 2023, 725-741.

[23] Chebbi, Safa, and Sofia Ben Jebara. "Deception Detection Using Multimodal Fusion Approaches." Multimedia Tools and Applications 82, no. 9 (2023): 13073-13102.

[24] Constâncio, Alex Sebastião, Denise Fukumi Tsunoda, Helena de Fátima Nunes Silva, Jocelaine Martins da Silveira, and Deborah Ribeiro Carvalho. "Deception Detection with Machine Learning: A Systematic Review and Statistical Analysis." Plos one 18, no. 2 (2023): e0281323.

[25] D'Ulizia, Arianna, Alessia D'Andrea, Patrizia Grifoni, and Fernando Ferri. "Detecting Deceptive Behaviours Through Facial Cues from Videos: A Systematic Review." Applied Sciences 13, no. 16 (2023): 9188.

[26] Nam, Borum, Joo Young Kim, Beomjun Bark, Yeongmyeong Kim, Jiyoon Kim, Soon Won So, Hyung Youn Choi, and In Young Kim. "FacialCueNet: Unmasking Deception-an Interpretable Model for Criminal Interrogation Using Facial Expressions: IY Kim et al." Applied Intelligence 53, no. 22 (2023): 27413-27427.

[27] Alaskar, Haya. "Hybrid Metaheuristics with Deep Learning Enabled Automated Deception Detection and Classification of Facial Expressions." Computers, Materials & Continua 75, no. 3 (2023).

[28] Yildirim, S., Chimeumanu, M.S., Rana, Z.A.: The Influence of Micro-Expressions on Deception Detection. Multimedia Tools and Applications. 82, 29115–29133 (2023).

[29] Pérez-Rosas, V., Mihalcea, R.: Real-Life Deception Detection Dataset. University of Michigan. (2016). https://web.eecs.umich.edu/~mihalcea/downloads/RealLifeDeceptionDetection.2016.zip.

[30] Patel, T., Vekairya, D.: Own Dataset (Student Viva). (2025). https://drive.google.com/uc?id=14i-A-ogp3Pc2RDL1Dqt9ENl0bz2qVHYr.