Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

# AI-Powered Lie Detection using Micro-Expressions

Mrs. R S Geethanjali
Assistant Professor, Dept. of Computer Science & Engineering
K.S. School of Engineering and Management
Bengaluru, India

Aditya Prakash Masabinal
Dept. of Computer Science & Engineering
K.S. School of Engineering and Management
Bengaluru, India

Bharath Kumar S C
Dept. of Computer Science & Engineering
K.S. School of Engineering and Management
Bengaluru, India

C Yuvaraj
Dept. of Computer Science & Engineering
K.S. School of Engineering and Management
Bengaluru, India

G Daewoo Sri Prasad
Dept. of Computer Science & Engineering
K.S. School of Engineering and Management
Bengaluru, India

***Abstract*** **- Deception detection has long been one of the most demanding challenges in behavioral science, forensic investigation, and security research. Conventional instruments such as the polygraph rely on invasive physiological measurement, remain susceptible to countermeasures, and produce results that many judicial systems refuse to admit as evidence. Meanwhile, human observers detect lies at rates barely exceeding chance. Facial micro-expressions—involuntary, fleeting muscular movements that surface when a person attempts to suppress an emotion—offer a compelling and non-invasive alternative indicator of deceptive intent. This survey paper systematically reviews five recent research contributions in AI-driven micro-expression analysis, covering Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) networks, hybrid CNN-LSTM architectures, handcrafted forensic feature extraction, multimodal audio-visual fusion, and comprehensive literature reviews. A comparative analysis of methodologies, datasets, performance metrics, and research gaps is presented. Building on identified gaps, a proposed real-time deception detection framework is described that integrates MediaPipe facial landmark extraction, CNN spatial feature learning, and LSTM temporal sequence modeling to generate calibrated confidence-scored Truth/Lie predictions suitable for criminal investigation, recruitment screening, and security surveillance.**

***Keywords*** **— Micro-expressions; deception detection; convolutional neural networks; LSTM; facial landmarks; deep learning; computer vision; MediaPipe; real-time processing.**

## I. INTRODUCTION

For thousands of years, humans have been able to communicate via spoken word. Nevertheless, scientists, law enforcement officials, and technology companies have struggled with developing ways to detect deception. Studies repeatedly demonstrate that human observers (whom do not receive specialized training) are correct in determining if a deception has taken place 54% of the time — which is equal to what would be expected from chance alone. The same holds true for police, investigators, and others who have undergone some degree of specialized training in detecting deception and/or analyzing human behavior in this regard; their percentages are only marginally higher than chance [5]. Therefore, experienced humans cannot be relied upon for successfully detecting deception using behavioral signals. The traditional polygraph (lie detector) has served as a basis for testing the deception of humans over the last several decades by measuring physiological reaction to perceived stress of being deceitful (blood pressure, heart rate, breathing rate, skin dryness). Though polygraph testing is widely used in law enforcement and employment screening, the many weak points of the testing process are substantial, including: being an invasive method; requiring specialized training in administering/reading the test; having a high false positive rate; as well as a skilled test subject being able to manipulate their physiological responses. Further demonstrating the limitations of polygraph testing as a useful forensic tool, results obtained from polygraph tests are inadmissible in a majority of global legal jurisdictions [1].

An alternative behavioral path for deception analysis lives in facial micro expressions rapid, almost involuntary facial muscle shifts that show up for a blink when someone is trying to hide or quiet a real emotional reaction. Typically these signals hang around for only 1/25 to 1/5 of a second, so they come off as effectively invisible to the unaided human eye, even when a person is really looking closely. And because micro expressions seem to pop out spontaneously from subcortical emotional processing , and they're so hard to manually restrain, they can work as a rather dependable doorway into what someone is actually feeling [4]. The meeting of Artificial Intelligence, Computer Vision, and Deep Learning has opened an unusual level of capability to find, examine, and interpret these fast facial indicators automatically, and often in real time. Convolutional Neural Networks pick up spatial facial cues with much finer granularity than human perception allows, and Long Short Term Memory networks handle the time based behavior of micro expression sequences as they play out across successive video frames [2][3]. In combination, these methods build the base of a new wave of intelligent , non invasive deception detection systems. This survey paper offers a sort of organized

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

rundown of five representative research contributions that help map the current state of the art. In Section II, there's a pointed analysis of each work. Section III points out key research gaps . Section IV lays out the aims of a proposed system. Section V explains the proposed methodology. Then Section VI and Section VII discuss applications and conclusions, respectively.

## II. LITERATURE SURVEY

### *A. Hybrid CNN-Machine Learning Framework for Deception Detection*

Hasanudin et al. [2] came up with a fusion based deception detection framework , that basically tries to blend the representational depth of Convolutional Neural Networks with the interpretability and computationally light weight nature of classical machine learning—meaning Support Vector Machines, Random Forest, and K-Nearest Neighbors. Their pipeline uses MediaPipe's 468 point, three dimensional facial mesh, to pull out geometric features which they claim have established behavioral links: mouth width, nose-to-mouth distance, brow elevation angle, lip corner displacement, and also left-right facial symmetry ratios. For the experiments, they used a corpus of 23,472 labeled facial expression samples taken from micro-expression video recordings, then split it into 17,634 for training and 5,838 for testing. The truth and deception labels were close to balanced, which is good I guess. The CNN reached a test accuracy of 99.96%, while KNN and Random Forest scored 91.79% and 90.92% respectively. Because the CNN result was so close to perfect, the authors didn't just stop there, they looked into overfitting risk more seriously. They tried to reduce it via stratified 80/20 train-test splitting, plus L2 weight regularization, and dropout layers too, which sounds pretty standard but still. The main contribution is kind of twofold: first, they show that facial geometric landmark coordinates can act as meaningful and discriminative indicators for deception. Second, they argue that pairing learned CNN representations with a more explainable classical classifier gives a better tradeoff between performance and transparency. The main limitation though is that all evaluations were done in controlled laboratory settings. So whether this model transfers well to different lighting conditions, ethnic variation, and real world camera setups, is still uncertain, or at least it's not fully validated.

**Key Gap:** Cross-dataset validation and real-world robustness testing are absent.

### *B. Lie Detection Using Facial Micro-Expression Features with Machine Learning*

Mohammed and Abdul Majeed [3] presented a computationally cheaper lie detection process, kinda based on classical image handling plus unsupervised grouping. Instead of using deep neural stuff, the authors seemed to go for something more approachable, building a multi step preprocessing flow that deals with the annoying stuff in video, like noisy content and blur. In their pipeline they run histogram equalization, Contrast Limited Adaptive Histogram Equalization (CLAHE), image sharpening , and noise reduction, then they move into feature extraction. For the feature side, they compared Sobel and Canny edge filters, focusing on how well each one could isolate gradient patterns that line up with micro-expression cues from facial images. After that, K-Means clustering was used as the last step for classification. Two results stood out. First, Sobel did better than Canny for extracting deception-relevant features again and again. Second, the mouth and nose regions were more informative than the periocular area , which matches neuroscientific ideas about lower-face emotional leakage that tends to be harder for people to hide on purpose. Their reported F1 scores were 0.996 for truth and 0.990 for deception. Even with those numbers, the overall setup still feels constrained by leaning on static image traits and conventional clustering. It also doesn't really keep track of timing, or anything about how micro-expressions actually change from one frame to the next, which is a big drawback since the dynamic unfolding of involuntary facial movements over time is one of the most distinctive and clinically useful cues they could have used.

**Key Gap:** Absence of temporal sequence modeling; video dynamics are not exploited.

### *C. Deception Discernment Using Audio and Micro-Expressions*

Elango et al. [6] tackled deception analysis in a sort of multimodal way by combining facial micro-expression cues together with acoustic speech characteristics. For the face side, Facial Action Units were taken out using OpenFace, and on the audio side, speech-derived Mel-Frequency Cepstral Coefficients were used to catch the prosodic plus spectral properties of the speaker's vocal output. The motivation behind this fusion, is fairly solid from behavioral science: deceivers often show those little incongruities between facial expression behavior and vocal patterning, so when you look at both at the same time, you can surface signals that neither channel on its own can really reveal.

Then Long Short-Term Memory networks were used, to track temporal patterns inside both of the feature streams, and the fused representations were evaluated across several datasets, such as courtroom video recordings, and the Bag-of-Lies collection. Overall, the multimodal approach showed noticeable gains compared to single-modality baselines, which lends support to cross-channel behavioral fusion as a strategy. Still, the reported accuracy is only around 62% , and that points to the inherent difficulty of deception detection outside the lab, messy recording conditions, and person-to-person behavioral variation can all weaken how well the model generalizes.

**Key Gap:** Moderate accuracy (~62%); complex preprocessing limits real-time scalability.

### *D. FTM: The Face Truth Machine—Handcrafted Features for Lie Detection*

The Face Truth Machine [4] is a rule-based deception detector inspired by forensic psychology, explicitly rejecting the opaque nature of neural networks and instead relying on hand-crafted facial features related to changes in landmarks' dynamic behavior. Specifically, the FTM tracks features indicative of particular patterns in brow movement, lip compression, periocular movement, and facial symmetry, classifying based on predetermined thresholds using behavioral characteristics of concealment of emotions.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

One of the main driving forces behind this approach was that of explainability. By delivering interpretable signals based on specific observable changes in behavior, this model offers understandable output to law enforcement officers, which can be challenged and verified—unlike many models that often prioritize precision but lack transparency. The FTM exhibited promising results across several datasets and did not require any GPU support, thereby ensuring practical application. The biggest downside of the FTM is its reliance on hand-crafted features, which, as a result, are inflexible and may not generalize well beyond a predetermined set of behavioral rules.

**Key Gap:** Limited adaptability under complex real-world variability; not scalable to new behavioral patterns.

*E. Unmasking Lies: A Literature Review on Facial Expressions and Machine Learning*

Sen and Deneckere [5] performed an exhaustive scientific literature review, comparing the extent to which machine learning and deep learning algorithms were utilized for deception detection through facial expressions. This review analyses several algorithms such as CNN, LSTM, Random Forest, 3D-CNN, Vision Transformers, and a combination of CNN-LSTM methods. Additionally, the Facial Action Coding System and Action Units were analyzed as the behavioral analysis systems employed in the studies under scrutiny. The study positions all the above technical approaches into a greater scientific controversy regarding whether facial expressions serve as universal deception indicators.

Overall, the review pinpoints five major issues common to all the research efforts on the topic: the lack of large deception databases; inadequate generalization performance of models; the absence of prediction reliability and explanation; inability to manage environmental differences; and the ethical aspect of large-scale biometric deception detection. The researchers conclude that, among others, deep learning techniques have the highest potential yet require significantly larger databases for further progress.

**Key Gap:** Dataset scarcity, low real-world robustness, and absence of explainable confidence-scored outputs.

*TABLE I. Comparative Summary of Reviewed Papers*

| Ref. | Method | Strength | Limitation |
|---|---|---|---|
| [2] | CNN+SVM/RF/KNN | High accuracy (99.96%) | Overfitting risk; controlled env. |
| [3] | Sobel+K-Means | Fast; low-cost | No temporal learning |
| [6] | LSTM+MFCC+AUs | Multimodal fusion | ~62% accuracy; complex pipeline |
| [4] | Handcrafted+Threshold | Explainable; lightweight | Limited adaptability |
| [5] | Systematic Review | Broad comparison | No new model proposed |

## III. PROBLEM IDENTIFICATION

The analyzed literature reveals the consistent presence of a number of constraints that prevent the practical implementation of the technology of AI-based lie detection. Firstly, the polygraph is still an invasive tool that requires human operators; secondly, the performance of human operators is no different from random guessing; thirdly, computational solutions that work well in controlled lab conditions are prone to degradation when introduced into real-life settings. Two expert consultations conducted within this research project provided even more proof of the necessity.

One of the HR professionals working for an international technology company pointed out that with the shift to online recruitment, it became much harder to judge candidates' body language, as the behavioral cues observable via video calls differ greatly from those available in person. One of the criminal investigators admitted that the success of interrogations was highly dependent on officers' experience and intuition. Thus, the main question of research is the following one:

How can facial micro-expression analysis based on deep learning and computer vision be developed to provide real-time, accurate, and environment-resistant deception detection overcoming the shortcomings of invasive measurements and static image processing?

## IV. GOALS AND OBJECTIVES

The overarching goal of the proposed work is to design and implement a real-time, non-invasive deception detection system that is computationally efficient and practically deployable. The following specific objectives guide the development:

- Develop a video acquisition and tracking pipeline using standard datasets and OpenCV, without requiring specialized hardware or physical sensors.
- Extract detailed facial landmark geometry and micro-expression features using MediaPipe's 468-point facial mesh, with particular emphasis on the mouth, nose, brow, and periocular regions.
- Train a CNN-LSTM hybrid deep learning model that captures both the spatial characteristics of instantaneous facial expressions and the temporal evolution of micro-expressions across consecutive video frames.
- Generate confidence-calibrated Truth/Lie predictions through a Softmax output layer, providing transparent probability scores rather than opaque binary classifications.
- Enhance model robustness through data augmentation covering illumination variation, pose transformation, horizontal flipping, and scale perturbation during training.

## V. PROPOSED METHODOLOGY

*A. Real-Time Video Acquisition and Face Detection*

The data pipeline starts off with real-time video stream capture via OpenCV. This is followed by identification of faces

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

of the subject using MediaPipe Face Detection module, which helps locate and track the face of the person within the spatial domain. Face region detection enables cropping and normalization of images with respect to a uniform resolution. Preprocessing tasks such as grayscale transformation, histogram equalization, and application of CLAHE technique are used to counteract effects of varying lighting conditions.

### *B. Facial Landmark and Micro-Expression Feature Extraction*

The MediaPipe Face Mesh provides 468 coordinates for landmarks corresponding to areas such as the eyes, eyebrows, nose, lips, cheeks, and jaw region within each face region detected. Geometric features based on these landmarks including inter-landmark distance measurements, angles, and symmetry measures are used as the main feature vector for classification of deception. In addition, CNN operations are performed on the facial image patches to identify high-level spatial features representing slight muscular movements within the face regions.

### *C. Temporal Learning via CNN-LSTM Architecture*

Since micro-expressions occur in a sequence of frames and not in just one still picture, understanding the dynamic nature of micro-expressions becomes crucial to ensure proper recognition. As the name implies, the architecture utilizes an LSTM-based model to analyze sequences of spatial feature vectors produced by the CNN and then detects patterns of changes in facial muscle activation throughout the entire process. The final CNN-LSTM feature vector gets classified using a fully connected Softmax layer which gives output values like Truth: 82%, Lie: 18%.

### *D. Robustness Enhancement Through Data Augmentation*

In order to train the model to be generalized beyond the lab environment, the training data set will include some artificially generated variations, such as random rotation (±15°), horizontal flip, variation in brightness and contrast, addition of Gaussian noise, and affine scaling. The aim of introducing these augmentations during training is that the CNN-LSTM model becomes familiar with different appearance characteristics in the training process, which would greatly alleviate the problem of performance loss in the case of deployment in an uncontrolled environment.

## VI. APPLICATIONS

The proposed system has direct applicability across several high-stakes domains:

- Criminal investigations and police interrogations, where the system serves as an AI co-analyst alongside experienced officers.
- Border security and immigration screening, enabling rapid behavioral assessment without physical contact.
- Online recruitment and virtual interview analysis, supporting HR professionals in evaluating candidate honesty.
- Fraud detection in financial institutions during high-value customer interactions.
- Online examination proctoring systems for real-time candidate monitoring.
- Behavioral and psychological research at large scale in ecologically valid settings.

## VII. CONCLUSIONS AND FUTURE WORK

In summary, this study has looked into the current AI deception detection literature, focusing on detecting deception via the analysis of facial micro-expressions in five representative works based on landmark classification using CNNs, ML algorithms' preprocessing pipeline, multimodal audio-visual model fusion, manual feature extraction, and a systematic literature review. Collectively, these reviewed articles have shown that facial micro-expressions contain a behavioral deceit signal, and deep neural networks vastly exceed both humans and conventional machine learning algorithms in detecting this signal under controlled conditions.

However, critical challenges still exist, including poor generalization to uncontrolled environments, lack of diverse datasets, and binary classification with no uncertainty scores. This paper's contribution includes a deployable CNN-LSTM network with facial landmark identification using MediaPipe, data augmentation, and output scoring using the Softmax function. Future research will focus on multimodal audio-visual fusion, transformer models for temporal modeling, and creating an API for deploying the trained model to the cloud.

## REFERENCES

[1] R. Kashyap, D. Bavkar, S. Rathore, A. Manhar, M. M. John, and J. G. Kotwal, "AI-Driven Micro-Expression Analysis for Deception Detection," in 2025 IEEE 5th Int. Conf. ICT in Business Industry & Government (ICTBIG), 2025, pp. 1–6.

[2] M. Hasanudin, Derisma, A. Wahab, Indrianto, E. R. Kaburuan, and S. Y. Dewi, "A Hybrid CNN-Machine Learning Framework for Deception Detection Using Facial Landmark Analysis," in 2025 13th Int. Conf. Orange Technology (ICOT), 2025, pp. 1–8.

[3] Z. T. Mohammed and I. O. Abdul Majeed, "Lie Detection Using Facial Micro-Expression Features with Machine Learning," MINAR Int. Journal of Applied Sciences and Technology, vol. 6, no. 3, pp. 50–59, 2024.

[4] M. De Marsico, G. Dionisi, and D. F. P. Stanco, "FTM: The Face Truth Machine—Hand-crafted Features from Micro-Expressions to Support Lie Detection," Computer Vision and Image Understanding, vol. 249, p. 104188, 2024.

[5] M. Sen and R. Deneckere, "Unmasking Lies: A Literature Review on Facial Expressions and Machine Learning for Deception Detection," in Proc. 28th Int. Conf. Knowledge-Based and Intelligent Information & Engineering Systems (KES 2024), 2024, pp. 1925–1935.

[6] A. K. Elango, V. Reddy, C. Padala, P. Badrinath, and P. Reddy, "Deception Discernment using Audio and Micro-expressions," in Proc. 2024 Int. Conf. IoT, Communication and Automation Technology (ICICAT), 2024, pp. 1498–1501.