IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.1510134

An AI-Powered Lie Detection System Using Facial Micro-Expressions

Mrs. Recharla Surendra Geethanjali[1], Aditya Prakash Masabinal[2], Bharath Kumar S C[3], C Yuvaraj[4], G Daewoo Sri Prasad[5]

Assistant Professor, Department of Computer Science and Engineering, K. S. School of Engineering and Management, Bengaluru, India[1]

UG Student, Department of Computer Science and Engineering, K. S. School of Engineering and Management, Bengaluru, India[2-5]

Abstract: Deciding whether a person is telling the truth matters in security screening, forensic questioning, recruitment interviews, and fraud inquiries, yet the available instruments are weak. The polygraph demands attached sensors and a trained operator, and unaided human judges perform only marginally above chance. Micro-expressions, which are involuntary facial movements lasting roughly a twenty-fifth to a fifth of a second, give a contactless cue, but they pass too quickly for people to follow reliably. This paper describes the design and prototype of a video-based deception analysis system built around the Real-life Deception Detection Dataset of 121 courtroom clips. It joins facial-landmark geometry with learned visual features, follows how both change over time with a recurrent network that carries an attention layer, and returns a Truth/Lie probability together with an account of the frames and facial regions that shaped it. The design rests on a short review of six recent studies and on informal consultations with a recruiter and a police officer. The upload, analysis, and explained-output interface of the prototype is working, whereas quantitative validation (accuracy, precision, recall, and F1-score) is still to be carried out, so no performance claim is made. The contributions are a documented architecture, an explicit evaluation protocol, and a discussion of ethical limits that treats the tool as support for a human examiner and not as a verdict.

Index Terms: lie detection, deception analysis, facial micro-expressions, MediaPipe, CNN-LSTM, attention mechanism, explainable AI

I. INTRODUCTION

Judging honesty is part of many professional routines. Investigators test statements made under questioning, recruiters weigh the confidence of candidates, and fraud analysts look for signs that an account has been invented. The traditional instrument, the polygraph, records physiological signals through attached sensors. It is therefore intrusive and costly, and its reliability is widely disputed. Observation by people is no better: human accuracy at spotting lies is close to 50%, which is the level of guessing.

Facial micro-expressions offer a different route. They are involuntary facial movements that are hard to hide and that the speaker does not intend to show. They typically last between one twenty-fifth and one fifth of a second, so they are easy to miss in a live conversation and in a recording watched at normal speed. A camera paired with a trained model, in contrast, can inspect each frame and measure small displacements of the lips, brows, and eyes that no observer could track by eye.

Published systems for this task differ widely. Some compute hand-crafted facial measurements and apply a threshold, some feed images to deep networks, and a few add speech. Our reading of the literature, summarised in Section II, suggests three recurring weaknesses: several models ignore the order of frames, few explain what drove their output, and results are often obtained in controlled settings. The system proposed here responds to these points. It works on recorded video from real trials, combines geometric and learned features, models time explicitly, and reports which frames and facial areas mattered.

The main contributions are as follows.

1. A fused feature representation that couples 468-point facial geometry with CNN-derived visual features for every frame.
2. A temporal classifier based on an LSTM with an attention layer, which produces a Truth/Lie probability and a ranking of influential frames.

 This work is licensed under a Creative Commons Attribution 4.0 International License

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.1510134

3. An explanation stage that links the prediction to specific timestamps and facial regions, and a web interface that presents it.
4. A planned evaluation protocol and an open statement of the ethical limits of the approach.

II. RELATED WORK

This section is a narrative summary of six recent studies, selected from the literature survey of the project. It is not a systematic review, and the selection is not exhaustive. Table I lists the studies with the strengths and limitations that the project records attribute to them.

Hasanudin et al. [1] measure the geometry of facial landmarks and classify deception with a CNN combined with SVM, random forest, and KNN models. They report very high accuracy and metrics that are easy to interpret, but they warn of possible overfitting and of tests run under controlled conditions. Mohammed and Majeed [2] take a lighter path, applying pre-processing, Sobel and Canny edge detection, and K-Means clustering to facial regions. The method is quick and simple, yet it relies on classical machine learning and does not learn from sequences. Elango et al. [3] merge facial action units from OpenFace with speech features and pass them to an LSTM; the reported accuracy is about 62%, and the authors point to imbalanced data and demanding pre-processing. De Marsico et al. [4] propose the Face Truth Machine, which analyses landmark dynamics, builds hand-crafted features, and flags deception through thresholds, giving alerts that an investigator can inspect and a low demand on hardware, although it adapts less well than deep models. Sen and Deneckère [5] review CNN, LSTM, random forest, 3D-CNN, vision-transformer, hybrid CNN-LSTM, and FACS-based approaches and identify open gaps. A further paper on AI-driven micro-expression analysis for deception detection [6] was consulted for general context only; no detail beyond its title and venue is relied on here.

TABLE I
STUDIES SUMMARISED IN THE RELATED-WORK REVIEW

Ref.	Technique	Strength noted	Limitation noted
[1]	CNN with SVM, RF, and KNN on facial-landmark geometry	Strong reported accuracy; readable facial measures	Possible overfitting; controlled tests
[2]	Pre-processing, Sobel, Canny, and K-Means on facial regions	Fast; simple to implement	Classical ML; no sequence model
[3]	OpenFace action units and audio features fed to an LSTM	Audio and visual cues combined	About 62% accuracy; imbalanced data
[4]	Landmark dynamics, hand-crafted features, threshold rule	Explainable alerts; low hardware needs	Less adaptable than deep models
[5]	Review of CNN, LSTM, RF, 3D-CNN, ViT, CNN-LSTM, and FACS methods	Structured comparison of models	Survey; proposes no system

Three observations follow. First, the spectrum runs from transparent, hand-crafted pipelines to opaque deep models, and the two ends rarely meet; the suggestion in [4] to combine hand-crafted and deep features is therefore an obvious direction. Second, models that learn sequences or add a second modality tend to need more data and more elaborate pre-processing, and the modest accuracy in [3] shows that added complexity does not guarantee better results. Third, explanation is handled mainly through hand-crafted or interpretable measures [1], [4], and frame-level explanation for a deep temporal model is not described in the records of these studies. The present design draws on all three points by fusing geometry with learned features, using an attention-equipped recurrent network, and giving every result an explanation.

Beyond deception work, a wider body of research deals with micro-expressions themselves. An updated review of trends, challenges, and solutions in micro-expression recognition is given in [7]. Shreve et al. [8] address the spotting of macro- and micro-expressions in long videos through spatio-temporal strain. CASME II [9] is an improved database of spontaneous micro-expressions with a baseline evaluation, and Pfister et al. [10] study how such spontaneous expressions can be recognised. Meyer and House [11] examine pre-processing and descriptor features for the same recognition task. These works supply background for the pre-processing and feature choices made here, although none of them addresses deception itself.

 This work is licensed under a Creative Commons Attribution 4.0 International License

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.1510134

III. PROBLEM IDENTIFICATION AND OBJECTIVES

A. Field Consultations

To anchor the requirements in practice, the team held two informal conversations: with a human-resources professional at an IT services company on 10 April 2026, and with a sub-inspector at a Bengaluru police station on 17 April 2026. These were not structured interviews and are not presented as a user study. The recruiter explained that virtual interviews conceal much of the candidate's body language, which makes confidence and honesty harder to judge. The police officer described suspects who often make false statements during questioning, with officers depending on personal experience and observation alone. Both accounts point toward a webcam- or video-based tool that assists, and does not replace, the human assessor.

B. Problem Statement

Polygraph tests and manual behavioural analysis are intrusive, slow, costly, and error-prone. People seldom detect deception reliably because micro-expressions last such a short time that the naked eye rarely catches them. A non-invasive, accurate, near-real-time, and explainable system that studies facial micro-expressions with deep learning is therefore needed for screening, interviews, and forensic use.

C. Objectives

1. Build a real-time deception-analysis system.
2. Raise detection accuracy by combining facial landmarks with micro-expression features.
3. Provide an explainable output that gives a confidence-based Truth/Lie prediction.
4. Remain robust under varied lighting and head poses.

IV. PROPOSED SYSTEM

A. Overview

The system has two phases, shown in Fig. 1. In the offline phase, a labelled set of video clips passes through pre-processing and augmentation, features are extracted, and a temporal classifier is trained; the resulting weights are saved. In the online phase, a user uploads a video through a web front end, a backend service passes it to the inference engine, and an explainability module turns the raw prediction into a readable result that is also logged in a database.

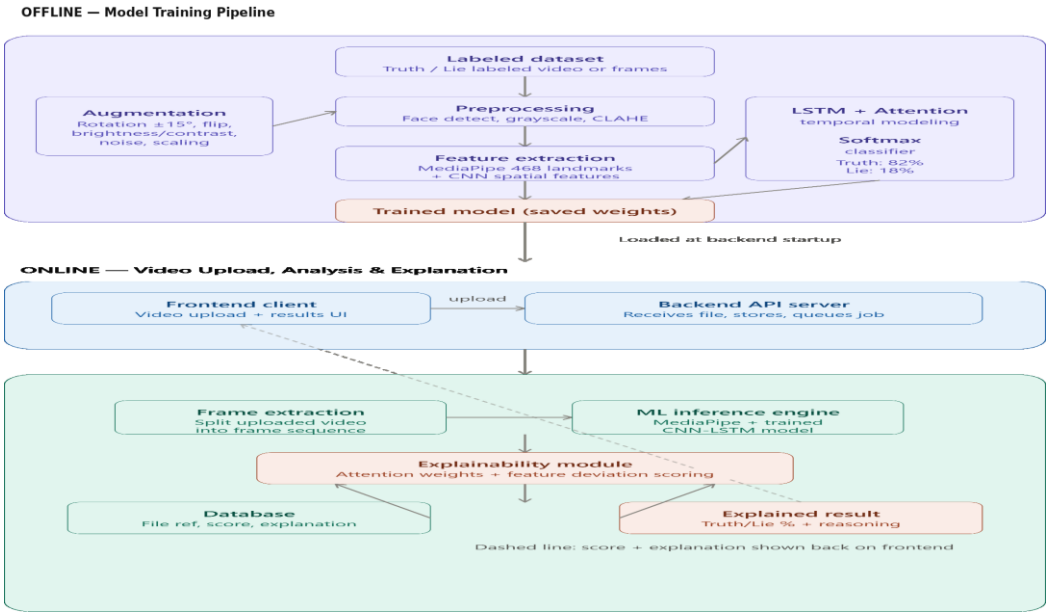


Fig. 1. Architecture of the proposed system, with the offline training pipeline above and the online upload, analysis, and explanation flow below.

B. Dataset

The system is built around the Real-life Deception Detection Dataset released by the University of Michigan. It holds 121 video clips, 61 deceptive and 60 truthful, taken from recordings of public court trials. The speakers are defendants or witnesses, and each clip is labelled from the trial outcome: guilty verdicts, not-guilty verdicts, and exonerations. These

 This work is licensed under a Creative Commons Attribution 4.0 International License



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026
DOI: 10.17148/IJARCCE.2026.1510134

labels follow the verdict and are not independently verified ground truth, since a verdict can be mistaken. The footage is unconstrained, so viewing angle, face size, lighting, and video quality all vary, which suits the aim of robustness. The collection is small and contains fewer distinct speakers than clips, so clips of one person must never appear in both the training and the test partitions. The clips are divided into training, validation, and test sets.

C. Pre-processing

Frames are read from each video with OpenCV at a fixed rate between 15 and 30 frames per second. A face detector isolates the facial region so that the background cannot distract the model. The crops are converted to grayscale and passed through contrast-limited adaptive histogram equalisation (CLAHE), which evens out lighting differences, and are finally resized and normalised to a common size.

D. Data Augmentation

Datasets labelled for deception are generally small, which invites overfitting. The training set is therefore enlarged with modified copies of each sample: rotation within ±15°, horizontal flipping, brightness and contrast changes, added noise, and rescaling. These variations push the network toward features that survive changes in pose and illumination and not toward memorised pixels.

E. Feature Extraction

Two feature streams are computed for every frame. The first uses the MediaPipe Face Mesh, which returns 468 three-dimensional landmarks. From them, geometric measures are derived: movement of the lip corners, the distance of brow raising, changes around the eyes, and how symmetric the two halves of the face are. The second stream is a CNN applied to the face crop, which learns texture and subtle muscle-movement patterns that are hard to describe by hand. The two outputs are concatenated into one vector per frame, so each frame carries geometric and learned cues.

F. Temporal Modelling and Classification

A single frame shows a static expression, whereas a micro-expression is defined by how quickly the face changes. The ordered sequence of per-frame vectors is therefore fed to an LSTM, which captures how an expression rises and fades. An attention layer placed on the LSTM outputs gives each frame an importance weight, so that the most informative frames contribute most to the summary of the clip. A final softmax layer then converts this summary into a probability for each of the two classes, Truth and Lie (for example, 82% against 18%). The attention weights are retained because they later identify the influential frames.

G. Training

The model is trained on the augmented data with a categorical cross-entropy loss. Early stopping, dropout, and learning-rate scheduling limit overfitting, and a held-out validation set is used periodically to tune hyperparameters. Once validation performance is satisfactory, the weights are saved. The backend loads them once at start-up, so no retraining is needed for individual requests.

H. Inference and Explanation

A FastAPI backend accepts the uploaded video, stores it, and processes it synchronously, which is sufficient at project scale because a clip takes only a few seconds. The frames are extracted with OpenCV and sent through the same MediaPipe and CNN pipeline used in training, followed by the LSTM, attention, and softmax stages. The explainability module then does two things. It uses the attention weights to list the frames and timestamps that influenced the decision, and it scores how far the brow, lip, eye-region, and symmetry measurements deviate, to show which facial change stood out most. The reference to the uploaded file, the confidence value, and the written explanation are saved together for later review, and the front end displays the probability and the reasoning behind it.

V. IMPLEMENTATION ENVIRONMENT

Table II lists the tools used. The front end is built with HTML, CSS, and JavaScript; Streamlit and Flask were considered as alternatives. The backend is written in Python with FastAPI, while Flask and Django were also considered. Model code uses TensorFlow/Keras, scikit-learn, OpenCV, and MediaPipe. Storage options evaluated for users, logs, and reports include MySQL, PostgreSQL, SQLite, and MongoDB. Development takes place in VS Code and Google Colab on Windows 10/11.

 This work is licensed under a Creative Commons Attribution 4.0 International License

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

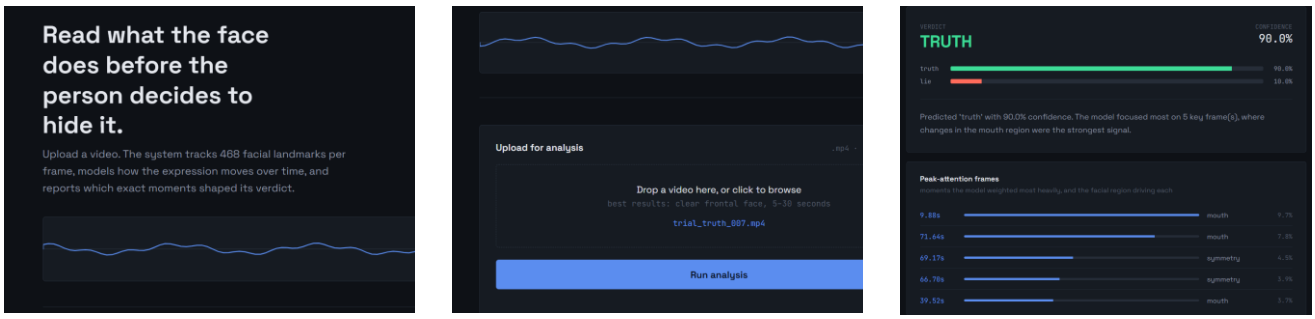
DOI: 10.17148/IJARCCE.2026.1510134

TABLE II
SOFTWARE AND HARDWARE REQUIREMENTS

Item	Specification
Front end	HTML, CSS, JavaScript (Streamlit, Flask)
Backend	Python, FastAPI (Flask, Django)
Database	MySQL, PostgreSQL, SQLite, MongoDB
ML libraries	TensorFlow/Keras, scikit-learn, OpenCV, MediaPipe
Tools, OS	VS Code, Colab; Windows 10/11
Processor	Intel Core i5 or i7
Memory, disk	8 GB RAM or more; 256 GB minimum

VI. PROTOTYPE INTERFACE AND RESULTS

Fig. 2 shows the prototype. The landing page (Fig. 2a) explains that the system tracks 468 facial landmarks per frame, models how the expression changes over time, and reports the moments that shaped its verdict. On the upload page (Fig. 2b), the user drops or selects a file in .mp4, .mov, .avi, or .mkv format and starts the analysis; the page advises a clear frontal face and a clip lasting 5 to 30 s.



(a) Landing page (b) Video upload (c) Analysis output

Fig. 2. Screens of the prototype: (a) landing page, (b) video-upload mechanism, and (c) explained output.

The output screen (Fig. 2c) shows the predicted class, here Truth, with a confidence of 90.0% and bars for both classes (90.0% and 10.0%). A sentence explains that the model concentrated on five key frames in which changes around the mouth gave the strongest signal. A second panel lists the five frames with the highest attention weight, from 9.7% down to 3.7%, and names the facial region behind each: mouth for three frames and symmetry for two. A footer states that results are logged locally and that the tool supports human judgement and does not replace it.

This example shows what the interface returns for one clip. It is not an accuracy measurement, and no system-level performance is claimed in this paper. The dataset is described in Section IV-B, but measured classification scores are not reported here.

VII. EVALUATION PLAN

Testing will run on videos unseen during training. The planned measures are accuracy, precision, recall, F1-score, and a confusion matrix, together with the false-positive rate, since wrongly labelling an honest person as deceptive is the costliest error in practice. Robustness will be probed by varying lighting and head pose. The protocol should be fixed before any run and should state the number of subjects, clips, and sources. Because the dataset has fewer speakers than clips, the split must be subject-independent. A cross-dataset test is also needed to expose overfitting of the kind flagged in [1], and ablations should compare landmark-only, CNN-only, and fused features, with and without attention. Runs should be repeated and reported with their spread, and processing time per clip should be recorded. Whether the highlighted frames actually matter to the decision can be checked by masking them and observing the change in output. No target figure will be claimed as achieved before it is measured.

 This work is licensed under a Creative Commons Attribution 4.0 International License



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026
DOI: 10.17148/IJARCCE.2026.1510134

VIII. ETHICAL CONSIDERATIONS AND LIMITATIONS

A deception score can harm a real person if it is over-trusted. Micro-expressions reveal concealed emotion, which is related to, but not the same as, lying; nervousness, fear, or cultural habits can produce similar movements in a truthful speaker. The output must therefore be handled as a lead for a human examiner and never as evidence on its own. Recordings should be taken with the subject's informed consent and stored securely. Performance may differ across skin tones, ages, and recording conditions, which the evaluation must measure and report.

The present work has further limits. The prototype handles uploaded videos and not a live camera feed, audio is not used, and the consultations summarised in Section III were informal. The dataset is small, which raises the risk of overfitting, and its labels come from trial outcomes and not from verified truth. The review in Section II covers six studies and cannot be taken as complete. Finally, quantitative validation is outstanding, so conclusions about accuracy would be premature.

IX. CONCLUSION AND FUTURE WORK

This paper presented a video-based system that estimates truthfulness from facial micro-expressions. It fuses MediaPipe landmark geometry with CNN features, models their evolution with an LSTM and attention, and returns a confidence-based Truth/Lie output with an explanation tied to frames and facial regions. The prototype interface demonstrates the full flow from upload to explained result. The next step is a documented evaluation under the protocol of Section VII, since stronger claims should wait for measured evidence.

Two extensions are planned. The first is real-time analysis of a webcam stream, so that the system can follow live interviews and conversations. The second is multimodal analysis that adds speech, voice characteristics, and textual answers to the facial cues, which the surveyed literature suggests can give a broader picture of deceptive behaviour.

REFERENCES

[1] 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 *Proc. 13th Int. Conf. Orange Technology (ICOT)*, 2025, pp. 1–8.

[2] Z. T. Mohammed and I. O. Abdul Majeed, "Lie detection using facial micro-expression features with machine learning," *MINAR Int. J. Appl. Sci. Technol.*, vol. 6, no. 3, pp. 50–59, 2024.

[3] A. K. Elango, V. Reddy, C. Padala, P. Badrinath, and P. Reddy, "Deception discernment using audio and micro-expressions," in *Proc. Int. Conf. IoT, Communication and Automation Technology (ICICAT)*, 2024, pp. 1498–1501.

[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," *Comput. Vis. Image Underst.*, vol. 249, p. 104188, 2024.

[5] M. Sen and R. Deneckère, "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, pp. 1925–1935.

[6] R. Kashyap, D. Bavkar, S. Rathore, A. Manhar, M. M. John, and J. G. Kotwal, "AI-driven micro-expression analysis for deception detection," in *Proc. 5th IEEE Int. Conf. ICT in Business Industry & Government (ICTBIG)*, 2025, pp. 1–6.

[7] K. M. Goh, C. H. Ng, L. L. Lim, and U. Sheikh, "Micro-expression recognition: An updated review of current trends, challenges and solutions," *Vis. Comput.*, vol. 36, pp. 445–468, 2020.

[8] M. Shreve, S. Godavarthy, D. Goldgof, and S. Sarkar, "Macro- and micro-expression spotting in long videos using spatio-temporal strain," in *Proc. IEEE Int. Conf. Automatic Face & Gesture Recognition (FG)*, Santa Barbara, CA, USA, 2011, pp. 51–56.

[9] W. J. Yan, X. Li, S. J. Wang, G. Zhao, Y. J. Liu, Y. H. Chen, and X. Fu, "CASME II: An improved spontaneous micro-expression database and the baseline evaluation," *PLoS ONE*, vol. 9, no. 1, p. e86041, 2014.

[10] T. Pfister, X. Li, G. Zhao, and M. Pietikäinen, "Recognising spontaneous facial micro-expressions," in *Proc. Int. Conf. Computer Vision (ICCV)*, Barcelona, Spain, 2011, pp. 1449–1456.

[11] R. Meyer and C. House, "Preprocessing and descriptor features for facial micro-expression recognition," Stanford Univ., Stanford, CA, USA, Tech. Rep., 2015.

 This work is licensed under a Creative Commons Attribution 4.0 International License