Grenze International Journal of Engineering and Technology, June Issue

Lie Detection using Facial Expressions and Voice Recognition

R. Mokshgna Krishna Koushik[1], Preetham Reddy PVS[2], Kanith Kumar P V[3], Sane Rohith Reddy[4] and Dr. Siavagamasundari G[5]
[1-5]Dept. of Computer Science and Engineering PES University, Bengaluru, India
Email: mokshagnar698@gmail.com, preethamsaireddy1214@gmail.com, kanithpathakamuri46@gmail.com, sanerohithreddy@gmail.com, sivagamasundari@pes.edu

***Abstract*— Lie detection is a critical area of study, especially within security and forensic applications. This project presents a multimodal lie detection system that analyzes facial micro expressions and vocal patterns to improve the accuracy of deception detection. The system leverages Long Short-Term Memory (LSTM) networks to process sequential data from facial expressions, detected via OpenFace, and audio features extracted through Librosa. Integrating these modalities allows the model to detect subtle cues in both the face and the voice, achieving an accuracy of 85 percent. By employing deep learning on multimodal data, this project advances traditional methods of lie detection by offering a noninvasive, automated approach. The system's high accuracy and real-time performance hold potential for real-world applications in interview analysis, security screening, and mental health diagnostics.**

***Index Terms*— Lie Detection, Facial Expressions, Real time, Voice Recognition, Machine Learning, Deception Detection, Feature Selection.**

I. INTRODUCTION

Detecting Deception is a complex challenge with significant implications in areas ranging from security and law enforcement to mental health and human resources. Traditional methods such as polygraph and body camera tracking, while useful, are nonetheless useful it often involves running techniques and can produce unreliable results due to factors such as anxiety or discomfort. Besides, these methods often drag attention focuses on controlling or masking physiological responses, limiting their effectiveness to real-world situations where biological reactions are needed.

Lie detection has long been an area of intense research due to its widespread use in law enforcement, psychology, security testing, etc. Traditional lie detection methods tend to be physiologically based cues such as heart rate, sweating, and dilated eyes, which can be intrusive, slow or difficult to interpret. Allows for non-invasive absorptive techniques, attention is focused on subtle cues in human behavior, especially facial expressions and vocal patterns.

This work introduces a multimodal system that combines subtle facial expressions and voice analysis for better fraud detection. Microexpressions are short, involuntary facial expressions that can convey hidden emotions, while vocal features such as tone, intonation, and rhythm can indicate tension or uncertainty. By exploring these cues together, we aim to increase the reliability of lie detection. Our model uses long-term and short-term memory (LSTM) networks, which are well suited for sequential and time-dependent data processing.

Grenze ID: 01.GIJET.11.2.592_2

The system uses OpenFace to capture and interpret subtle expressions, providing information on facial markings and muscle movements. At the same time, Librosa extracts audio features that reveal nuances in language structure. By combining these two data streams, the model can detect fraud with high precision. Unlike methods that analyze a single communication channel, this method uses the overlap of facial and vocal data to perform a detailed analysis. Our preliminary results show an accuracy rate of 85 percent, suggesting that the combination of facial and voice information can significantly improve lie detection. This study demon- strates the potential of a noninvasive, AI-powered lie detection tool that can be used in high-risk environments, from face-to- face screening to security screening to medical settings on which understanding of underlying emotions is critical. The main goal of this work is to form an independent system, work with an unlimited database collection environment, and have no physical connection to it in the human body.

II. Related Work

Deception Detection System(DDS) It is important to in- dividuals without researchers. It is possible to find someone Deception is based on an automated facial screening process if recognition of signs of fraud. Creating computer vision system towards deception detection based on facial expressions can can be divided into three important stages. The first step is video capture and pre-processing that is video recording to the features extraction stage, this stage involves video recording of the subject's face during the interview. Editing captured video, face recognition, and Vital facial landmarks are searched in facial images to eliminate facial symptoms. Second, feature extraction From the front image sequence, this phase uses it for removal Facial cues representing potential clues to Cheating. Finally Qualified Decision Maker The classifier was used based on the participant's condition.

A. Video Capturing and Pre-processing

The process begins with recording videos of the client's facial expressions. The high-definition camera is positioned to focus on the subject's face, ensuring that subtle details can be spotted. The lighting conditions are optimized to reduce shadows and glare, as these can obscure important details in facial movement. The system constantly captures video, breaking it down into individual images for analysis.

Once the face in each frame is detected, OpenFace is used to normalize the face to a neutral, correct position in a step called face normalization. This system standardizes facial expressions across frames, allowing continuous analysis of facial expressions without interference from head rotation or rotation Furthermore, frames are sampled at intervals to consume data process efficiently, and appropriately sample microexpressions while avoiding redundancy.

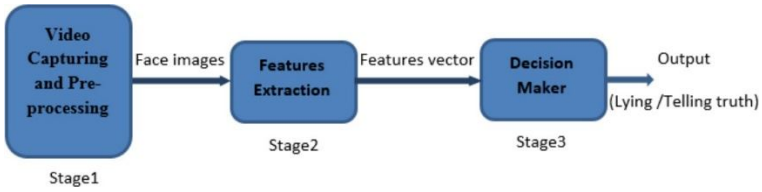

Fig. 1. "General block diagram illustrating the architecture of deception detection systems utilizing facial expression analysis."

Once sorted, basic preprocessing is applied to the images. This often includes a grayscale transformation, which reduces computational burden by focusing only on light information and avoiding color information Grayscale images allow facial markings and emotions to be seen more clearly, since shadows are the highlights is more obvious. At this point, each frame is ready for feature extraction, with key areas of the face normalized and enhanced for microexpression recognition.

B. Feature Extraction

Feature extraction is the stage where the system analyzes and extracts meaningful facial expression and voice information. The data is divided into two categories: visual and auditory. Visual information is extracted from the facial stimuli in each image, while auditory information is acquired from the voice system in real-time speech.

For facial features, OpenFace identifies and tracks specific facial features, including eyes, eyebrows, mouth, and cheeks. These markers correspond to subtle nerves capable of revealing hidden emotions. Transient, involuntary microexpressions are particularly valuable because they can reveal the truth behind deceptive responses. OpenFace records changes in these landmarks over time, creating a temporal snapshot of facial movement.

These temporal elements are arranged in sequences that can subsequently be processed by the LSTM network to detect patterns of deception.
At the same time, audio is captured and processed using the Librosa library of voice feature filters. These include tone, volume, intensity, and rhythm, all of which can indicate tension or hesitation, which are common signs of fraud. Librosa's filtering techniques focus on time-frequency-dependent audio characteristics, unconsciously capturing changes in pitch, vibrations, and irregularities in speech structure on.

C. Decision Maker

The decision-phase determines the probability of construction by combining visual and auditory information. Two pieces of information—sequential facial expressions and vocal segments—are transferred to the short-term memory (LSTM) network. LSTMs are more effective for this purpose because they contain sequences of information, allowing the model to detect temporal patterns that occur over time, such as the occurrence of vocal tension or the repetition of facial displacements.
The LSTM network is trained on labeled data containing both fraudulent and non-fraudulent information, allowing it to discriminate between true information and lies The model gives more importance to signals, such as sudden changes in voice tone inside or subtle short statements of concern Specifics learns to associate patterns with deceptive or truthful behavior.
Once the inputs have been processed, LSTM provides an output indicating the probability of cheating. This process is then translated into a confidence score, where a higher score indicates a greater likelihood of being deceived. In real-time applications, this determination is continuous, enabling the system to actively monitor possible fraud while the subject is talking.
This multi-way decision-making by combining facial and voice cues increases the robustness and accuracy of the model, as it relies on multiple independent indices of deception Through subtle adjustments facial and voice capture, the sys- tem offers a comprehensive, non-intrusive, and highly accurate solution for lie detection in real-world situations.

III. METHODOLOGY

This lie detection system approach includes various tech- niques including data acquisition, pre-processing, feature ex- traction, and classification By combining visual and auditory signals, the system is designed to capture and identify complex behavioral cues associated with deceptive behavior under.

A. Datasets Collection

We use the Bag of Lies data set, which is a rich repository of human complexity Cheating is telling the truth. This edited collection captures the responses of 35 people Different users, each offering their unique perspective through a series of search queries. Delving into the nuances of human behavior, this series works in an interesting way An examination of honesty, deception, and the shadow of truth in between

B. Data Acquisition: Video and Audio Capture

The system begins with simultaneous video and audio recording, collecting data that will serve as the basis for fraud detection. Video capture focuses on the subject's facial expressions, capturing specific high-resolution images that subtly identify his or her facial features—short involuntary facial captures associated with emotions that hidden around This is done through a high-definition camera positioned to capture the face, ensuring that minimal interference from head movement occurs. At the same time, audio is recorded with a microphone that can indicate stress or anxiety.

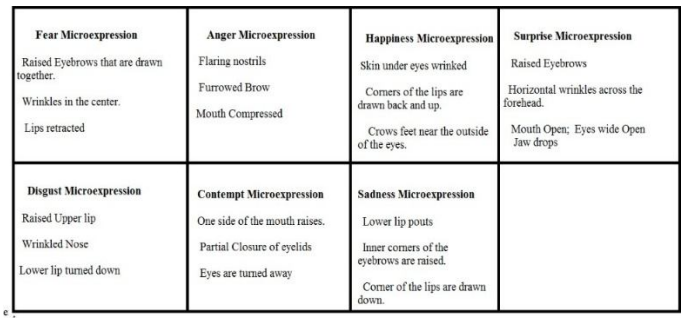

Fig. 2. "Illustrative summary of microfacial cues associated with deception, as studied in behavioral analysis."

The quality of audio and video data is critical to the success of subsequent segments. The environment has been optimized for maximum reliability, with controlled lighting and minimal background noise. This ensures audio and visual clarity, reducing the risk of misinterpretation during feature removal.

C. Pre-processing of Visual and Auditory Data

Once acquired, video and audio data are first processed to prepare them for feature extraction. The goal here is to normalize the data to improve quality, ensure accuracy, and improve the model's ability to detect cues associated with fraud.

For video data, frames are extracted periodically, balancing the need for detailed analysis with computational effort.

OpenFace, an open-source facial behavior analysis tool that normalizes the subject's face to a neutral position in the frame Known as normalization, this process of normalization is then provided consistent collision of facial features followed to compensate for small head movements or tilts does Furthermore, the frames become grayscale, complication Reduces brightness and focus on, allowing subtle facial features such as wrinkles, grooves and slight movement to be detected.

Audio preprocessing is done in parallel. Audio data are filtered using the Python library Librosa for audio-track anal- ysis to remove background noise and normalize pitch. This library extracts necessary audio features in preparation for the feature extraction phase, ensuring that excessive noise does not interfere with the recognition of valid vocal signals.

D. Feature Extraction: Facial and Vocal Cues

Feature extraction is a basic process in which meaningful information is derived from pre-processed data, where facial and vocal cues help to accurately identify the target's behavior.

To extract facial features, OpenFace recognizes and tracks specific facial landmarks such as eyes, eyes, nose, mouth, and cheeks. These markers represent key areas of microexpressions. The system monitors frame-rate movements of these facial regions, detecting minute, involuntary movements that are often associated with hidden emotions. For example, a sudden movement of the mouth or a rapid tightening of the eye muscles can indicate stress or dishonesty. The information from these landmarks is then organized into sequences that capture the temporal aspect of facial expression, allowing the system to recognize patterns that occur over time.

At the same time, Librosa processes the audio data and extracts voice features including pitch, pitch, intensity, and speed of speech. For example, changes in tone are often associated with stress or tension, while changes in tone or energy can reflect emotional instability. The temporal dimension is again key, as gradual changes in these vocal components can reveal underlying fear or deception. By focusing on these dynamic features rather than permanent traits, the system develops a nuanced understanding of the subject's emotional state.

E. Data Integration and Feature Fusion

To maximize detection accuracy, the system adds visual and hearing properties to a single multimodal information set. This approach, known as Feature Fusion, integrates facial landmarks and vocal properties in integrated data structure, and captures both views and hearing aspects of subject be- havior over time. By combining these currents, representation achieves rich information, as it can identify correlations and interactions between facial expressions and vocal signals that are signs.

Temporal alignment of audiovisual data ensures correspon- dence between facial expression and vocal changes, crucial for deception detection. Inconsistencies between physical be- havior and speech are key indicators of deceit.

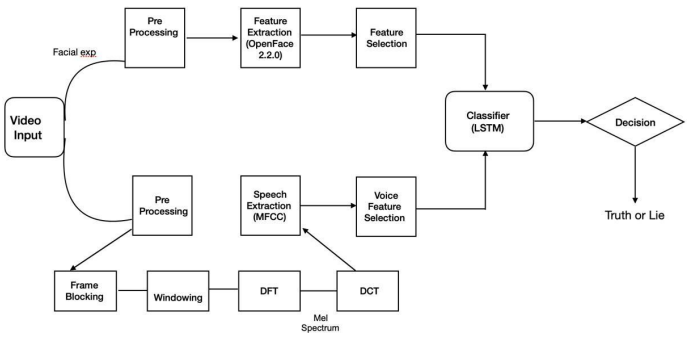

Fig. 3. "Architecture of a multimodal deception detection system utilizing facial expression and speech analysis."

F. Classification Using Long Short-Term Memory (LSTM) Networks

The combined data is then passed to Long Short-Term Memory (LSTM), which is chosen for its ability to manipulate data sequentially over time. LSTM is particularly useful in this regard. This is because it stores information about long connections.

This makes it possible to detect emerging patterns that may indicate fraud.

The LSTM model is trained on a labeled dataset consisting of real models and fraudulent models. during training The model learns to classify specific combinations of facial and audio signals as false or real. The main features include sudden inflections of sound. quick comma expression Or the content of words and facial expressions do not match. The model develops an extensive memory of these patterns through repeated learning. To be able to distinguish between misleading responses and accurate responses.

In real-time processing, LSTM processes input data and outputs authentication levels that indicate potential fraud. This probability is obtained by evaluating the weighted importance of the various indicators in the equation, giving higher weight to those most relevant to fraud.

G. Decision Making and Output Interpretation

The final output from the model is a deception probability score, which ranges from 0 to 1. This score represents the system's confidence in whether the subject is lying, with scores closer to 1 indicating a higher probability to be deceived Depending on the application, this score as a direct binary outcome (e.g., "cheat" or "deceiver.") or can be expressed as a level of confidence, whereby Users receive an assessment of the validity of the model.

The decision-making process is designed to be interpretable. It identifies specific facial and speech signals that contribute to fraud scores. Such an understanding is critical to the ap- plication's usability. This is because it helps users understand why a topic was flagged as potentially fraudulent. This will increase confidence in the system's results.

IV. RESULTS

It captures microexpressions and speech texture features in real-time to determine the present emotional state of the subject and identify deception. It keeps track of ups and downs in emotion by using OpenFace for visual stimulus measurements and Librosa for speech measurements that depend on other physiological features that might indicate deceitful behavior. The system offers real-time feedback and an analysis of emotional trends making it useful for security applications.

Evaluated on a labeled dataset, our hybrid LSTM-based based model on a labeled dataset, outperforming unimodal approaches based on facial expressions (78%) or voice analysis (80%). In contrast to SVM (72%) and Random Forest (74%), our hybrid model achieved better accuracy via a combination of multiple modalities, reducing false positives and increasing confidence in classification. Some challenges were overlapping emotion expressions, speech variability, and light conditions, which affected performance in some cases. Future development will center on working with transformer-based models and adaptive feature extraction methods to further improve robustness and accuracy.

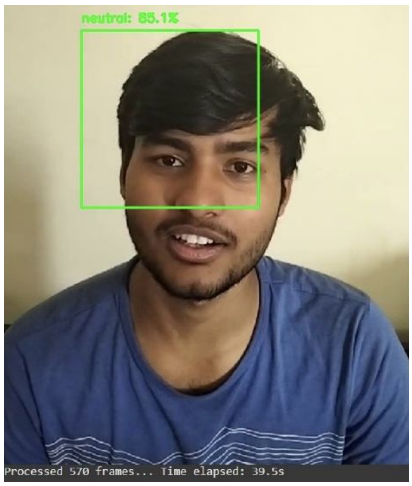

Fig. 4. "Facial microexpression analysis frame extracted from the dataset, showcasing real-time emotion recognition for deception detection."

V. CONCLUSION

This research proposes a comprehensive and multimodal deception detection strategy that incorporates facial micro- expression analysis and speech-based acoustic features. Contrary to the traditional lie-detecting methods that have been marred by invasiveness and ambiguity, the system used advanced deep learning techniques to capture the subtle behavioral signals associated with deception. It integrates facial expression analysis using OpenFace and speech feature extraction through Librosa into a steadier assessment of a person's emotional and cognitive state. Real-time processing assures the speedy detection of results and offers the potential for application in security screenings, psychological assessments, and interview-based deception analyses. The emotion distribution analysis offered by the chart visualizing the examined emotional response furthers the interpretability of this approach by exhibiting how various emotionally charged states are related to deceptive or truthful responses.

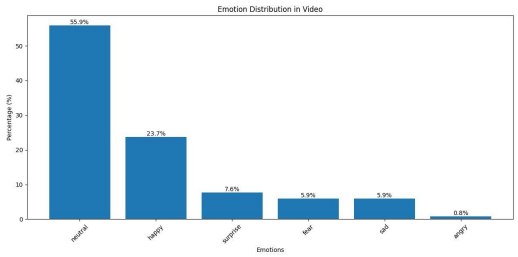

Fig. 5. "Emotion distribution chart representing the detected microex- pressions during deception analysis."

The LSTM-based integrated model achieved an 85% accu- racy rating, surpassing unimodal methods focusing on either facial expressions (78%) or acoustic analysis of speech (80%). Compared to traditional machine learning methodologies such as SVM (72%) and Random Forest (74%), the deep learning model outperformed in terms of robustness in classification performance with fewer false positives and increased confi- dence levels. Yet, more work is needed to handle variations in speech patterns, overlapping emotional expressions, and environmental factors such as lighting. Future work should focus on improvements in robustness through transformer- based architectures and adaptive feature extraction techniques. This work contributes to the increased adoption of behavioral AI in demonstrating that multimodal deception detection has great promise for heightening the overall understanding of automated human behavior analysis.

VI. FUTURE WORK

Future work could explore how to extend the flexibility of the system to different areas, increased definition, and add additional physiological cues to increase detection accuracy Such improvements would improve the deception detection capabilities of the system has succeeded with greater reliability and ethical clarity in real-world applications.

A. Integration of Additional Physiological Signals

If it extends beyond subtle facial expressions and vocal cues, future iterations of this system may integrate physiological cues such as heartbeat, eye dilation, and skin perfusion These cues are common is associated with stress and fear, providing valuable information for detecting deception. Remote photo- plethysmography (rPPG) can be used to measure heart rate from video, increasing the system's sensitivity to the subject's mental state without the need for physical sensors.

B. Robustness in Unconstrained Environments

Although the current system works well on a controlled schedule, it would expand its usefulness if it were optimized for more dynamic environments. With noise reduction in audio, shadow control in video, and training in different envi- ronments, the system can achieve more reliable performance in real-world applications such as an interview or security demonstration with varying lighting, background noise and camera angles.

C. Enhanced Model Interpretability

To provide transparency and trust, future iterations could incorporate interpretable AI techniques. Through attention or presentation techniques, the system can identify the specific cues (e.g., sudden changes in volume or specific facial stimuli) that influenced its decision Thus this definition is important in applications where users need to understand the logic of the model In and believe in capital investment programs.

REFERENCES

[1] B. Singh, P. Rajiv and M. Chandra, ”Lie detection using image processing,” 2015 International Conference on Advanced Computing and Communication Systems, Coimbatore, India, 2015, pp. 1-5, doi: 10.1109/ICACCS.2015.7324092.

[2] Barathi, C. Soumya. ”Lie detection based on facial micro expression body language and speech analysis.” International Journal of Engineer- ing Research Technology 5.2 (2016): 337-343.

[3] Bareeda, E Mohan, B Muneer, K. (2021). Lie Detection using Speech Processing Techniques. Journal of Physics: Conference Series. 1921. 012028. 10.1088/1742 6596/1921/1/012028.

[4] M. -T. Tran-Le, A. -T. Doan and T. -T. Dang, ”Lie Detection by Facial Expressions in Real Time,” 2021 International Conference on Decision Aid Sciences and Application (DASA), Sakheer, Bahrain, 2021, pp. 787- 791, doi: 10.1109/DASA53625.2021.9682390.

[5] A. A. Perdana et al., ”Lie Detector with Eye Movement and Eye Blinks Analysis Based on Image Processing using Viola-Jones Algorithm,” 2021 IEEE International Conference on Internet of Things and Intel- ligence Systems (IoTaIS), Bandung, Indonesia, 2021, pp. 203-209, doi: 10.1109/IoTaIS53735.2021.9628413.

[6] Rastogi, Rohit Anand, Tushar Sharma, Shubham Panwar, Sarthak. (2023). Emotion Detection via Voice and Speech Recognition. Inter- national Journal of Cyber Behavior, Psychologand Learning. 13. 1-24. 10.4018/IJCBPL.333473

[7] A. Akyakı, B. Karaca, D. Altınel and F. Gu¨rkan, ”Image Processing Based Lie Detector Model,” 2024 Innovations in Intelligent Systems and Applications Conference (ASYU), Ankara, Turkiye, 2024, pp. 1-5, doi: 10.1109/ASYU62119.2024.10757042.

[8] M. A. Khalil, M. Babinec and K. George, ”LSTM Model for Brain Control Interface Based-Lie Detection,” 2024 IEEE First Interna- tional Conference on Artificial Intelligence for Medicine, Health and Care (AIMHC), Laguna Hills, CA, USA, 2024, pp. 82-85, doi: 10.1109/AIMHC59811.2024.00024.

[9] B. Wang and X. Shen, ”The Application of Machine Learning in Deception Detection : a study based on the effect of empathy on deception detection,” 2022 4th International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI), Shanghai, China, 2022, pp. 234-237, doi: 10.1109/MLBDBI58171.2022.00052.

[10] M. -T. Tran-Le, A. -T. Doan and T. -T. Dang, ”Lie Detection by Facial Expressions in Real Time,” 2021 International Conference on Decision Aid Sciences and Application (DASA), Sakheer, Bahrain, 2021, pp. 787- 791, doi: 10.1109/DASA53625.2021.9682390.