Kufa Journal of Engineering
Vol. 16, No. 2, April 2025, P.P. 180 -196
Article history: Received 16 July 2024, last revised 20 September 2024, accepted 20 September 2024

3D FACIAL LANDMARK-BASED DECEPTION DETECTION IN VIDEO USING GRU MODEL

Amira Abbas Hussein [1] and Israa H. Ali [2]

[1] Master's degree Student, The University of Babylon, College of Information Technology, Software Department, amiraabbash.sw@student.uobabylon.edu.iq

[2] Professor, The University of Babylon, College of Information Technology, Software Department, Israa_hadi@itnet.uobabylon.edu.iq

https://doi.org/10.30572/2018/KJE/160211

ABSTRACT

Deception detection is an interdisciplinary field that has researchers from psychology, criminology, and computer science. We propose the automated detection of deception based on facial micro expressions which occur spontaneously in response to the attempt to mask the inner emotion. It has received significant attention as an indicator of deceit, it reveals the genuine emotions that are concealed. In this paper, we first proposed a 3D 478 Mediapipe Face Mesh Model to extract facial landmarks that reflect facial micro expression, this is contrary to the traditional method, which relies on human judgment and the use of devices to detect facial micro expression. Second, a feature selection-based multivariate mutual information method was proposed to select facial landmarks that are most related to the deceptive cues and have critical influence on the classification task. Finally, a gated recurrent unit model was trained to predict deceptive behavior on a real-life trial dataset. The model successfully achieved 97% accuracy, outperforming other state-of-the-art methods.

KEYWORDS

Facial micro expression, Facial action coding system, Mediapipe face mesh model, Feature selection-based multivariate mutual information method, Gated recurrent unit (GRU) model.

1. INTRODUCTION

Deception is a complex social behavior in which the deceiver attempts to influence others (i.e., the deceived person) by changing their perception of the situation to make them more consistent with the deceiver's views and behaviors (Jakubowska and Białecka-Pikul, 2020). Lying can damage relationships and hinder communication, which can have harmful consequences. Therefore, deception detection is a key component in many fields, such as healthcare, court trials, and security (D'Ulizia et al., 2023). Deception detection involves determining whether a particular communication contains truth. It is an active and evidence-based reasoning process (Levine, 2014). Lie detection has been the subject of intensive research for decades, with deception researchers primarily focused on detecting deception automatically without the use of special equipment or based on human judgment, such as Facial Action Units (FACS). Because humans have a limited ability to detect deception (Monaro et al., 2022), the average accuracy of lie detection without special aids is reported to be 57%, which is only slightly better than chance. Even physiological methods such as polygraphs or newer methods based on functional magnetic resonance imaging (fMRI) do not always correlate with deception (Farah et al., 2014). Furthermore, the usefulness of these devices for real-life deception detection is limited by the cost of the equipment and the overt nature of the methods. The importance of digitalization and machine-learning-based approaches has also become critical in multiple disciplines (Mohammed, Kareem and Mohammed, 2022) (Alaa, Hussein and Al-libawy, 2024). Machine learning methods have the ability to automate the detection of deception, utilizing multiple methods and possessing multiple modes of information (Prome et al., 2024). Many indicators are used to differentiate liars from truth tellers; these include verbal indicators like voice and text analysis, as well as non-verbal indicators like facial expressions and body movement (D'Ulizia et al., 2023). One of the most significant indicators of a liar is the analysis of the micro expressions on the face. Micro-expressions on the face are non-verbal signals that are instantaneous and are also short-lived. These signals are often undetected by untrained observers because of their short duration and low intensity (Verma et al., 2019). They express the feeling that the individual is attempting to hide. These involuntary facial movements are important because they can expose crucial information about the person's actual emotions. Because of their nature, these micro expressions have difficulty in their recognition, which necessitates the measurement and analysis of facial movements. In this case, landmarks on the face are critical. Face landmarks are spatial features that represent significant facial locations; these include the center of the chin, the eye's corner, the nose's tip, and the eye's corner. The locations of the landmarks' faces can indicate alterations to the facial muscles, which are

represented by the micro expressions of the face. The landmarks on the face are categorized into three groups based on their number: sparse, moderate, and dense. Sparse models have fewer than 50 landmarks, moderate models have 50-100 landmarks, and dense models have over 100 landmarks (Chen et al., 2015). In comparison to the 2D method that extracted 68 facial landmarks, the 3D method that extracted over 100 facial landmarks can include a larger variety of the facial landmarks, including its emotional state, position, occlusion, and lighting conditions (Jabberi et al., 2023). Therefore, the key contributions of our approach include the following:

1. Proposing a deep learning technique to extract 478 facial landmarks with three coordinates (x, y, z) per frame for each video in the used dataset to analyze and capture a wider range of facial micro expression for more accurate full automated deception detection and minimize human error and subjectivity.
2. Proposing a robust method that is capable of accurately measuring the degree of the mutual information between features to identify the most relevant cues of deception.
3. Proposing a deep learning model that can handle sequential data and accurately predict deceptive tellers from truth tellers.

The paper organizes its subsequent sections as follows: Section 2 presents a review of the pertinent literature on deception detection, with a focus on deep learning techniques. In Section 3, the methodology is explained. This includes the dataset and how it was preprocessed, as well as 3D facial landmark extraction, feature selection, and prediction using the GRU model. In Section 4, the proposed method's performance is fully discussed, with evidence from experiments. Finally, Section 5 presents the study's conclusion.

1.1. Aim of the Study

Detecting deception is particularly crucial in scenarios such as security screenings, police interrogations, and courtroom testimonies, where the consequences of deceit can be severe. Therefore, the main aim of this study is to propose an efficient system for detecting liars using deep neural networks. The system aims to scale and analyze facial micro expressions, automatically extracting and selecting 3D facial landmarks (features) in a way that surpasses the human capability of identifying deceit.

2. LITERATURE REVIEW

To accurately detect deception, researchers used a variety of different methodological techniques, such as (Yildirim, Chimeumanu and Rana, 2023) developed a deep learning model that achieved a 74.17% accuracy in classifying deception based on micro-expressions. Moreover (Tsuchiya, Hatano and Nishiyama, 2023) use machine learning in detecting

deception through facial expressions and pulse rate achieving an accuracy and F1 value of 0.75-0.8. While (Islam et al., 2021) (Shen et al., 2021) and (Stathopoulos et al., 2023) identify deceit from the subject's natural response to truth and lie by evaluating Facial Action Units (FACS), which is a taxonomy of human facial muscle movements based on how they appear on the face and splits the face into multiple action units (AUs), which are basic movements caused by a single muscle or a group of muscles in response to a face expression (Martinez et al., 2017). Where (Islam et al., 2021) achieved accuracy of 61.54%, 80% by (Shen et al., 2021), and 92.36%. by (Stathopoulos et al., 2023). Also, (Nam et al., 2023) used Multimodal deep neural network FacialCueNet that employed action units and micro expressions to reveal an individual's intentions and feelings. Additionally, short-term memory and convolutional neural networks construct the spatial-temporal attention module. The approach achieved an evaluation accuracy of 88.45%. Another study (Alaskar, 2023) uses the hybrid metaheuristics and deep learning for deception detection, developed a self-adaptive population-based firefly algorithm for deception detection, achieving a high accuracy of 99%. Also, (Khan et al., 2021) highlights the importance of eye movements by identifying them as a key feature for distinguishing between truthful and deceptive behavior, employing various classifiers. However, RF produces better results with 78% accuracy. When (Venkatesh, Ramachandra and Bours, 2019) using extracted micro expression features provided with the dataset yields the best accuracy of 88% in an individual-level model. While in (Şen et al., 2020), semi-automatic lying detection using OpenFace to automatically extract facial action units and the visual features that were given with the dataset was used to attain a best accuracy of 80.97% using nearest neighbor classifier and fully automatic lying detection that employed OpenFace to automatically extract facial action units got an accuracy of 61.58% use a random forest classifier. Moreover, (Nikbin and Qu, 2024) presents a well-structured approach that uses hybrid deep neural network (HDNN) to detect deception. The method reveals promising advancements in accuracy, and the methodology achieves a 91% accuracy rate in detecting fear-related micro-expressions. Another approach (Dinges et al., 2024) uses multiple sets of facial cues, each predicted by its own Convolutional Neural Network (CNN). The method involves CNN models analyzing gaze, head pose, and facial emotions, as well as Action Units (AUs). The research found that the effectiveness of deception detection varies by dataset context, with the best results on high-stake datasets. It revealed that integrating multiple modalities and classifiers led to an average accuracy of 67%. While (Kang et al., 2024) presents a novel method for deception detection that incorporates both global and local facial features. The study employs shallow CNNs to extract local features and utilizes a Video Transformer with spatiotemporal separation attention

to extract global ones. Both of these methods are effective at capturing complex facial dynamics and do well on existing datasets.

3. METHODOLOGY

We experimented with different approaches to find the ideal preprocessing and techniques for detecting deception as shown in Fig.1.

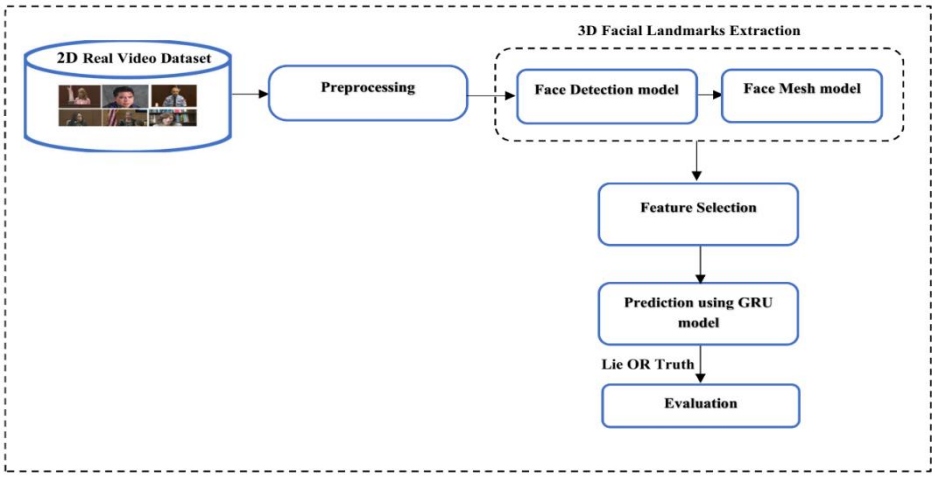


Fig. 1. Block diagram of the proposed system

3.1. The Dataset

We employ the real-life trial dataset (Pérez-Rosas et al., 2015). It is a multimodal deception detection dataset (video, audio, and text). The videos are real scenarios that have been downloaded from YouTube as shown in Fig. 2. These videos are factual police interrogation and courtroom videos, comprised of raw, 60 truthful, and 61 deceitful videos. Also, researchers have extracted certain features from these videos. These extracted features are facial, audio and text features which are indicators of deceptive behavior. In this study, one model-based raw video was used to extract the 3D 478 facial landmarks.

Fig. 2. Samples of the real video dataset

3.2. Preprocessing

Preprocessing aims to prepare a raw video dataset and extract features into a simple and efficient format, which are:

1-BGR Conversion

Involves converting an image from the RGB color space, commonly used in many image processing tasks, to the BGR color space. The RGB video frames are converted to BGR color

space by processing them with OpenCV. This conversion ensures compatibility with the model, preventing color misinterpretation that could lead to inaccurate facial landmark detection. This step is crucial in preparing the frame correctly for subsequent analysis, aligning with the expectations of the face mesh model for optimal performance.

2-Frame Deletion

The videos in the used dataset, from which we extract 3D facial landmarks, are real-life scenarios. Therefore, some faces in these videos are difficult to detect due to occlusion by subtitles or other objects and faces far from the camera may be too small or blurry to detect effectively. As a result, we exclude some frames that have zero values based on the following equations:

$$R_i = \sum_{j=1}^{n} U_{ij} \quad (1)$$

Where R_i is the total sum of values in the *i-th* row. U_{ij} represents the value in the *i-th* row and *j-th* column, in the matrix *B* that is $m \times n$, which stores 3D facial landmarks. After compute the summation of each row in the matrix *B*, now filters the row by delete the rows that have summation equal to zero according to this condition:

$$A_i = \begin{cases} 1 \text{ if } R_i \neq 0 \\ 0 \text{ if } R_i = 0 \end{cases} \quad (2)$$

Where *A* is a binary factor of length *m*, where each element A_i is 1 if $R_i \neq 0$ and 0 otherwise. Thus, we have now matrix $B'$where rows with a sum of zero are removed based on value of binary vector *A* using this equation:

$$B' = B \cdot A \quad (3)$$

In this context, the multiplication $B \cdot A$ s not standard matrix multiplication but rather a filtering operation where rows in B corresponding to $A_i = 0$ are removed.

3.3. 3D Facial Landmarks Extraction

The facial landmark detection technique used an end-to-end neural network Mediapipe franework. Mediapipe framework used a set of pre-trained models. It makes 3D facial landmark estimations forming a mesh of 478 points (see Fig. 3) from the input of 2D video frames. For extracting the 478 facial landmarks with three coordinates (x, y, z), it employed two deep learning models. It first used a face detector model which drew a rectangle around the detected faces to locate faces, and extracted main face landmarks including center of mouth, nose tip, left eye trigon, right eye trigon. The face mesh model uses cropped detected faces from the face detector model as input, without additional depth information. The face mesh model establishes

a metric 3D space and uses the positions of facial landmarks on the screen to estimate facial transformations in that space (Lugaresi et al., 2019). The face transformation data consists of conventional 3D primitives, including a face pose transformation matrix and a triangular face mesh. It uses regression to estimate the approximate 3D surface and generates a vector of the desired 478 facial landmarks, each with three coordinates (x, y, z) per frame for each video in the dataset. Furthermore, the model was trained using synthetic visualized data and 2D semantic contours from annotated real-world data. The resulting model provided us with reasonable predictions of 3D landmarks, not only on synthetic data but also on real-world data. This combination of data helps the model accurately infer the depth of facial features, resulting in robust 3D landmark detection from 2D inputs. This training enables the model to understand how 2D features map to a 3D face mesh (Li et al., 2024).

After extracting these 3D facial landmarks, we vectorized columns for each coordinate (x, y, z) per frame in every deceptive and truthful video, storing landmarks (Landmark0, Landmark477) in rows within an CSV file. Next, we store both the deceptive and truthful samples in a single CSV file. Furthermore, giving label zero value to the truthful samples and label one value to the deceptive samples.

3.4. Feature Selection

Feature selection is critical for improving a model's performance and interpretability by choosing relevant features. It focuses on the most informative aspects of features, and reduced dimensionality speeds up training and inference. The feature selection methods can be divided into four categories: filter methods, embedded methods, hybrid methods, and wrapper methods (Pudjihartono et al., 2022). In this paper, we utilized a multivariate mutual information (MMI) method. It is a filter technique that is used to analyze more than two features at once. The aim is to find patterns and dependence between several features simultaneously, allowing for a much deeper and more complex understanding of a given scenario than with two features. In deception detection, the system selects the most significant facial landmarks (features) that reflect deceptive cues, enabling it to distinguish between truthful and deceptive behaviour. These sets of selected facial landmarks also improve prediction model accuracy because they are most relevant to the classification process (deceptive vs. truthful). MMI is a measure of the dependence between a set of features. Consider three random features X, Y, and Z. MMI quantifies the amount of information that X, Y, and Z share. More specifically, MMI measures how much knowledge of one feature decreases uncertainty about the other. The calculation of

MMI involves integrating over the joint probability density function (PDF) and marginal density functions, as shown in the following equation (Batina et al., 2011):

$$MMI(X,Y,Z) = \sum_{x\in X} \int_{y^2} p[x,y,z].\log\left(\frac{p[x,y,z]}{p[x].p[y].p[z]}\right) dy$$

where $p[x,y,z]$ is the joint density distribution (PDF) and $p[x], p[y], p[z]$ are the marginal density distributions of the random variables (X,Y,Z), respectively. MMI measures the amount of information that one random feature contains about another. This measure is derived from the PDF, as it reflects the combined behavior of these features. In addition, the marginal distributions are the individual distributions of these random features, obtained by integrating the PDF over the other feature. So, if X, Y and Z are independent, their PDF would equal the product of their marginal distribution, leading to MMI (X, Y, Z) =0, Non-zero MMI indicates dependence. Since, to estimate the MMI between the features, the estimation of the PDF between them is needed and because density functions between high-dimensional features is a hard task in practice. Therefore, another alternative is simply not to estimate densities, while directly estimating the MMI by using the nearest neighbor estimator (NNE) which computed by the following equation (Nguyen, Xue and Andreae, 2016):

$$MMI(S) = \psi(\mathrm{k}) - \frac{m-1}{k} + (m-1) \times \psi(N) - \frac{1}{N} \times \sum_{i=1}^{N} \sum_{j=1}^{m} n_{ij} \qquad (5)$$

Where $\psi(\mathrm{k})$ is the digamma function used to adjust for the bias in the estimation process, m is the number of dimensions in the set of features (S), and features mean 3D facial landmarks. N is the number of samples in the dataset. K is the number of neighbors. n_{ij} is the number of neighbors for the *i*th instance. The idea is that if the neighbors of a specific observation in X space correspond to the same neighbors in the Yand Z space, there must be a strong relationship between X, Y and Z. These estimators locally estimate distributions based on distances between features. NNE methods use geometrically regular volume elements, the closet neighbor of facial landmark can help to estimate the structure around each landmark. So, the multivariate mutual information-based NNE involves calculating the MMI between each feature and the target label to identify the most informative features by using the nearest neighbor approach. Following the calculation of their MMI values, we sort the 3D facial landmarks in descending order based on their MMI values. The number of selected 3D facial landmarks was determined to be used with the GRU model. We repeatedly experimented with various numbers of selected 3D facial landmarks during the training of the GRU model until achieved the highest accuracy. This accuracy was achieved when used 35 facial landmarks, which represent facial muscles with the following indices: 361, 376, 132, 141, 360, 5, 459, 401, 288, 440, 94, 354, 25, 435, 209, 126,

64, 458, 241, 433, 49, 46, 456, 309, 275, 344, 102, 438, 26, 23, 125, 44, 19, 4, 1, as shown in Fig.4. They serve as indicators of deceitful behavior and have a major impact on the model, resulting in the GRU model's accuracy of 97%.

Fig. 4. Selected facial landmarks using multivariate mutual information method

3.5. Prediction using Gated Recurrent Unit (GRU) Model

GRU is an advancement of the standard recurrent neural network (RNN) that learns the dependencies between time steps in videos. It preserves longer sequences by employing two gates: the reset gate and the update gate (Ramasamy et al., 2020). These gates control what information is permitted to pass through to the output and can be trained to retrieve information over an extended period of time. This enables it to pass on related information along a chain of events, resulting in better predictions, which is critical for processing facial landmark sequences. The reset gate (r_t) determines how much memory is forgetting from previous hidden state $h_t - 1$ before proceeding to the next GRU cell. It takes the previous hidden state output and the current input (x_{-t}) as input and applies a sigmoid function to it. The reset gate equation as follows (Xing and Xiao, 2019):

$$r_t = \sigma(W_r[h_t - 1 + x_{-t}] + b_r) \quad (6)$$

Where, the weight, sigmoid function, and bias denoted by W_r,σ, b_r respectively. The update gate (z_t) consolidates the roles of the forget gate and input gate in an LSTM, determining the proportion of previous information that should be transmitted to the next GRU cell. The update gate equation as follows (Xing and Xiao, 2019):

$$z\_t = \sigma(W_z[h_t - 1 + x_t] + b_z \quad (7)$$

Where, the weight, sigmoid function, and bias denoted t by W_z, σ, b_z respectively. We used many layers in the model to acquire knowledge of the intermediate features that exist between the input data and the high-level classification, each one with a distinct role :

1- Batch Normalization layers are used to normalize the activations of the preceding layer. It also acts as a regularizer to help prevent overfitting.

2- Dropout layers are employed to mitigate overfitting by randomly excluding a portion of input units during the training phase.

3- The flattening layer converts the multidimensional output of convolutional layers into a one-dimensional vector. This step aggregates spatial features into a format suitable for dense layer processing and prediction.

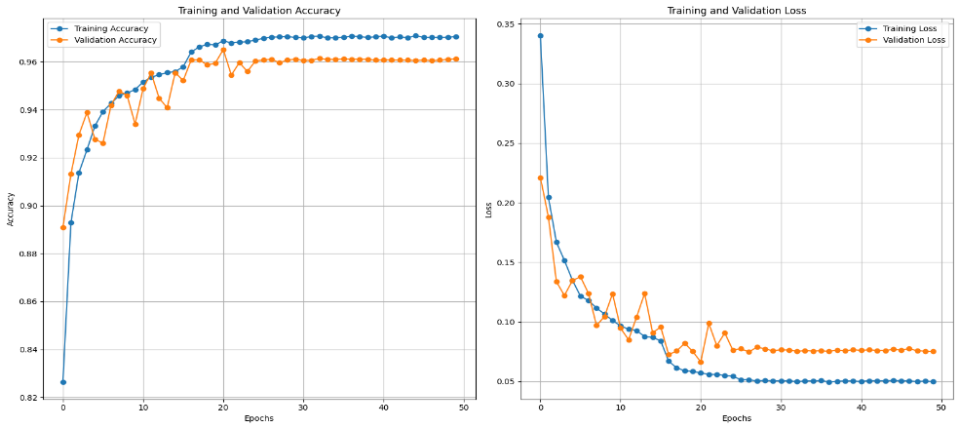
4- The dense layers, or fully connected layers, learn high-level, abstract representations and complex patterns by connecting each neuron to every neuron in the previous layer. They are crucial for making final predictions in the network, whether for classification, regression, or other outputs.

The GRU model has several advantages over other recurrent neural networks in certain situations. GRU is characterized by faster processing and lower memory consumption, which reduces training time.

4. RESULT AND DISCUSSION

The model was trained using robust 3D facial landmarks (features), extracting 478 of these features per frame from each video in the dataset. Their coverage extends to a broader spectrum of facial landmarks, allowing for a more comprehensive analysis of facial micro expressions that serve as indicators of deceitful behavior. In addition, we selected the important features using MMI that have the most influence on detecting liars from truthful tellers, which are 35 facial landmarks. So, the training process involves concatenating samples (3D facial landmarks) and their labels, 35 selected facial landmarks. The used dataset is split into training, validation, and test sets randomly with 70%, 15%, and 15%, respectively. The weights are fine-tuned to reduce the difference between actual and predicted outputs. This process is repeated until the model achieves a high level of accuracy in predicting the truth or lie. As a result, these informative 3D facial features enable the model to produce a satisfactory fit by striking a balance between overfitting and underfitting. Fig. 5 displays the learning loss graph for the proposed GRU model. The model is a good match because the training and validation losses are reduced to a point of stability. It has consistent learning features and illustrates the potential for better results from our strategy. As shown in Fig. 5, in the initial phase (epochs 0-10), the model rapidly improves, with training accuracy increasing quickly and validation accuracy rising with some fluctuations. Training loss drops sharply, while validation loss decreases but remains unstable, indicating early effective learning with some generalization challenges. While the middle phase (epochs 10–30) validation accuracy levels off with fewer fluctuations, training loss flattens, showing the model nearing convergence. Validation loss continues to

decrease, with minor fluctuations suggesting better generalization. Moreover, the final phase (epochs 30-50) Training accuracy stabilizes around 0.97, and training and validation losses remain steady, confirming the model has fully converged.



(A) **(B)**

Fig. 5. (A) Training and validation accuracy. (B) Training and validation loss.

Furthermore, the confusion matrix was used (Luque et al., 2019), to assess the effectiveness of a classification algorithm. It allows visualization of an algorithm's performance by comparing the actual and predicted classifications. The matrix consists of four quadrants true positive (TP) which represent the number of times the model correctly predicted the positive class, true negative (TN) which represent the number of times the model correctly predicted the negative class, false positive (FP) which represent the number of times the model incorrectly predicted the positive class when the actual class was negative, and false negative (FN) which represent the number of times the model incorrectly predicted the negative class when the actual class was positive. Also, it provides several key metrics for evaluating a classifier's performance: accuracy, precision, recall, and F1 score. In our case, using a confusion matrix to evaluate deception detection through facial expressions involves comparing the predicted labels with the actual labels. This helps in understanding how well the model distinguishes between deceptive and truthful tellers as shown in Tabel 1.

Tabel 1: Confusion matrix of deception detection

			Predictive Label	
			No	Yes
Actual Label	Truthful	No	True Negative 6050	False positive 241
	Deceptive	Yes	False Negative 54	True Positive 6302
			Truthful	Deceptive

The classification report based on confusion matrix that displays the numerical values of metrics accuracy, recall, F1-score, and precision for the truthful samples and for deceptive samples as shown in Tabel 2. The model has a high level of accuracy and performs well in both classes, with only minor discrepancies between precision and recall. The balanced F1-scores for both classes further support the idea that the model is effective in both detecting deception and correctly identifying truthful tellers. The slight differences in precision and recall across the classes suggest that the model is slightly better at correctly identifying deceptive instances (with higher recall for deceptive) but is more cautious with truthful predictions (higher precision for truthful). Overall, this indicates a well-performing model with a strong ability to differentiate between truthful and deceptive tellers. Consequently, this indicates that GRU model has demonstrated its capacity to capture long-term dependencies by preserving information over numerous time steps and from past steps. the GRU model adopted because of the nature of our problem, which requires tracking the movement of facial muscles in a sequential manner.

Table 2: Classification report of GRU model

Samples	**Precision= TP/(TP+FP)** (Kulkarni et al., 2020)	**Recall= TP/(TP+FN)** (Kulkarni et al., 2020)	**F1-Score = 2*(Recall*Precision)/ (Recall + Precision)** (Kulkarni et al., 2020)
Truthful Label	0.99	0.96	0.97
Deceptive Label	0.96	0.99	0.97
Accuracy= (TP+TN)/(TP+FP+FN+TN) (Kulkarni et al., 2020)		97%	
Top–k (the number of selected 3D facial landmarks)		35	

Moreover, to evaluate the effectiveness of the suggested model, the comparison analysis with several state-of-the-art methods has been implemented as shown in Tabel 3. Although the dataset includes multiple modalities, which are extracted facial, audio, and text features from real videos and provided with the dataset. Our model uses raw videos to extract 478 facial landmarks with coordinates (x,y,z) per frame from each video; we only used facial features. The decision-making process for deception is based on the patterns identified in facial landmarks during the model's training phase. These patterns are learned by the model through training, where it associates specific facial micro expressions and movements with deceptive behavior. The selection of relevant facial landmarks is guided by the Multivariate Mutual Information (MMI) method, which identifies the most informative features related to deception.

The detection of deception, especially through the extraction of 3D facial landmarks, is a particularly effective approach due to the detailed and comprehensive analysis it provides. Unlike 2D methods, 3D facial landmarks can capture subtle and intricate facial expressions from a variety of angles, making it possible to detect micro expressions that might indicate deception. These 3D landmarks offer a more robust solution in real-world conditions, as they can account for variations in lighting, head movement, and facial orientation, which are common challenges in video analysis. Additionally, 3D analysis enables continuous tracking of facial movements, allowing for a dynamic and context-sensitive understanding of facial behavior over time. With this depth and spatial awareness, video, enhanced with 3D landmark extraction, is a powerful tool for detecting deception because it provides a richer, more accurate set of data for analysis. Thus, we have presented comparative studies using only unimodal accuracy-based facial features. As shown in Table 3, the GRU model works the best out of all the comparison methods, with an accuracy rate of 97%. The comparisons (Stathopoulos et al., 2023) and (Şen et al., 2020) that used the well-known facial action coding system (FACS) rely on 2D techniques to find specific facial action units (AUs) that are linked to muscle movements. While FACS is effective in many contexts, its focus on predefined AUs and operation within a 2D plane make it limited. This can lead to missed subtle expressions, inaccuracies in capturing facial details, and difficulties in handling head rotations or varying lighting conditions. In addition, it is also based on human.

Table 3: Comparative methods of deception detection

References	Dataset	Feature	Method	Accuracy
(Stathopoulos et al., 2023)	Multimodal real-life trial	Facial action units	OpenFace, Temporal Convolutional Network, Attention module.	92.36%
(Şen et al., 2020)	Multimodal real-life trial	Facial action units, visual features provided with the dataset	OpenFace, nearest neighbor classifier, random forest classifier	Semi-automatic 80.79%, Fully automatic 61.58%.
(Venkatesh, Ramachandra and Bours, 2019)	Multimodal real-life trial	Micro expression provided with the dataset	AdaBoost classifier	88%
Our prediction model	Multimodal real-life trial	3D facial landmarks	GRU model	97%

judgments, which introduces another layer of potential error. FACS requires trained coders to manually identify and categorize facial action units (AUs), which can lead to subjective and be prone to biases and inconsistencies across different coders.

5. CONCLUSION

In this paper, we present a new method for facial micro expression detection using the 3D facial landmark-based deep neural network. The Mediapipe Face Mesh model was employed to obtain 478 facial landmarks with three coordinates (x, y, z) per frame for each video in the dataset. They capture more of the facial muscle movement and tracking of facial movements in a variety of angles and light conditions, which enables more analysis of the micro expressions that are used as indicators of deceitful behavior. Furthermore, to improve the accuracy of the classification, the feature selection-based multivariate mutual information method is used for selecting the 3D facial landmarks that have significant impact because they are most relevant to the movements that depict the facial micro expression. Besides, to take full advantage of GRU, we construct a model that can capture temporal dependencies and nuances to achieve 97% accuracy. The model's findings demonstrate how successful our method is in the real world and how accurate it is compared to other methods. In contrast to conventional techniques that involve human observation and judgment in analyzing and detecting the facial micro expression such as FACS, the characteristic cognitive load and the inability to process big data make human-based deception detection unreliable. Additionally, the use of equipment like fMRI for deception detection is limited due to its basic functionality and the need for human involvement to identify deception. Furthermore, the tension and interference of the electrodes attached to the face can complicate fMRI's application, making it unsuitable for widespread use.

6. REFERENCES

Alaa, R., Hussein, E. and Al-libawy, H. (2024) ‘OBJECT DETECTION ALGORITHMS IMPLEMENTATION ON EMBEDDED DEVICES: CHALLENGES AND SUGGESTED SOLUTIONS’, Kufa Journal of Engineering, 15(3), pp. 148–169.

Alaskar, H. (2023) ‘Hybrid Metaheuristics with Deep Learning Enabled Automated Deception Detection and Classification of Facial Expressions.’, Computers, Materials & Continua, 75(3).

Batina, L. et al. (2011) ‘Mutual information analysis: a comprehensive study’, Journal of Cryptology, 24(2), pp. 269–291.

Chen, F. et al. (2015) '2D facial landmark model design by combining key points and inserted points', Expert Systems with Applications, 42(21), pp. 7858–7868.

D'Ulizia, A. et al. (2023) 'Detecting Deceptive Behaviours through Facial Cues from Videos: A Systematic Review', Applied Sciences, 13(16), p. 9188.

Dinges, L. et al. (2024) 'Exploring facial cues: automated deception detection using artificial intelligence', Neural Computing and Applications, pp. 1–27.

Farah, M.J. et al. (2014) 'Functional MRI-based lie detection: scientific and societal challenges', Nature Reviews Neuroscience, 15(2), pp. 123–131.

Islam, S. et al. (2021) 'Non-invasive deception detection in videos using machine learning techniques', in 2021 5th International Conference on Electrical Engineering and Information Communication Technology (ICEEICT). IEEE, pp. 1–6.

Jabberi, M. et al. (2023) '68 landmarks are efficient for 3D face alignment: what about more? 3D face alignment method applied to face recognition', Multimedia Tools and Applications, 82(27), pp. 41435–41469.

Jakubowska, J. and Białecka-Pikul, M. (2020) 'A new model of the development of deception: Disentangling the role of false-belief understanding in deceptive ability', Social Development, 29(1), pp. 21–40.

Kang, J. et al. (2024) 'Deception Detection Algorithm Based on Global and Local Feature Fusion with Multi-head Attention', in 2024 3rd International Conference on Image Processing and Media Computing (ICIPMC). IEEE, pp. 162–168.

Khan, W. et al. (2021) 'Deception in the eyes of deceiver: A computer vision and machine learning based automated deception detection', Expert Systems with Applications, 169, p. 114341.

Kulkarni, A., Chong, D. and Batarseh, F.A. (2020) 'Foundations of data imbalance and solutions for a data democracy', in Data democracy. Elsevier, pp. 83–106.

Levine, T.R. (2014) 'Truth-default theory (TDT) a theory of human deception and deception detection', Journal of Language and Social Psychology, 33(4), pp. 378–392.

Li, S.Z. et al. (no date) Handbook of Face Recognition Third Edition. Available at: https://doi.org/https://doi.org/10.1007/978-3-031-43567-6.

Lugaresi, C. et al. (2019) ‘Mediapipe: A framework for perceiving and processing reality’, in Third workshop on computer vision for AR/VR at IEEE computer vision and pattern recognition (CVPR).

Luque, A. et al. (2019) ‘The impact of class imbalance in classification performance metrics based on the binary confusion matrix’, Pattern Recognition, 91, pp. 216–231.

Martinez, B. et al. (2017) ‘Automatic analysis of facial actions: A survey’, IEEE transactions on affective computing, 10(3), pp. 325–347.

Mohammed, H., Kareem, S. and Mohammed, A. (2022) ‘A COMPARATIVE EVALUATION OF DEEP LEARNING METHODS IN DIGITAL IMAGE CLASSIFICATION’, Kufa Journal of Engineering, 13(4), pp. 53–69.

Monaro, M. et al. (2022) ‘Detecting deception through facial expressions in a dataset of videotaped interviews: A comparison between human judges and machine learning models’, Computers in Human Behavior, 127, p. 107063.

Nam, B. et al. (2023) ‘FacialCueNet: unmasking deception-an interpretable model for criminal interrogation using facial expressions’, Applied Intelligence, 53(22), pp. 27413–27427.

Nguyen, H.B., Xue, B. and Andreae, P. (2016) ‘Mutual information for feature selection: estimation or counting?’, Evolutionary Intelligence, 9, pp. 95–110.

Nikbin, S. and Qu, Y. (2024) ‘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(3), pp. 14–20.

Pérez-Rosas, V. et al. (2015) ‘Verbal and nonverbal clues for real-life deception detection’, in Proceedings of the 2015 conference on empirical methods in natural language processing, pp. 2336–2346.

Prome, S.A. et al. (2024) ‘Deception detection using ML and DL techniques: A systematic review’, Natural Language Processing Journal, p. 100057.

Pudjihartono, N. et al. (2022) ‘A review of feature selection methods for machine learning-based disease risk prediction’, Frontiers in Bioinformatics, 2, p. 927312.

Rengasamy, D. et al. (2020) ‘Deep learning with dynamically weighted loss function for sensor-based prognostics and health management’, Sensors, 20(3), p. 723.

Şen, M.U. et al. (2020) ‘Multimodal deception detection using real-life trial data’, IEEE Transactions on Affective Computing, 13(1), pp. 306–319.

Shen, X. et al. (2021) ‘Catching a liar through facial expression of fear’, Frontiers in Psychology, 12, p. 675097.

Stathopoulos, A. et al. (2023) ‘Deception Detection in Videos Using Robust Facial Features with Attention Feedback’, in Handbook of Dynamic Data Driven Applications Systems: Volume 2. Springer, pp. 725–741.

Tsuchiya, K., Hatano, R. and Nishiyama, H. (2023) ‘Detecting deception using machine learning with facial expressions and pulse rate’, Artificial Life and Robotics, 28(3), pp. 509–519.

Venkatesh, S., Ramachandra, R. and Bours, P. (2019) ‘Robust algorithm for multimodal deception detection’, in 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR). IEEE, pp. 534–537.

Verma, M. et al. (2019) ‘LEARNet: Dynamic imaging network for micro expression recognition’, IEEE Transactions on Image Processing, 29, pp. 1618–1627.

Xing, Y. and Xiao, C. (2019) ‘A GRU model for aspect level sentiment analysis’, in Journal of Physics: Conference Series. IOP Publishing, p. 32042.

Yildirim, S., Chimeumanu, M.S. and Rana, Z.A. (2023) ‘The influence of micro-expressions on deception detection’, Multimedia Tools and Applications, 82(19), pp. 29115–29133.