Salah, S, Elbatanouny, H, Sobuh, A, Almajali, E, Khan, W, Alaskar, H, Binbusayyis, A, Hassan, T, Yousaf, J and Hussain, A

Explainable AI for Unraveling the Significance of Visual Cues in High Stakes Deception Detection

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 8 March 2025, accepted 28 March 2025, date of publication 8 April 2025, date of current version 21 April 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3558875

RESEARCH ARTICLE

Explainable AI for Unraveling the Significance of Visual Cues in High Stakes Deception Detection

SUHAIB SALAH[1], HAGAR ELBATANOUNY[2], ABRAR SOBUH[2], EQAB ALMAJALI[2,3], (Member, IEEE), WASIQ KHAN[4], (Senior Member, IEEE), HAYA ALASKAR[5], ADEL BINBUSAYYIS[5], (Member, IEEE), TAIMUR HASSAN[6], (Senior Member, IEEE), JAWAD YOUSAF[6], (Senior Member, IEEE), AND ABIR HUSSAIN[2,4], (Senior Member, IEEE)

[1]Department of Computer and Information Engineering, Khalifa University, Abu Dhabi, United Arab Emirates
[2]Electrical Engineering Department, University of Sharjah, Sharjah, United Arab Emirates
[3]School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, ON K1N 6N5, Canada
[4]School of Computer Science and Mathematics, Faculty of Engineering, Liverpool John Moores University, L3 3AF Liverpool, U.K.
[5]Department of Computer Sciences, Prince Sattam Bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia
[6]Department of Electrical, Computer, and Biomedical Engineering, Abu Dhabi University, Abu Dhabi, United Arab Emirates

Corresponding authors: Suhaib Salah (eng.suhaib.salah@gmail.com), Abir Hussain (abir.hussain@sharjah.ac.ae), and Eqab Almajali (ealmajali@sharjah.ac.ae)

This work was supported in part by Prince Sattam Bin Abdulaziz University under Project PSAU/2024/R/1445, and in part by the University of Sharjah under Project 22020403199.

ABSTRACT Deception, a widespread aspect of human behavior, has significant implications in fields like law enforcement, security, judicial proceedings, and social areas. Detecting deception accurately, especially in high-stakes environments, is critical for ensuring justice and security. Recently, machine learning has significantly enhanced deception detection capabilities by analyzing various behavioral and visual cues. However, machine learning models often operate as opaque ''black boxes,'' offering high predictive accuracy without explaining the reasoning behind the decisions. This lack of transparency necessitates the integration of Explainable Artificial Intelligence to make the models' decisions understandable and trustworthy. This study proposes the implementation of existing model-agnostic Explainable Artificial Intelligence techniques—Permutation Importance, Partial Dependence Plots, and SHapley Additive exPlanations—to showcase the contributions of visual features in deception detection. Using Real-Life Trial dataset, recognized as the most valuable high-stake dataset, we demonstrate that Multi-layer Perceptron achieved the highest accuracy of 88% and a recall of 92.86%. Along with the aforementioned existing techniques, Real-Life Trial dataset inspired us to develop a novel technique: 'set-of-features permutation importance'. Additionally, this study is novel in the sense of that it extensively applies XAI techniques in the field of deception detection on Real-Life Trial dataset. Experimental results shows that the visual cues related to eyebrow movements are most indicative of deceptive behavior. Along with the new findings, our work underscores the importance of making machine learning models more transparent and explainable, thereby enhancing their utility for human-in-loop AI and ethical acceptability.

INDEX TERMS Deception detection, human-in-loop AI, trustworthy AI, permutation importance, partial dependence plots, shap, explainable machine learning, black-box models, model-agnostic techniques, multi-layer perceptron.

I. INTRODUCTION

Deception detection refers to the act of using behavioral and psychological cues to determine that a person is deliberately trying to deceive the receivers of the information [1]. The practice of deception detection holds significance in various fields, including border security [2], [3], law enforcement [4], criminal investigations [5], corporate matters [6], clinical psychology [7]. A person's deceptive state can be determined through a machine (or a system) equipped with cognitive

The associate editor coordinating the review of this manuscript and approving it for publication was Orazio Gambino.

capabilities for detecting visual, vocal, textual, and psychological indicators associated with deception. People, when facing situations that are against their own benefit, lie once to twice a day in average [8]. Most lies are small and harmless. However, this research is directed to detecting lies with major threats to the person and to the society. Considering this, it is necessary to implement the sophisticated techniques to tackle deception detection accurately and reliably. Research studies reveal that people can easily deceive others when compared with the technological lie detectors. For instance, an average person can detect only 54% of lies [9], which justifies the need of more efficient and effective deceit detecting systems.

Understanding the psychology of deception is crucial for developing an effective AI-driven methods. The cognitive load associated with lying triggers specific physiological and psychological responses. The associated non-verbal cues, including body language, facial expressions, and physiological responses, play a pivotal role in deception detection [10]. Research shows that under the strain of deception, certain behaviors show significant association, such as involuntary facial expressions known as microexpressions [11], changes in posture, fidgeting, rigidity [12], and eye behaviors like blinking rates and pupil dilation [13]. On the other hand, verbal cues also play important role in identifying deception, involving both speech content and style. Hesitations, speech errors, and changes in speech rate can indicate cognitive overload [5], and deceptive statements often lack detail or contain inconsistencies due to the cognitive demands of fabricating a lie [14]. Linguistic cues such as reduction in first-person pronouns and increased negations may also indicate deception [15].

The evolution of deception detection methodologies spans from traditional, contact-based, human-centric approaches to sophisticated, contactless, AI-driven systems. Traditional approaches rely on psychological assessments and physiological measurements. A key method is the polygraph, a device that detects lies by connecting sensors to the person's body to measure signals like respiration, blood pressure, heart rate, and sweat [16]. The examiner, an expert, analyzes these measurements to classify the person's answers as truthful or deceptive. However, polygraphs have a number of pitfalls. For instance, the subject must cooperate and allow sensors to be attached. The test can be biased against honest people, while compulsive liars can train themselves to give false statements as 'truth'. Additionally, polygraph tests require professional examiners where a ten-minute interrogation needs hours of analysis, requiring substantial efforts and resources. Likewise, the physiological measurements can also be affected by the nature of the questions [16].

On the other hand, non-invasive methods offer a less intrusive alternative with broader applications. They encompass various techniques eliminating the direct physical interaction. Verbal analysis is one of such methods examining speech or written content, identifying linguistic patterns, hesitations, and inconsistencies as signs of deceit [17]. Likewise, visual analysis techniques detect micro-expressions, small facial gestures revealing underlying emotions [18], while thermal imaging detects temperature variations on the face during deceit [18]. Behavioral cues, such as shifts in posture, offer additional insights [17], and acoustic analysis, evaluating pitch, tone, and speech rhythm, provide insights into the individual's emotional state and truthfulness [17].

Despite the benefits of non-invasive methods, the associated cues are highly complex to analyze using traditional techniques, requiring machine learning (ML). The rapid development of ML technologies has significantly enhanced deception detection systems, especially in analyzing the vast and complex array of data sources associated with deception, ranging from linguistic nuances to facial gestures [19].

A variety of ML algorithms are employed in this domain, each suited to different aspects of deception detection. Supervised learning algorithms, such as Support Vector Machines (SVM) and Neural Networks, are adept at classifying data into deceptive or truthful categories based on labeled training datasets [20]. Unsupervised learning explores data without predefined labels, identifying anomalous patterns that could signify deception [21]. Semi-supervised learning leverages the strengths of both, utilizing limited labeled data to guide the analysis of larger unlabeled datasets [22], [23], allowing for a nuanced understanding and detection of deceptive behaviours.

While ML offers great capabilities in terms of accuracy and performance, it comes with a challenge: the ambiguity of its decision-making process. ML models, often described as ''black boxes,'' produce results without clear explanations on how these results are derived. This lack of transparency can be problematic in the realm of deception detection, where understanding the rationale behind a decision is crucial for several reasons. For instance, refining the model with respect to errors. Similarly, human understandability for expert-in-loop decision. To resolve these challenges, XAI might be utilized in this field for better transparency and interpretability of machine-based deception detection methods. Incorporating XAI in deception detection ensures that the expert (such as investigative officers) can comprehend and trust the determinations made by ML model. This transparency enhances accountability, enabling potential biases or errors to be identified and rectified [24].

This study contributes to the state-of-the-art by extensively applying XAI techniques in the field of deception detection on real-life trial dataset. This work underscores the study's contribution to expanding the methodological toolkit available for analyzing and understanding deception cues within real-world contexts. The study [24] enabled us to integrate the following XAI techniques: integrating permutation importance [25], Partial Dependence Plots (PDP) [26], and SHAP [27], the study establishes a nuanced correlation between specific visual features and the classification outcomes of deception or truthful behavior. This approach of offering both global and local interpretations

of the dataset enriches the analytical depth, allowing for a detailed exploration of the significance of particular features and cues in the detection process.

Furthermore, the research introduces a novel permutation importance technique, specifically designed as a set of features permutation importance. This innovative method is inspired by the nature of the real-life trial dataset, which suggests that features should be conceptualized and analyzed as groups rather than as isolated units. This approach acknowledges the complex nature of deception cues, providing a more accurate and holistic understanding of their role and significance. In summary:

- This study contributes to the field by extensively applying XAI models to the high-stake deception detection field, which marks a novel advancement in the field.
- We provide a novel set-of-features permutation importance XAI technique, inspired by the nature of the Real Life Trial dataset. This novel technique realistically tackles the Real Life Trial dataset's visual cues.
- The holistic methodology of this study emphasizes the essence of ML interpretations in security and justice sectors. This is approached by the implementation of a wide range of local and global model agnostic XAI techniques in the realm of deception detection.

The remainder of this study is organized as follows: Section II examines related works, encompassing AI models, features, datasets, and sensors. Section III describes the 'Methodology' milestones. Section IV presents and analyzes the results, offering insightful conclusions. Finally, Section V summarizes the study's key findings and proposes directions for future research.

II. LITERATURE REVIEW

Deception detection areas demonstrate significant diversity in AI models, including Neural Networks [5], [6], SVMs [19], [28], Convolutional Neural Networks (CNNs) [29], [30], and Decision Trees [31]. This variety underscores the multifaceted nature and complexity of deception detection tasks and approaches. The reported accuracies of these models vary notably across studies, with Neural Networks often showing high performance [5], [6]. However, high performance raises questions about potential overfitting and the models' ability to generalize across diverse datasets. The use of transfer learning, as demonstrated in research works such as [32], reflects an effort to harness pre-existing AI capabilities for specific deception detection tasks, thereby enhancing overall model efficacy. Additionally, the incorporation of model-agnostic XAI approaches, particularly in research studies [33], [34], is a growing trend, aiming to enhance the transparency and interpretability of AI models in a domain where understanding the rationale behind decisions is crucial. Recently, a variety of research studies proposed deep learning methods as an effective alternative to conventional ML for the deception detection task [5], [6], [29], [30], [32], [35], [36], [37], [38], [39], [40], [41], [42], [43], [44], [45], [46], [47], [48], [49]. However, deep learning models require large amounts of data as well as substantial resources, which is why some other studies still use traditional ML techniques K-Nearest-Neighbours (KNN), SVM, decision tree, and random forest [3], [19], [28], [31], [34], [50], [51], [52], [53], [54], [55], [56], [57].

It's crucial to emphasize that high accuracy does not necessarily imply high quality. For instance, study [5] reported 100% accuracy. however, subjectivity should be considered. For instance, the training and test datasets were identical in terms of the one male suspect, questions, and answers, with the only difference between the training and the test sets being the time of day when the recordings were made. This example illustrates that accuracy alone is not a conclusive indicator of the model's success.

The features utilized in deception detection, as detailed in the reviewed studies, reveal a wide array of methods to capture deception. Audio features, such as stress patterns in speech, are central in some approaches [5], [37], whereas textual analysis plays a critical role in others, particularly in examining linguistic patterns [33], [34]. Similarly, visual features encompass facial expressions and body language, have been reported as effective for enhancing accuracy in deception detection. These features are predominant in several studies [29], [31], [39], reflecting the importance of non-verbal cues in deception detection. A recent work [3] reported that eye micro-movements serves as more effective indicators of deception than facial micro-expressions. Some research employs thermal imaging [38], showcasing the use of innovative sensors to detect physiological changes associated with deceit. Notably, the integration of multiple modalities, such as combining audio-visual data [6], offers a comprehensive approach to enhance detection accuracy. These diverse feature sets underscore the complexity of deception detection and highlight the need for multifaceted approaches to effectively interpret and analyze deceptive behavior.

In addition to ML models, the datasets used in the literature also play a crucial role, encompassing a wide range of types and sources. Audio datasets, which capture vocal nuances and stress patterns, are integral to studies like [5] and [37]. Similarly, textual datasets are central in text-based deception detection, as evidenced in [33] and [34]. Moreover, visual datasets that focus on visual cues such as facial expressions and body language are prominently featured in the field. This diversity, highlighted in the reviewed studies, extends to the categorization of datasets into real-life and mock datasets. Real-life datasets, for instance, provide authentic settings for model testing and are exemplified by studies such as [6], which utilize the Real-Life Trial dataset [58] for a realistic context in deception detection. Conversely, mock datasets are created in controlled environments, like games or simulated interviews, to study deceptive behavior in a structured manner. Examples of such datasets include [40] and [59]. Additionally, the use of custom datasets,

particularly facial datasets combined with transfer learning from large-scale databases like ImageNet [32], illustrates the trend towards creating more specific and effective datasets for targeted deception detection tasks. The most frequently deception detection dataset used in the literature is Real-Life Trial dataset [58] which is visual and textual. Real-Life Trial dataset is the most utilized deception detection dataset in the literature [6], [29], [36], [39], [41], [52], [60]. Other popular datasets include bag of lies dataset [61] and Ott Deceptive Opinion Spam Dataset [62].

Various types of sensors have been used in the literature to collect deception dataset, however common practice is to use secondary data instead of primary data collection. A research presented in [37] utilized microphones to capture nuanced audio data, such as stress patterns in speech, underscoring the importance of auditory cues. Similarly, textual data analysis plays a crucial role in deception detection without the need for physical sensors, as demonstrated in studies like [33], [34], and [63], where linguistic patterns and textual cues are analyzed. Visual sensors, including RGB cameras, are utilized in [32] for detailed facial analysis and micro-expression detection. Thermal imaging is innovatively used for physiological change detection associated with deceit, with [38] employing the Seek Compact Thermal Imager for Android and [28] utilizing a FLIR C2 compact thermal camera to capture detailed thermal data. The integration of multiple sensor types, as observed in [28], illustrates the diverse methodologies in this field.

Table 1 provides a concise comparison of related works in the literature.

III. DATASET DESCRIPTION

Real-life Trial Dataset comprises video clips from court trials, offering a balanced mix of deceptive and truthful statements in realistic environments converted into structured raw format. The pre- processing steps include audio transcription and behavioral annotation, to ensure the dataset's readiness for analysis. This preparation enables leveraging the dataset for developing accurate deception detection models, setting a solid foundation for this research.

A. DATASET SELECTION: REAL-LIFE TRIAL DATASET

This study centers on the Real-life Trial Dataset, identified as the most extensively utilized dataset for deception detection research within the existing literature [58]. Its selection was driven by the critical need to analyze deception in high-stakes environments, where the consequences of deceit can significantly impact trial outcomes and judicial decisions. As shown in Table 2, the dataset comprises 121 video clips, split almost evenly between deceptive and truthful statements, collected from publicly available court trial recordings. These clips include a rich variety of deceptive and truthful examples, featuring defendants and witnesses, making it uniquely suited for studying the multifaceted nature of human deception.

The Real-life Trial Dataset is distinguished by its real-world applicability and its multimodal nature, incorporating both verbal and non-verbal cues to deception. This aspect is crucial for developing a comprehensive understanding of deception, as previous research suggests that a multimodal approach can significantly enhance detection accuracy. The dataset was painstakingly compiled from a variety of sources, ensuring that each clip met specific quality standards in terms of visual clarity and audio comprehensibility. This careful selection process ensured that the dataset would be both representative of real-life scenarios and suitable for detailed analysis using computational methods.

In preparing the dataset for analysis, several steps were undertaken to optimize its utility for deception detection research. These included the transcription of audio content to text, enabling the analysis of verbal cues, and the detailed annotation of non-verbal behaviors, such as facial expressions and hand movements. The researchers in [58] achieved this annotation by leveraging the Multimodal Multilingual Information Management (MUMIN) coding scheme to categorize gestures and facial expressions. Such steps are critical for facilitating the extraction of meaningful features from the dataset, which can then be used to train and evaluate deception detection models.

Furthermore, the Real-life Trial Dataset's application in this study is underpinned by a thorough examination of its characteristics, including the balance between deceptive and truthful clips, the demographic diversity of the individuals featured in age and sex, and the contextual variety of the deception instances it contains. This analysis reaffirms the dataset's relevance to the study's objectives and highlights its potential to contribute valuable insights to the field of deception detection. Table 2 summarizes real-life trial dataset.

B. DATA PREPROCESSING

The preprocessing stage is crucial for ensuring the dataset's readiness for effective model training and evaluation. This stage involved a series of meticulously executed steps aimed at refining the dataset for the subsequent analysis. Initially, an exhaustive check for missing values was conducted across the dataset, revealing no instances of missing data. This absence of missing values assures the dataset's completeness and reliability, facilitating a straightforward analysis process.

Another critical aspect examined was the balance between deceptive and truthful classes within the dataset. A balanced distribution is vital for preventing model bias towards the more prevalent class. Upon evaluation, the dataset was found to be balanced, with the classes of deceptive and truthful instances being nearly equal. This balance enhances the generalizability and fairness of the models developed from this dataset.

Feature validation was carried out with meticulous attention to ensure the integrity and correctness of the features within each combination. A comprehensive review of all rows and groups was undertaken, revealing a singular discrepancy

TABLE 1. Comparison of deception detection studies.

Ref	AI Model	Accuracy	Dataset(s)
[37] (2023)	Neural Network	0.98	Audio recordings from interviews with a randomly selected group
[3] (2021)	Random Forest	0.73	126,291 deceptive and 128,735 truthful instances
[53] (2022)	SVM	0.78	60 videos in low-stakes situations
[43] (2023)	Neural Network	0.79	Three datasets from: Personal opinions, autobiographical memories, etc.
[44] (2023)	Neural Network	0.99	MU3D
[54] (2023)	ML Methods	0.71	Card game interaction with iCub robot
[45] (2023)	Neural network	0.99	Real-Life Trial dataset, Box of Lies dataset
[46] (2023)	Neural Network	0.61	"Cheat-Game" dataset, 10788 samples
[55] (2023)	KNN	1.00	RLT
[30] (2023)	CNN	0.74	FER-2013 dataset
[47] (2022)	Neural network	0.88	RLT and Live-Action Program datasets
[56] (2020)	Different ML	0.66	LieCatcher game, large corpus of interviews
[48] (2024)	Neural Network	0.81	Interview-style corpus, CSC corpus
[49] (2021)	Neural Network	0.78	Bag-of-Lies dataset
[57] (2020)	Machine Learning	0.90	THEPHY dataset, mock crime
[33] (2022)	Several ML methods	0.73	Crowdsourced dataset of 1640 statements on planned activities [64]
[34] (2019)	SVM	0.87	Ott Deceptive Opinion Spam dataset
[63] (2022)	Several ML methods	0.69	Dataset of 1487 statements (757 typed, 730 transcribed) from a project on deception detection
[39] (2021)	Neural network	0.97	Real-Life Trial dataset, Bag-of-Lies dataset, and a dataset of long videos from The Resistance game
[32] (2022)	Neural Network	0.98	Custom facial dataset, with ImageNet dataset
[6] (2021)	Neural Network	0.97	BoL, RL trail, MU3D databases
[5] (2021)	Neural Network	1.00	One suspect police interrogation
[38] (2021)	Neural network	0.61	Interviews with ten participants
[29] (2021)	CNN	0.68	Real-Life Trial, Low-Stakes Deceit datasets
[31] (2020)	Decision tree	0.70	Youtube political statements
[19] (2021)	SVM	0.62	Bag-of-Lies dataset
[28] (2022)	SVM	0.64	Deception Detection and Physiological Monitoring (DDPM) dataset
[2] (2020)	Several ML models	Various	Image Vector dataset (86584 vectors)
[35] (2020)	Neural Network	0.94	Ott Deceptive Opinion Spam dataset
[36] (2022)	Neural Network	0.83	Real-life trial data (public court trial videos)
[40] (2020)	Neural Network	0.60	Controlled games in English and Hebrew
[41] (2019)	Neural Network	0.97	Real-life trial
[60] (2019)	Combined ML methods	0.97	Real life trial
[50] (2019)	Random Forest	0.90	smartphone surveys (47 participants)
[51] (2019)	SVM	0.82	KWOLF dataset with 388 speech samples
[52] (2019)	SVM	0.77	Real-life Trial
[42] (2019)	Neural Network	0.75	Daily Deceptive Dialogues Corpus of Mandarin (DDDM), 7504 utterances, 96 speakers
[59] (2019)	Combined ML methods	0.71	Resistance game videos, 285 players from 44 games

TABLE 2. Summary of real-life trial dataset for deception detection.

Attribute	Detail
Number of Instances	121 (61 deceptive, 60 truthful)
Number of Subjects	56 (21 female, 35 male)
Age Range	Approximately 16-60 years
Number of Features	39

in the 6th instance regarding hand movements, where all values were zeros, which is invalid. We noted that the number of features is unequal to other groups. This issue was attributed to the inherent challenge in visually capturing hand movements across all videos, resulting in this group totaling 68 instead of 121. All other feature groups, including mouth, eyes, gaze, eyebrows, head, gestures, and hand, correctly summed to 121, affirming the dataset's overall consistency and reliability.

For model training and evaluation, categorical class labels were converted to numerical values, with 'deceptive' instances labeled as 1 and 'truthful' instances labeled as 0. This conversion facilitates the use of computational models that require numerical values. Furthermore, the 39 features identified were meticulously grouped into seven categories: mouth, eyes, gaze, eyebrows, head, gestures, and hand. This grouping strategy aims to capture the multifaceted nature of deceptive behavior through various non-verbal cues [58].

To guarantee reliability and robustness, the dataset was split into training and testing sets following a leave-out cross-validation strategy, ensuring that a substantial portion of the data is used for model training while reserving a representative subset for evaluation. This requires a 'subjective split', meaning that 20% of the samples contain subjects that are unseen in the 80% training data. This technique involves partitioning the training dataset into five subsets, with the model being trained on four subsets and validated on the remaining one. This process is recursively implemented five times.

IV. EXPLAINABLE MACHINE LEARNING METHODOLOGY

The proposed XAI-based deception detection method is a composite of dataset preparation, the selection and hyperparameter tuning of various ML models, and XAI techniques to enhance the interpretability and identification of novel

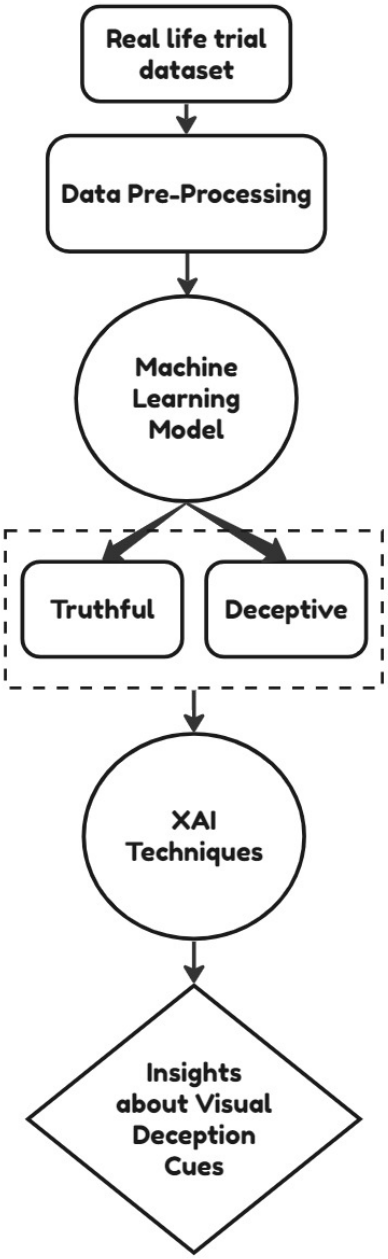

FIGURE 1. Methodology block diagram.

insights. Our comprehensive methodological framework aims to advance the understanding of deception, contributing to the broader application of AI in high-stakes settings, and includes the following main components. Figure 1 presents the details of the implemented XAI methodology.

A. MODELS' TRAINING AND EVALUATION

In this study, a diverse array of ML models was selected to explore their effectiveness in deception detection. The models include Multi-layer Perceptron (MLP), SVM, Decision Trees, Random Forests, Logistic Regression, KNN, Naive Bayes, Light Gradient Boosting Machine (LGBM), XGBoost, and CatBoost. This selection spans a wide range of approaches, from simple linear models to complex ensemble methods, offering a comprehensive examination of various strategies in tackling the problem of deception detection.

TABLE 3. Hyperparameters tuned for each classifier.

Classifier	Hyperparameters
MLP	hidden_layer_sizes, activation, solver, alpha, learning_rate
XGBoost	n_estimators, max_depth, learning_rate, subsample, colsample_bytree, min_child_weight, gamma
CatBoost	iterations, learning_rate, depth, l2_leaf_reg, border_count, loss_function, eval_metric, random_seed, silent
Random Forest	n_estimators, max_features, max_depth, min_samples_split, min_samples_leaf, bootstrap
Decision Tree	max_depth, min_samples_split, min_samples_leaf, criterion
Logistic Regression	C, solver, penalty
KNN	n_neighbors, weights, metric, leaf_size
Naive Bayes	var_smoothing
LGBM	n_estimators, max_depth, learning_rate, num_leaves, boosting_type, objective, colsample_bytree, subsample

To optimize the performance of each model, a systematic hyperparameter tuning was conducted using grid search and randomized search. This process involves training each model multiple times with different combinations of hyperparameters, allowing for the identification of the optimal set that yields the best performance. The grid search and the randomized search were meticulously designed to cover a broad spectrum of hyperparameters for each model, ensuring a thorough exploration of the parameter space. Table 3 summarizes the hyperparameters tuned for each model.

Afterwards, model evaluation is a critical step in the process of validating the efficacy of ML models. In this study, each model's performance was rigorously assessed using a range of metrics that provide insights into various aspects of prediction quality. These metrics included accuracy, precision, recall, and the area under the curve (AUC). Accuracy measures the proportion of true results among the total number of cases examined, precision reflects the proportion of true positive results in the dataset, recall indicates the fraction of relevant instances that have been retrieved over the total amount of relevant instances. The AUC curve is a performance measurement for classification problems at various threshold settings, representing the degree of separability achieved by the model.

The models demonstrating superior performance across the metrics will be subjected to XAI techniques to shed light on the contributing factors leading to their decisions. By applying XAI, this study aims to unravel the black-box nature of complex models, particularly those like neural networks and ensemble methods, which often offer limited interpretability despite their high accuracy. The insights gained from XAI will be pivotal in understanding model

behavior and will be instrumental in the advancement of the field of deception detection.

B. EXPLAINABLE AI ANALYSIS

This study employs several XAI methods to interpret the models selected for their exemplary performance in the deception detection task.

1) PERMUTATION IMPORTANCE

Permutation importance is utilized as the primary feature ranking technique. Unlike other methods, permutation importance does not require retraining the model multiple times, thus providing a computationally efficient means of evaluating feature significance. Two phases of permutation importance analysis are conducted. The first phase assessed individual features to determine their independent impact on the model's predictions. The second, a novel phase, evaluated groups of features, providing insight into how combinations of related features contribute to the detection of deception.

- **Single-feature Permutation Importance:**
  1) Initialize the model with a test set and evaluate its accuracy.
  2) Shuffle a single feature within the test set, ensuring the use of a specific random seed for reproducibility.
  3) Re-evaluate the model's accuracy on the shuffled test set.
  4) Compute the permutation importance as the difference between the original and new accuracies, denoted as:

$$PI = OA - NA \tag{1}$$

  where **PI** is the permutation Importance, **OA** is the accuracy of the model applied on the original dataset, and **NA** is the accuracy of the model applied on the shuffled dataset.
- **Novel Proposed Set-of-features Permutation Importance:**
  1) Start with the same model and its corresponding accuracy on the test set.
  2) Group the 39 features into 7 categories and shuffle a whole group together, maintaining the reproducibility of the shuffle. Adhere to the dataset's structure where for every group there is one '1', and the remaining are '0's, maintaining the integrity of the real-life trial dataset structure during permutation.
  3) Assess the model's accuracy post-shuffling of the feature group.
  4) Determine the permutation importance in the same manner described in equation 1.

2) PARTIAL DEPENDENCE PLOTS (PDP)

The analysis is further extended by the use of PDPs, which graphically depict the relationship between selected features and the predicted outcome. Two approaches are used in the PDP analysis:

- **Single Feature Impact:** The first approach explores how variations in a single feature's values affect the model's predictions, holding all other features constant. This method highlights the influence of individual features on the decision boundary between deceptive and truthful classes.
- **Feature Interaction Impact:** The second approach investigates the interactions between pairs of features and how these interactions alter the predicted outcome. This two-way PDP analysis is instrumental in understanding the combined effect of feature interactions on the model's predictions.

3) SHAP VALUES

Finally, SHapley Additive exPlanations (SHAP) values are employed to measure the contribution of each feature to individual predictions. SHAP values offer both local and global interpretability:

- **Local Interpretability:** On a local scale, SHAP values provide insights into individual predictions by quantifying the impact of each feature. This reveals the directionality of feature influence, indicating whether a feature pushes the model's output towards deception or truth.
- **Global Interpretability:** Globally, SHAP values aggregate the effects of features over a dataset, highlighting the overall importance and impact of features across numerous instances. This holistic view aids in identifying consistent patterns and trends in feature contributions.

By leveraging these XAI techniques, this study aims to improve the predictive performance of deception detection models and to enhance the transparency and trustworthiness of AI in high-stakes decision-making scenarios. The combined use of permutation importance, PDPs, and SHAP values provides a comprehensive suite of tools for interpreting complex model behaviors and substantiating the factors driving their predictions.

V. RESULTS AND DISCUSSION

A comprehensive analysis of the results is performed for multiple ML algorithms. The performances are evaluated and discussed based on a range of metrics including accuracy, precision, recall, and AUC. This discussion extends to a comparative analysis of the models' results against existing benchmarks and theoretical expectations derived from the literature. In addition to quantitative performance metrics, this section delves into the qualitative insights yielded by the application of XAI techniques. Permutation importance, PDP, and SHAP values have been employed to interpret the models and understand the underlying factors that contribute to their predictive accuracy. The results from these interpretability

techniques are discussed to provide a deeper understanding of the feature contributions and model behaviors.

The discussions aim to bridge the gap between raw predictive performance and the interpretive understanding necessary for practical application in real-world scenarios.

A. MACHINE LEARNING CLASSIFIERS PERFORMANCE

Table 4 presents the performance of each classifier. MLP emerged as the top-performing model in all metrics with the highest accuracy of 88%, indicating a superior overall classification rate. Its precision of 86.67% and recall of 92.86% demonstrate its effectiveness in identifying deceptive instances with a low rate of false positives and high true positive rate. The AUC of 84.42% further confirms the MLP's robustness in distinguishing between classes.

SVM and XGBoost both achieved an accuracy of 80.00%. However, XGBoost exhibited a higher AUC score of 88.96%, suggesting its greater capability in class discrimination compared to SVM's 81.17%. This may be attributed to XGBoost's ensemble approach, which typically provides better generalization. However, SVM and XGBoost have equal precision and recall. Recall and precision do follow similar trends to the accuracy's trend across the array of models used in this study.

CatBoost, another ensemble method, showed slightly lower performance with an accuracy of 76.00% and the lowest AUC of 68.83% among the top four models. Random Forest and Decision Tree classifiers displayed identical accuracies of 72.00%. The Random Forest model had a higher AUC score of 76.36%, compared to the Decision Tree's 66.23%, which might be due to the ensemble nature of Random Forest providing a more nuanced decision boundary.

Logistic Regression and KNN both reported accuracies of 68.00%, with Logistic Regression achieving a slightly higher AUC. This indicates that the linear decision boundaries of Logistic Regression are relatively effective for this task, despite the complex feature space. Naive Bayes showed a lower accuracy of 60.00% and an AUC of 62.99%, reflecting its challenges with the dataset's features, which may not meet the naive conditional independence assumption of this classifier.

Lastly, LGBM recorded the lowest accuracy of 56.00% and an AUC of 58.44%, suggesting that the model's configuration was not optimal or that this approach is less suited to the characteristics of the dataset. Furthermore, accuracy and AUC being close to 50% indicate chance level. Having such performance suggests that LGBM is not functional in this task.

We conducted another experiment in which we designed the custom deep learning model, which takes both video and text input from the given 'Real-Life' dataset to predict whether the given case is deceptive or truthful. It is to be noted here that the 'Real-Life' dataset does not explicitly contain the text and video input of each case. Instead, it comprises different text and video samples for various cases that do not have any relationship with each other. This lack of video and textual for each case within the dataset creates a problem for the multimodal deep learning model as it expects both video and text inputs to be passed for each case to recognize the deceptive and truthful categories accurately.

TABLE 4. Performance of various classifiers.

Classifier	Accuracy	Precision	Recall	AUC
MLP	88.00%	86.67%	92.86%	84.42%
SVM	80.00%	80.00%	85.71%	81.17%
XGBoost	80.00%	80.00%	85.71%	88.96%
CatBoost	76.00%	75.00%	85.71%	68.83%
Random Forest	72.00%	76.92%	71.43%	76.36%
Decision Tree	72.00%	73.33%	78.57%	66.23%
Logistic Regression	68.00%	75.00%	64.29%	70.13%
KNN	68.00%	71.43%	71.43%	66.08%
Naive Bayes	60.00%	66.67%	57.14%	62.99%
LGBM	56.00%	61.54%	57.14%	58.44%

To overcome this limitation, we tried manually cleaning the data and establishing the link between text and video inputs to train the model. Once the model was trained, we applied it to the test cases to evaluate its performance. Moreover, during training, we also validated the model after each epoch using the validation set that is composed of 20% of the unseen training data (i.e., the data that which model does not see during training). The training and validation performance of the model is shown in terms of accuracy and loss curves reported in Figure 2. Moreover, the model's performance at the inference stage is reported in Figures 3, 4

From Figure 2, we can observe that although the model produces lower model prediction error and performance variance. It cannot outperform state-of-the-art works. For example, at the inference stage, the proposed model achieved an accuracy of 65%, with precision, recall, and AUC scores of 71.43%, 50%, and 56%, respectively. The difference in the performance of the model during the training and testing phase can be explained by the fact that the training dataset was not rich and large enough to allow the model to fully learn the feature representations of both deceptive and truthful classes, which led the model to produce overfitting results during training and validation stage.

Furthermore, the used 'Real-Life' dataset is not designed to train the multimodal networks as it does not contain the multimodal training, validation, and test samples of both deceptive and truthful classes. In the future, we envisage the model to perform better once it's trained on a better quality, and large-scale dataset containing multiple video and textual inputs for each case.

These results underscore the importance of model selection in the field of deception detection. The MLP's strong performance across all metrics suggests that it is well-suited for this dataset, potentially due to its ability to model non-linear relationships between features. On the other hand, models with lower AUC scores may be intrinsically less capable of handling the complexity of the deception detection task.

B. PERMUTATION IMPORTANCE

Permutation importance technique assesses feature importance by evaluating the decrease in a model's performance

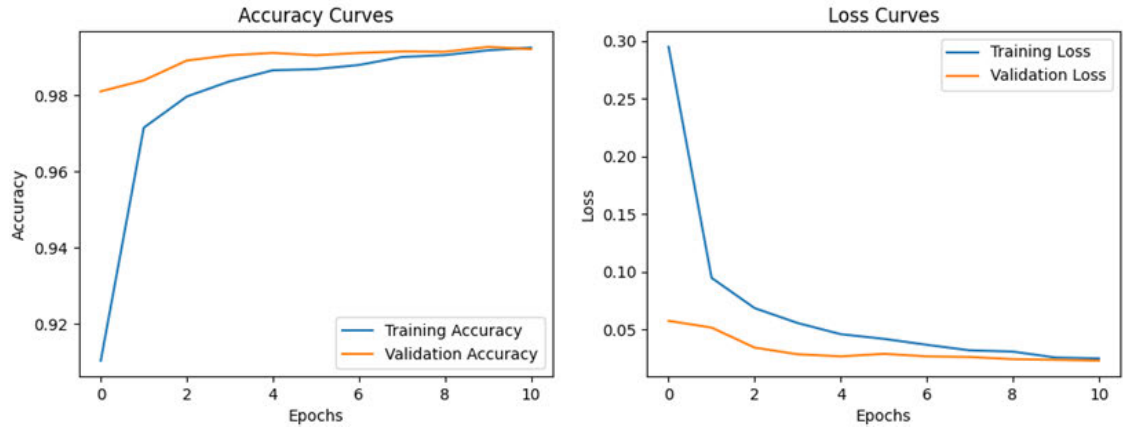

FIGURE 2. Training and validation Performance in terms of accuracy and sparse cross-entropy loss curves.

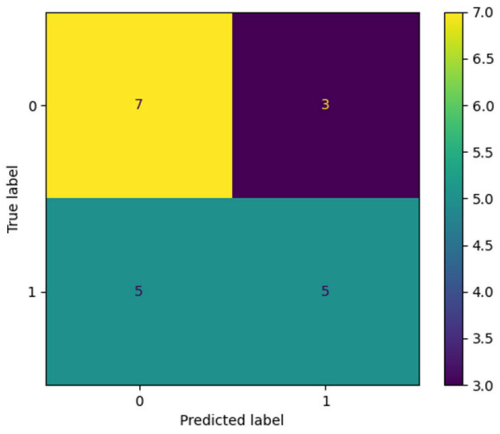

FIGURE 3. Confusion Matrix. 0: Truthful Events, 1: Deceptive Events using multimodal.

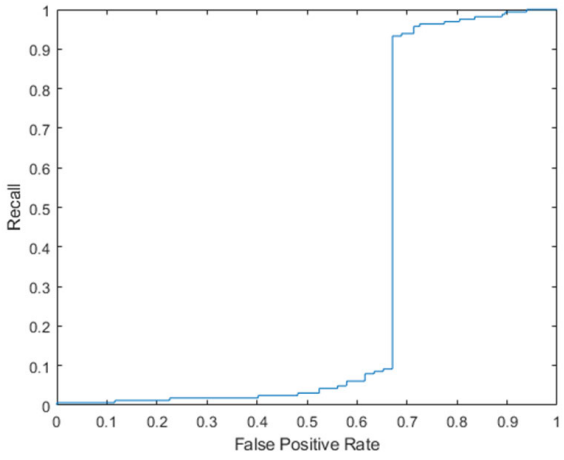

FIGURE 4. ROC curve using multimodal.

when the values of a single feature, or a group of features - as this study proposed- are randomly shuffled. This is extensively explained in Methodology section. Three specific methodologies are explored: using eli5 library [65], analyzing single-feature permutation importance created in this study, and introducing a novel approach for set-of-features permutation importance.

1) FEATURES CORRELATION

An important aspect to be noted is that permutation importance will not yield authentic results if a feature, A, is highly correlated with another feature, B. This is because feature B will compensate A's permutation effect. Therefore, it is important to find the correlation matrix, as shown in Fig. 5.

It can be noticed that the common trend is unrelated features. However, there is an occasional high negative correlation between features that are exact opposites to each others, such as open mouth and close mouth. This emphasises on the need of the proposed novel group permutation importance.

2) PERMUTATION IMPORTANCE USING LIBRARY ELI5

An open-source library is eli5 library that computes permutation importance. This method involves a systematic shuffling of each feature's values and measuring the impact on the model's accuracy. The permutation importances for the top four models—MLP, SVM, XGBoost, and Catboost—are presented in Fig. 6, with features organized in ascending order based on their importance, transitioning from red to green. Only two features with a permutation importance of zero (white-colored) are displayed, but all features with non-zero permutation importance are shown. In this context, a green hue indicates a positive permutation importance, whereas red indicates a negative permutation importance. In Fig. 6 the permutation importances for the four leading models are arranged from left to right: MLP, SVM, XGBoost, and Catboost.

The analysis of permutation importances reveals that attributes related to eyebrow movements—raising and frowning, specifically—are significant contributors to model

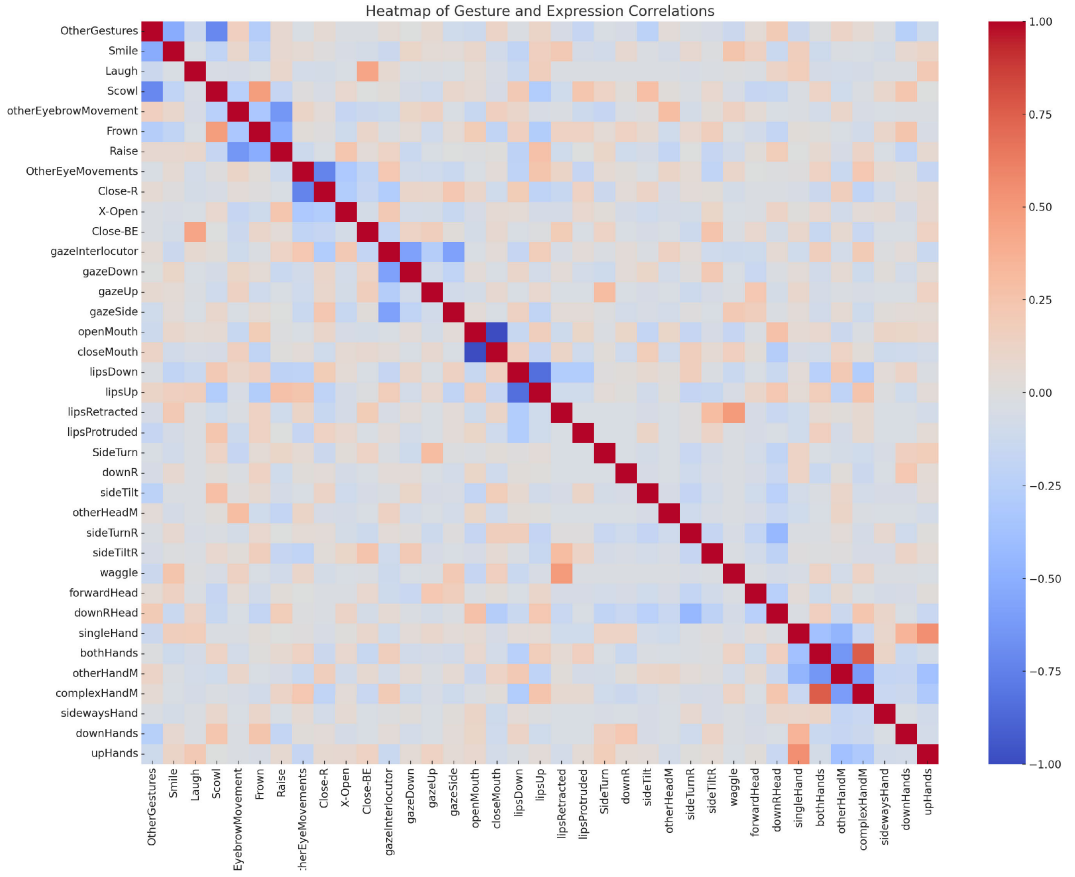

FIGURE 5. Real-life Trial dataset correlation matrix.

predictions. This consistency across all four models, MLP, SVM, XGBoost, and Catboost, strongly suggests the pivotal role these features play, beyond coincidence.

MLP distinguishes itself by utilizing the largest number of important features, with the most influential feature contributing 16%, followed by two features at 12%, six features at 8%, and nine features at 4%. This indicates that the MLP has a total of eighteen notable features. Notably, in the MLP model, no feature exhibits a negative permutation importance, indicating an efficient utilization of attributes without any negative impact on MLP's performance. This could be reflective of MLP's complex architecture, which draws inspiration from neural networks in the human brain, allowing it to exploit a broader range of inputs effectively.

In contrast, the XGBoost model demonstrates a more selective feature reliance, with a smaller number of features with positive permutation importance and only one feature, 'Smile,' presenting a negative value. This suggests a focused approach in XGBoost's decision-making process, utilizing a small number of features. On the other hand, Catboost, while leveraging a broader array of features than XGBoost and SVM, also includes high number of features that negatively influence its predictions.

SVM displays a comparatively higher ratio of features with negative permutation importance, which may be attributed to its algorithmic structure. Unlike MLP, the SVM's linear nature may not capture complex patterns without sufficient margin for error, potentially leading to misclassifications when influenced by certain features.

Across all models, a majority of features exhibit zero permutation importance. This highlights a general limitation in ML models' capacity to integrate a high number of features effectively. This analysis underscores the importance of feature selection and model tuning. Understanding the nature and impact of features can drive more refined modeling approaches and encourage the development of strategies to mitigate the inclusion of non-contributing or negatively impacting features in ML models.

3) SELF-IMPLEMENTED SINGLE-FEATURE PERMUTATION IMPORTANCE

The purpose of self-implemented single-feature permutation importance technique is to replicate eli5 outcomes and to gain a comprehensive understanding of permutation importance, a single-feature permutation importance method is implemented manually. The purpose of this

FIGURE 6. eli5 library Permutation importance results for four models. From the left: MLP, SVM, XGBoost, Catboost.

implementation is not to showcase new insights other than the ones provided in the previous section, but to validate our grasp of the permutation importance concept. This approach is considered as a bridge towards the development of the innovative method for set-of-features permutation importance in the next subsection. In short, the implementation serves as a tool for reinforcing theoretical concepts through practical application and paving the way for the introduction of our novel approach. It is important to note that the results from our permutation importance implementation may not replicate identically those from the `eli5` library, due to the inherent randomness of the permutation process.

Figure 7 presents the permutation importance by the MLP model, identifying 15 features with positive importance and none with negative importance. For the SVM model, as illustrated in Figure 8, there are 13 features with positive importance and 6 with negative importance.

Figure 9 showcases the XGBoost model's permutation importance, featuring four positively important features and an equal number of negatively important features. In contrast, the Catboost model, depicted in Figure 10, comprises ten

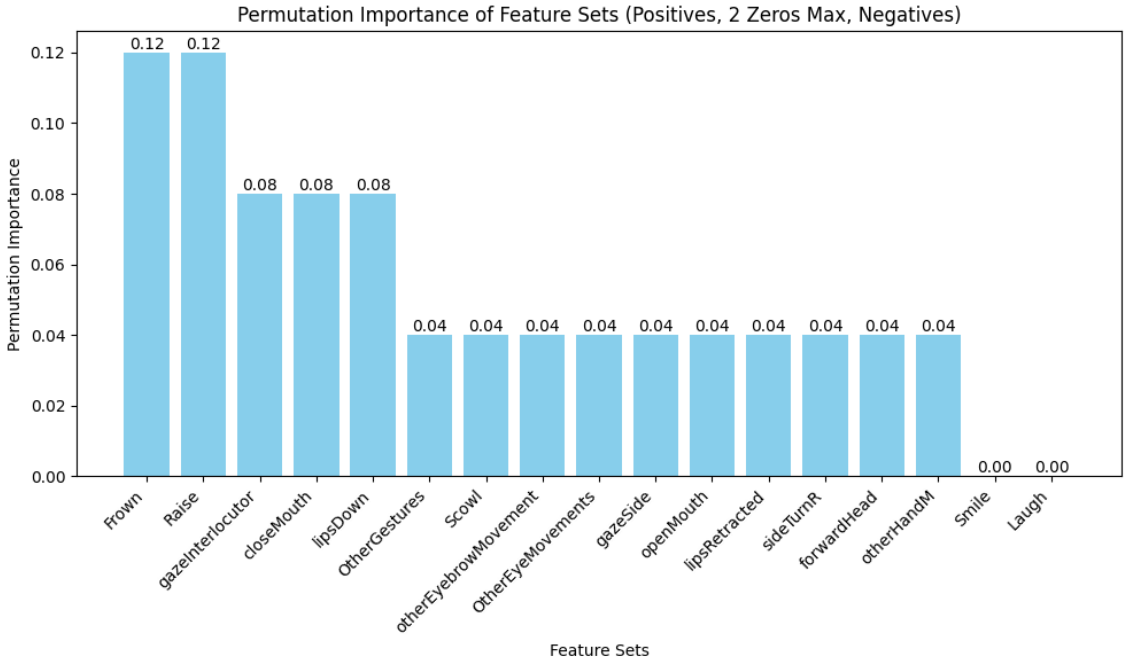

FIGURE 7. MLP model results: self-implemented permutation importance.

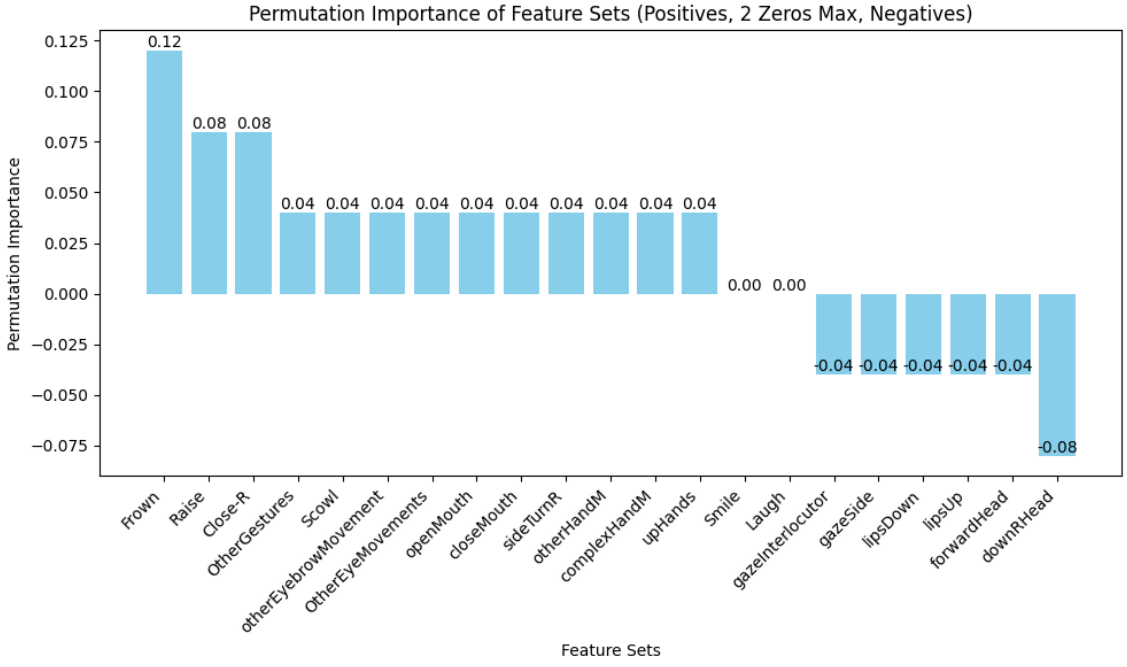

FIGURE 8. SVM Model Results: Self-Implemented Permutation Importance.

features with positive importance and two with negative importance.

The goal of this self-implemented technique is achieved: As with the `eli5` findings, the features related to eyebrow movements, specifically raising and frowning, are consistently deemed significant across all four models, underscoring their substantial importance. Table 5 corroborates the accuracy of our understanding regarding single-feature permutation importance and provides a strong foundation for proceeding with the novel set-of-features permutation importance technique.

4) NOVEL SET-OF-FEATURES PERMUTATION IMPORTANCE

The motivation for this step is that, the dataset consists of features in groups, as shown in Table 6. Each group

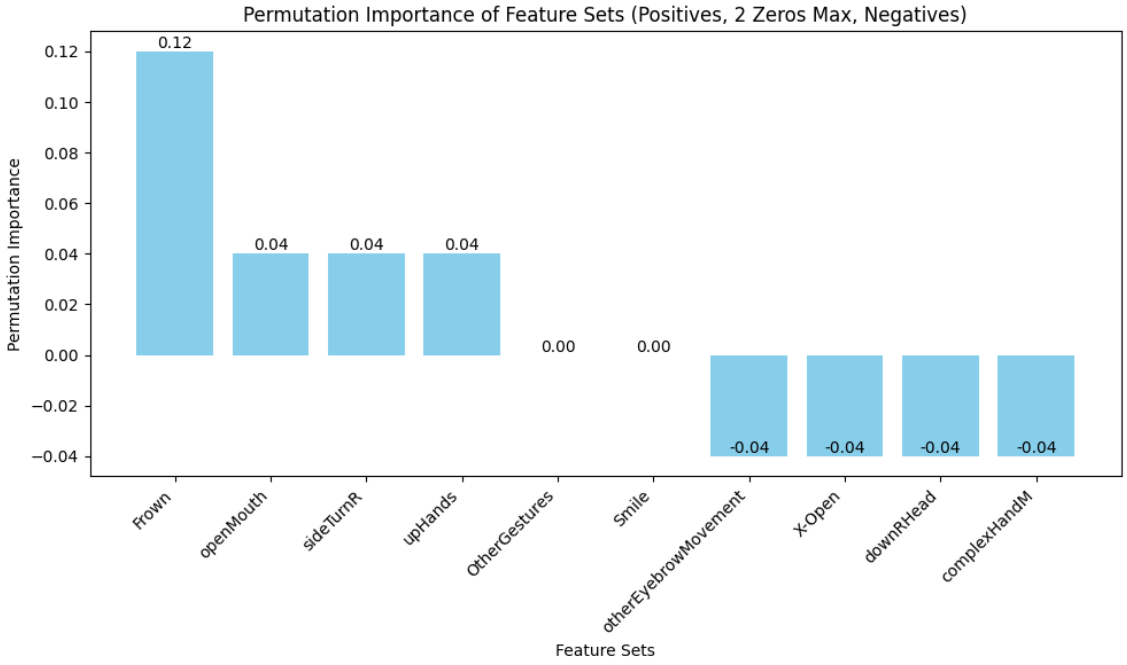

FIGURE 9. XGBoost model results: self-implemented permutation importance.

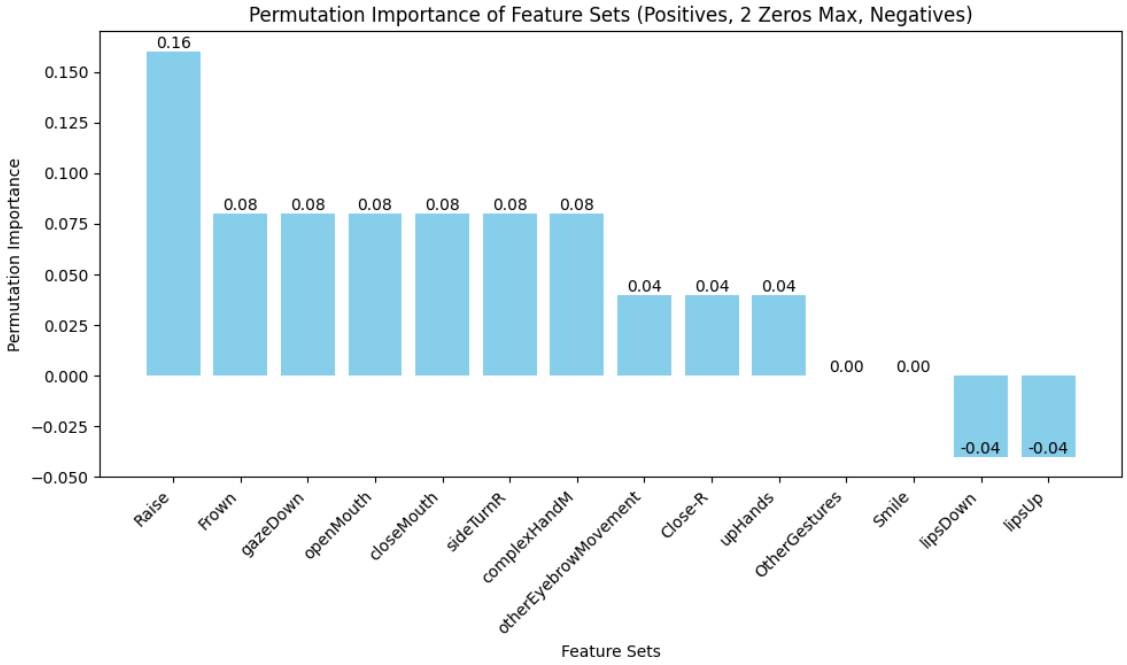

FIGURE 10. Catboost model results: self-implemented permutation importance.

shall have only one feature activated (1), and all remaining features deactivated (0s). For example, in Gestures group, if Scowl is activated, Laugh, Smile, and OtherGestures must be deactivated. Similarly, in Eyebrows group: if Raise is activated, Frown and other eyebrowmovement shall be deactivated. This means, that permutating each feature individualy, will provide unrealistic inputs, such as someone raising and frowning her eyebrows at the same instance, which is impractical in a realistic scenario. To address this, we need to permute each group as a whole rather than on an individual feature basis.

TABLE 5. Comparison of permutation importance results aspects between eli5 library and our implementation.

Permutation Importance Aspect	eli5	our Implementation
MLP Highest number of positive features	✓	✓
MLP scarcity of negative features	✓	✓
XGBoost lowest number of positive features	✓	✓
SVM Highest number of negative features	✓	✓
Eyebrows frowning and raising top importance	✓	✓

TABLE 6. Real Life Trial Dataset's Groups, and their corresponding features.

Group	Features
Gestures	OtherGestures, Smile, Laugh, Scowl
Eyebrows	otherEyebrowMovement, Frown, Raise
Eyes	OtherEyeMovements, Close-R, X-Open, Close-BE
Gaze	gazeInterlocutor, gazeDown, gazeUp, otherGaze, gazeSide
Mouth	openMouth, closeMouth, lipsDown, lipsUp, lipsRetracted, lipsProtruded
Head	SideTurn, downR, sideTilt, backHead, otherHeadM, sideTurnR, sideTiltR, waggle, forwardHead, downRHead
Hand	singleHand, bothHands, otherHandM, complexHandM, sidewaysHand, downHands, upHands

For the MLP model, Figure 11 demonstrates that permutations within the 'Eyebrows' group result in a significant accuracy reduction of 36%. In addition, all groups positively influence MLP's decisions, indicative of the model's ability to effectively utilize all input features.

The group permutation importance for the SVM model is presented in Figure 12. The 'Eyebrows' group emerges as the most influential, with a permutation importance of 0.32, followed by the 'Hand' group at 0.24. Although there are no non-contributing groups, the 'Mouth' group is observed to have a negative impact on the model's performance, with a permutation importance of −0.04.

For XGBoost, as shown in Figure 13, the 'Eyebrows' group remains the most significant with a permutation importance of 0.20. The 'Eyes' and 'Head' groups both have an importance of 0.08. Notably, the 'Gestures' and 'Gaze' groups do not contribute to the model's accuracy, and the 'Hand' group worsens it, with a negative permutation importance of −0.04.

Lastly, Figure 14 illustrates the permutation importance for the Catboost model. Here, the 'Hand' group is the most contributing with a permutation importance of 0.24, while the 'Eyebrows' group is the second most significant at 0.20. The 'Gaze' group does not contribute to the model, and the 'Mouth' group negatively affects the model with a permutation importance of −0.08.

The permutation importance statistics, as detailed in Table 7, emphasize the significance of certain feature groups in the best-performing models. Notably, 'Eyebrows' consistently ranks at the forefront, except in the Catboost model where it is second. The 'Hand' group also demonstrates high importance, consistently appearing in the top half of all models except for XGBoost. 'Eyes' generally features in the top half, with the exception of SVM where it ranks fourth. This pattern suggests that the 'Eyebrows,' 'Hand,' and 'Eyes' groups are likely to be highly indicative of deceptive behavior.

TABLE 7. Group's permutation importance occurrence summary.

Groups	Best One	Best Half	Positive	Zero	Negative
Eyebrows	3	4	4	0	0
Eyes	0	3	4	0	0
Hand	1	3	3	0	1
Head	0	2	4	0	0
Gestures	0	0	3	1	0
Gaze	0	0	2	2	0
Mouth	0	0	2	0	2

Conversely, there is no consistent pattern of negative or zero contribution across any group, indicating a general utility in all groups. The 'Mouth' group is the least beneficial, with negative contributions in two instances. Nevertheless, the overall trend shows positive contributions from all groups, indicating that no single group hinders the input set. Based on these observations, it is advisable not to exclude any group from the dataset in future modeling efforts.

Table 8 demonstrates the impact of permutation on the performance of the best four models across three distinct stages: eli5, self-implemented, and group permutation importance techniques. Notably, MLP consistently leads in positive utilization according to the permutation importance criterion across all three techniques. This dominance is likely due to MLP's complex architecture, which enables it to process a large number of inputs effectively. This observation is further supported by MLP's lack of reliance on any features with negative importance in all three techniques, underscoring its robustness in feature selection.

Conversely, SVM exhibits the highest degree of negative permutation importance across all techniques, hinting at its simplicity and potential limitations in handling all available inputs within its algorithmic structure. This shall not be understood as a huge drawback, but instead, as a potential limitation of SVM.

XGBoost displays the lowest levels of positive permutation importances, indicating a selective approach to input utilization. This characteristic indicates that XGBoost may prioritize a smaller set of highly impactful features over a broader but less effective range. On the other hand, Catboost demonstrates a capacity to engage a significant number of inputs, albeit with a moderate occurrence of negative importance. This suggests that while Catboost is adept at incorporating a wide array of features, it may occasionally incorporate inputs that detract from model performance.

Overall, these findings illuminate the diverse strategies employed by different classifiers in navigating the complexities of feature permutation, each with its unique strengths and areas for improvement.

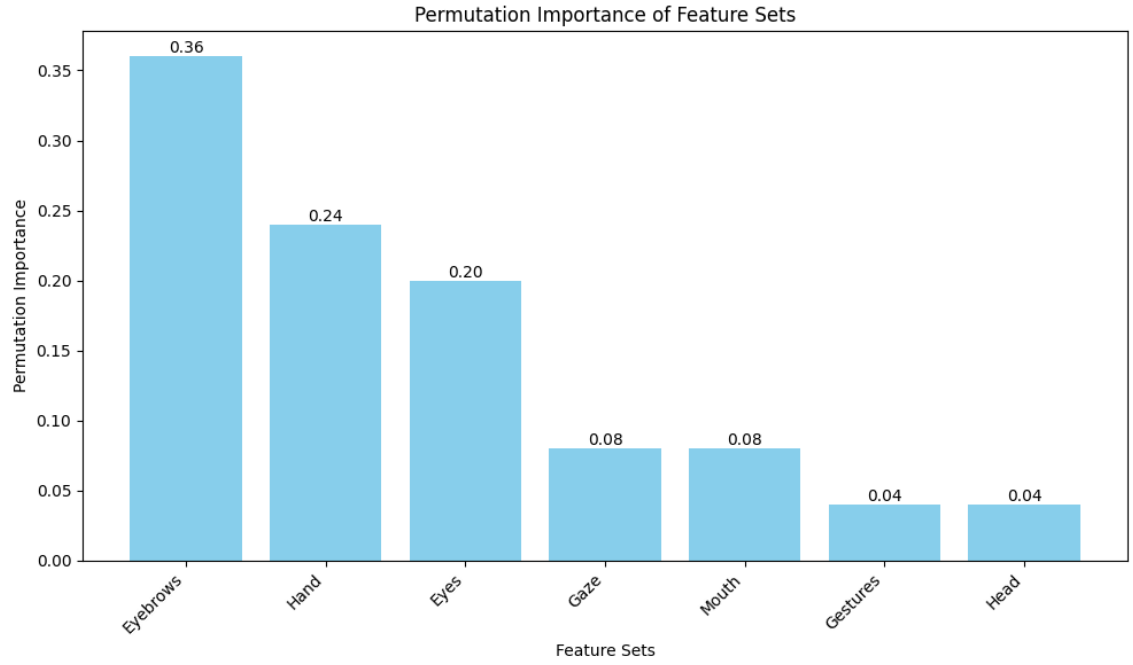

FIGURE 11. MLP model results: Set-of-features permutation importance.

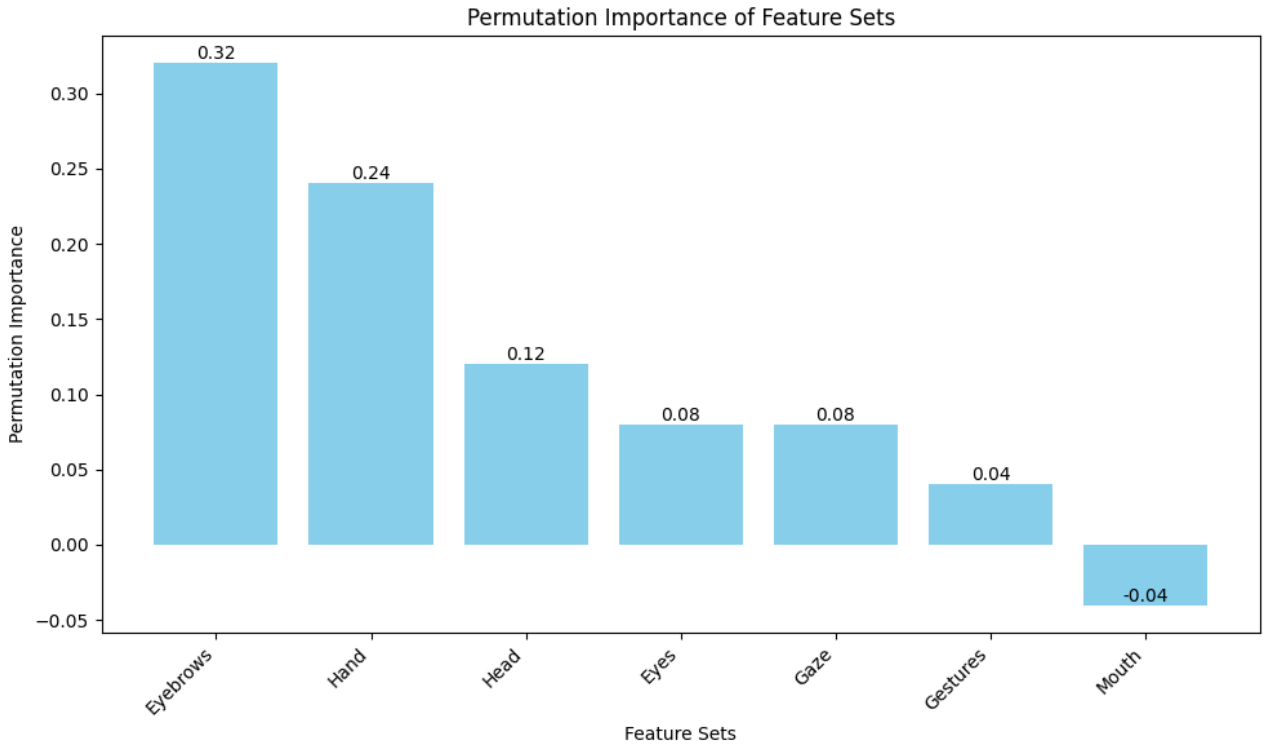

FIGURE 12. SVM model results: Set-of-features Permutation Importance.

C. PARTIAL DEPENDENCE PLOTS

PDP as extensively explained in Methodology section is utilized to examine how the class predictions depend on features. PDP technique facilitates this analysis through two primary approaches. The first approach evaluates the influence of a single feature on the target class, providing

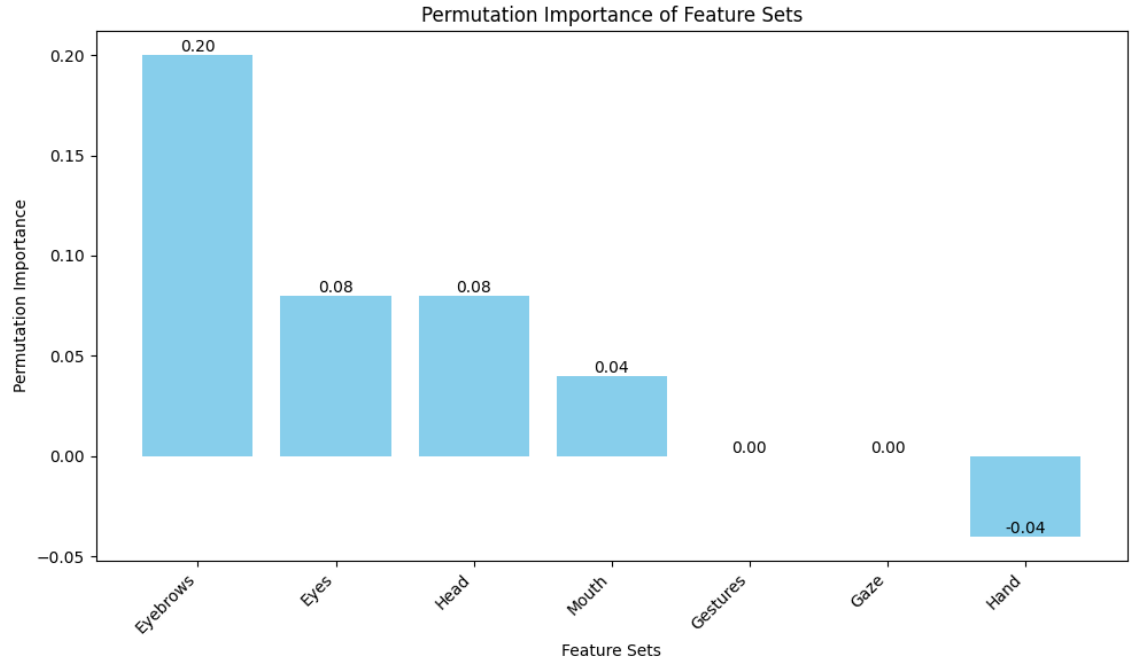

FIGURE 13. XGBoost model results: Set-of-features permutation importance.

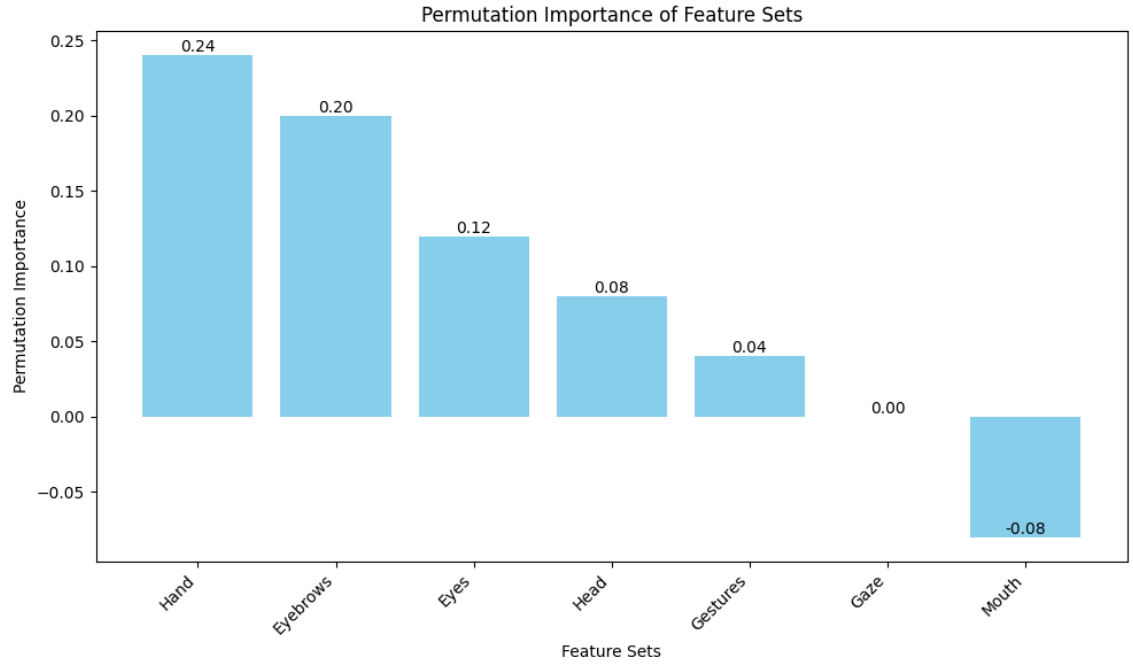

FIGURE 14. Catboost Model Results: Set-of-features Permutation Importance.

insights into the direct relationship between the feature and the predicted outcome. The second method explores the interaction between two features and how this interaction impacts the target class, offering a deep understanding of how feature combinations affect predictions.

1) SINGLE FEATURE IMPACT

This subsection examines how individual features influence the model's predictions. By altering a single feature's values and observing the variance in class predictions, the specific contribution of each feature can be identified, highlighting

TABLE 8. Comparative analysis of feature utilization by classifiers.

Technique	Importance	MLP	SVM	XG-Boost	Cat-boost
Groups (out of 7)	Positive	7	6	4	5
	Zero	0	0	2	0
	Negative	0	1	1	1
eli5 (out of 39)	Positive	18	8	6	13
	Zero	21	27	32	17
	Negative	0	4	1	4
Self-Implemented (out of 39)	Positive	15	13	4	10
	Zero	22	20	31	27
	Negative	0	6	4	2

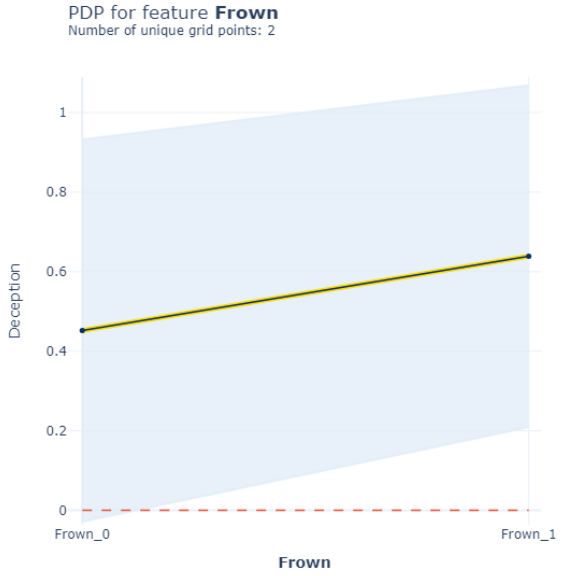

FIGURE 15. MLP model's PDP result of 'Frown' feature.

the features with the most significant impact on the model's output.

a: MLP'S PDPS

In the MLP's classifier, Figure 15 is the PDP of the 'Frown' feature. Following are important aspects to note in the PDPs:

b: SIGN OF PDP SLOPE

If the PDP's slope is positive, as observed in the PDP for 'Frown' in Figure 15, this indicates that 'Frowning' tends to signal deception. Conversely, a negative slope suggests that the feature is more likely associated with truthful instances. An example of this is illustrated in Figure 16, where the 'Raise' feature demonstrates a negative slope, implying its association with truthfulness. From these observations, it can be deduced that frowning is indicative of deception, whereas raising the eyebrows suggests truthfulness.

c: MAGNITUDE OF PDP SLOPE

The magnitude of the slope in PDP serves as an indicator of a feature's impact on the model's predictions. According to the results derived from permutation importance techniques, 'Frown' and 'Raise' emerge as the most significant features

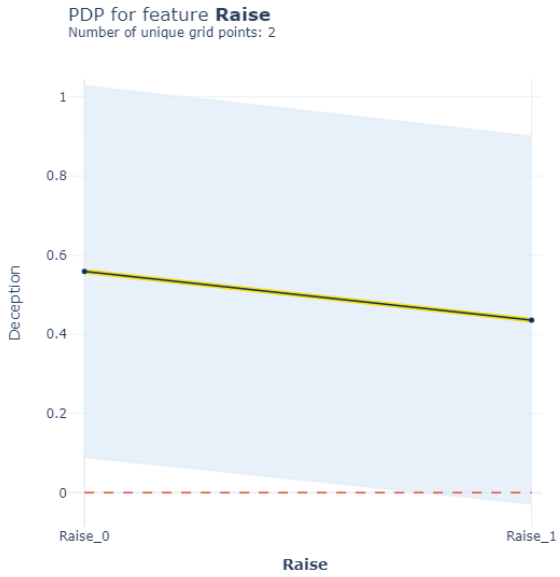

FIGURE 16. MLP model's PDP result of 'Raise' feature.

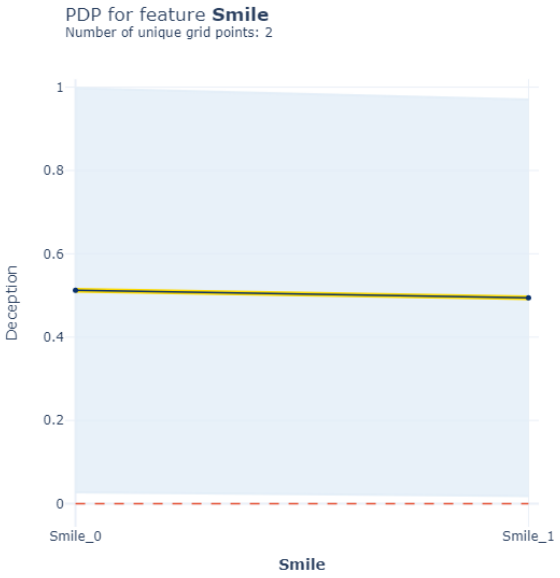

FIGURE 17. MLP model's PDP result of 'smile' feature.

as shown in Figure 15. To further explore this, consider examining a feature that exhibits no permutation importance on MLP's performance, such as 'Smile'. Figure 17 presents a nearly horizontal slope for 'Smile', signifying its minimal impact on the target class.

On the other hand, the steep lines of 'Frown' and 'Raise' assure their high impact on the output.

d: LIGHT SHADE OF BLUE

The light blue shading in PDP reflects the model's confidence level in its decision-making process. Given that all the PDPs discussed herein are derived from the MLP model, it's

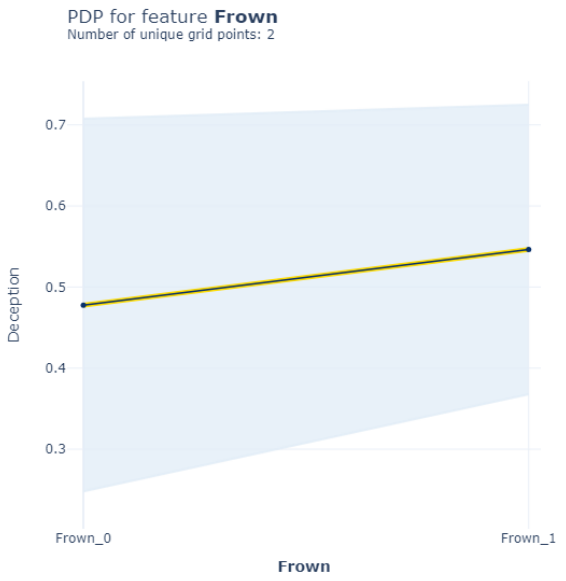

FIGURE 18. Random forest model's PDP result of 'frown' feature.

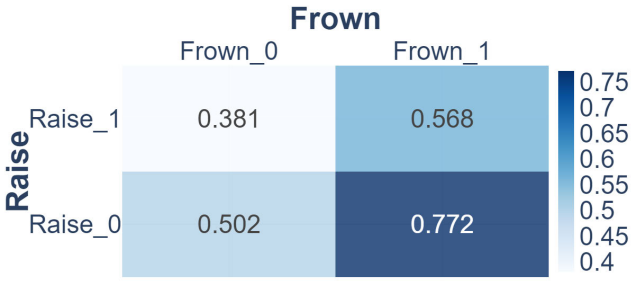

FIGURE 19. MLP Model's PDP interact result of frown&raise features.

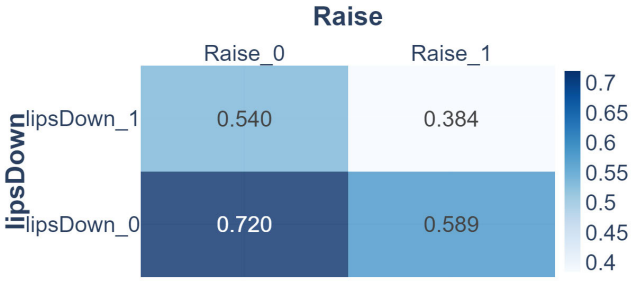

FIGURE 20. MLP Model's PDP Result Interact of Lips&Raise Features.

noteworthy that the MLP consistently exhibits a high degree of confidence in its decisions, irrespective of their accuracy.

e: RANDOM FOREST'S PDPS

Despite the Random Forest classifier not being among the top four classifiers in this study, we performed its PDP analysis due to its lower confidence levels in decision-making. Figure 18 illustrates the PDP for 'Frown' with a notably thinner light blue shaded area, indicating reduced confidence in its predictions. While the impact of 'Frown' on indicating deception the effects observed in the MLP and other top classifiers, Random Forest exhibits less confidence. It is noteworthy that the PDPs for the best-performing models are similar to those of the MLP, suggesting little motivation to present them individually due to their similarity.

2) FEATURE INTERACTION IMPACT

The feature interaction impact analysis goes beyond single features to explore how two features work together to affect model's predictions. This analysis identifies interactions that significantly influence the model's performance, revealing complex dependencies not apparent when considering features in isolation.

Figure 19 presents the Interact PDP for 'Raise' and 'Frown' as analyzed by the MLP model. The plot elucidates that in scenarios where 'Frown' is active (1) and 'Raise' is inactive (0), the prediction leans towards deception with a scale of 0.772. Conversely, activating 'Raise' (1) while 'Frown' remains inactive (0) shifts the prediction towards truthfulness, indicated by a scale of 0.381. These observations are in alignment with the insights derived from individual PDP analyses.

Figure 20 illustrates the MLP's interaction PDP for two influential features from different groups: 'Raising' eyebrows and the corners of the 'Lips' turning down. These features, each associated with truthful decisions, create four real-life scenarios when combined. Notably, when both features are active, the model leans towards a deception probability of 0.384. Conversely, deactivating both features increases the deception probability to 0.72. In cases where one feature is active and the other is not, the outcome resides in between, slightly tilting towards deception. This analysis underscores the significant impact each feature holds within the model's decision-making process. The interaction PDP between 'Frowning' and 'Smiling' encompasses four realistic scenarios, with one feature exhibiting a notably lower impact. Utilizing the XGBoost model, Figure 21 reveals the limited influence of 'Smiling' on the predictive outcome. Specifically, when 'Frown' is active, the deception probability remains at 0.686, irrespective of 'Smile's' presence or absence. Conversely, deactivating 'Frown' shows that variations in 'Smile' adjust the deception probability by a mere 0.006, underscoring its minimal effect. This highlights the dominant role of 'Frowning' over 'Smiling' in affecting the model's decision-making process.

Across these interactions, it's evident that certain features play pivotal roles in shaping the model's decision-making process, with some interactions revealing a significant influence on the predictive outcome. The findings underscore the complexity of model behavior, highlighting how specific feature combinations can either increase or decrease the probability of deception or truthfulness. These insights illuminate the nuanced understanding of feature interplay within models and underscore the utility of PDP Interact analyses in unraveling these complex relationships.

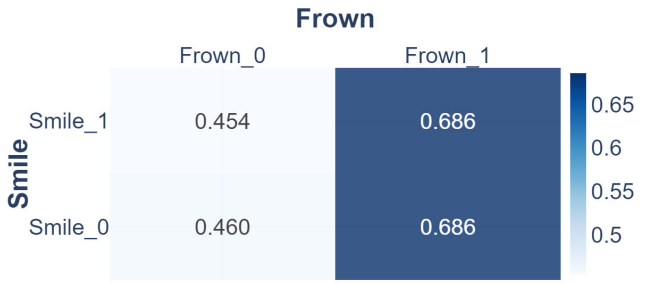

FIGURE 21. XGBoost Model's PDP Interact Result of Smile&Frown features.

D. SHAP (SHAPLEY ADDITIVE EXPLANATIONS)

As extensively explained in Methodology section, SHAP is a game theory-based approach for explaining the output of any ML model. It decomposes predictions into contributions from each feature, offering both local and global explanations.

1) LOCAL SHAP

Local SHAP explanations provide insights into the decision-making process for individual predictions. By attributing a SHAP value to each feature, it elucidates how each feature contributes to the specific prediction, offering a detailed explanation at the instance level.

As an elaboration, Figure 22, represents SHAP explanation of a truthful instance. The red rows are the features adding weight to 'Deception', and blue ones add weight to the 'Truthful' class. 'Frown'=1 and 'Raise'=0 indicate that this row should be deceptive. However, blue features, such as 'Scowl'=1, and others, aggregate to have the higher impact on the decision in this local instance. Therefore, this instance has a score of 0.03, which is highly truthful. As previously discussed, MLP often produces extreme scores due to its high confidence.

As shown in Figure 23, the decision made by the SVM on the same instance corroborates the findings of the MLP, albeit with minor discrepancies. For instance, the SVM classifies 'Gazedown' as blue and omits the 'closed mouth' feature, which is marked as red. The associated score of 0.16 indicates that the prediction is considered 'Truthful', yet it reflects a lower confidence level compared to the MLP, which yields a score of 0.03. It is important to note that a high confidence level does not necessarily imply accuracy of a prediction.

2) GLOBAL SHAP

Global SHAP, in contrast, aggregates SHAP values across all instances to provide a holistic view of feature importance. This global perspective highlights overall trends and patterns in the data, offering a comprehensive understanding of the model's reliance on different features for making predictions. Figure 24 showcases the Global SHAP values for the MLP classifier. Each feature is represented by 121 dots, signifying individual instances, with colors indicating feature activity: blue for inactive and red for active states. Notably, 'Frowning' exhibits a significant impact when active, whereas 'Raising' demonstrates importance regardless of its state. Features like 'OtherGestures' and 'Gaze Interlocutor' exhibit the ability to influence predictions towards both deception and truthfulness, active or not, showcasing the model's non-linear behavior. Some features, such as 'SideTurn', typically show minimal SHAP impact when inactive but reveal substantial influence upon activation, highlighting their conditional importance.

Conversely, Figure 25 reveals notable differences, corroborating insights from the 'Permutation Importance' section. Studying MLP's Global SHAP with that of XGBoost, the comparison reveals that the lower-ranked features by the MLP model exhibit noticeably higher SHAP values than those in the XGBoost model, underscoring XGBoost's selective feature utilization. Meanwhile, features with the highest SHAP values in XGBoost indicate a more focused exploitation of key features compared to MLP, aligning with the observed behavior that XGBoost prioritizes a narrower set of features for making predictions.

E. DISCUSSIONS

The exploration of feature interaction through PDP and SHAP analyses revealed nuanced insights into how combinations of features influence model predictions. For instance, the interaction between 'Frowning' and 'Smiling' highlighted the significant impact of 'Frowning' on deception detection, underscoring the importance of understanding feature interactions in improving model interpretability and performance.

Moreover, the application of global SHAP analysis offered a holistic view of feature importance, reaffirming the critical roles of 'Frowning' and 'Raising'. These global insights are crucial for understanding the overall behavior of models and guiding feature selection and model tuning processes.

In line with the objectives of this report, a literature review was conducted to identify the gap that needs addressing, which is the implementation of XAI techniques to gain insights into real-life visual cues of deception. The XAI techniques employed include permutation importance, PDP, and SHAP. These techniques are applied to the best-performing classifiers selected from a diverse set of relevant ML models, highlighting model behavior and the significance of specific features, such as eyebrow characteristics. The top-performing classifiers are MLP, SVM, XGBoost, and CatBoost, achieving accuracies of up to 88%.

In conclusion, the findings from this section enhance the understanding of the predictive dynamics of ML models in deception detection and emphasize the value of XAI techniques in bridging the gap between raw model output and intrepretable insights. The demonstrated superiority of the MLP model, coupled with the detailed interpretative analyses, lays a solid foundation for future research aimed

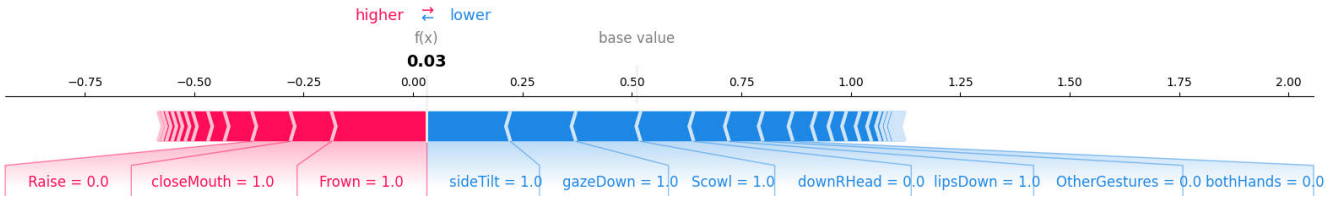

FIGURE 22. MLP model's local shap result on a truthful instance.

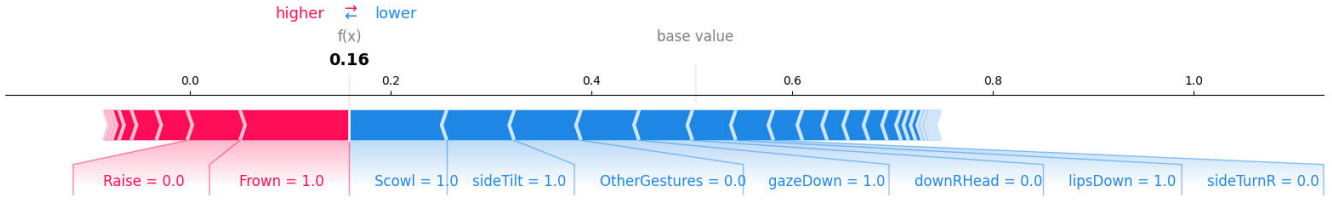

FIGURE 23. SVM model's local shap result on a truthful instance.

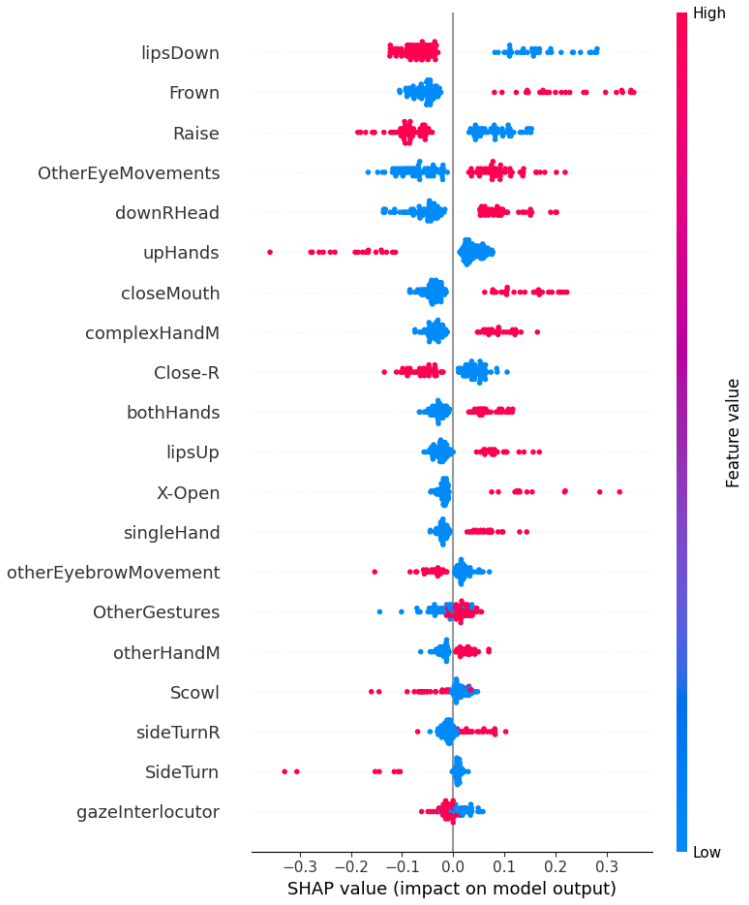

FIGURE 24. MLP model's global shap result.

at refining models for greater accuracy and interpretability in the field of deception detection. Table 9 showcases the advancements our study contributes to the literature compared to state-of-the-art studies, particularly those that employed XAI techniques and used the Real-Life Trial dataset.

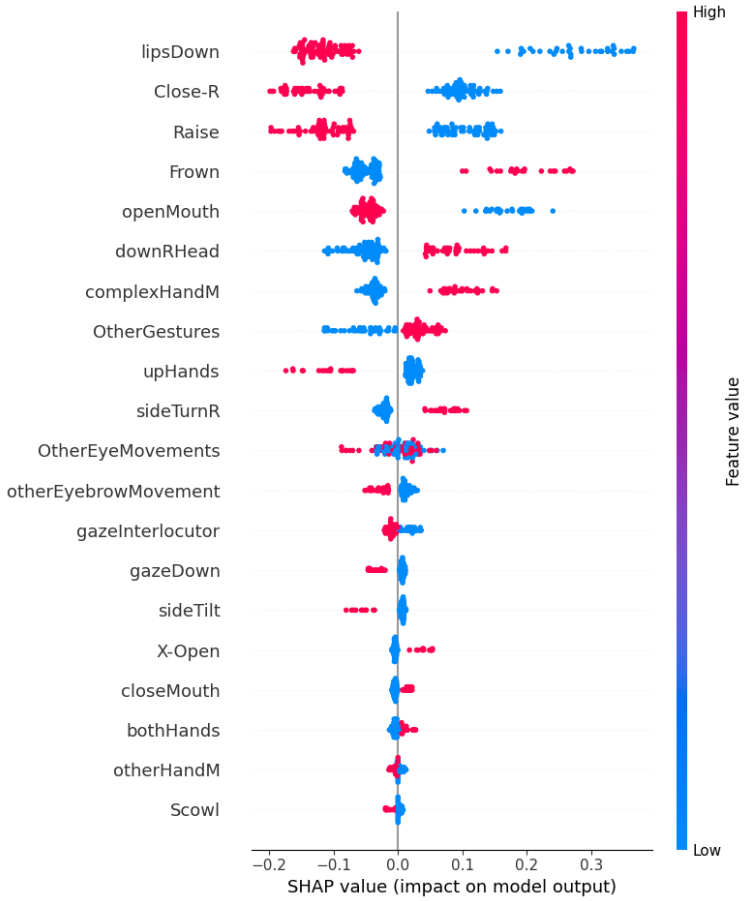

FIGURE 25. XGBoost model's global shap result.

TABLE 9. Comparison of this work with the state-of-the-art.

Ref	Algorithm Used	Accuracy	XAI Techniques Used	Dataset Used
This Work	MLP, SVM, XGBoost, Catboost	**88%**	Permutation Importance, PDP, SHAP, Set-of-Features Permutation Importance	Real-Life Trial Dataset
[29]	CNN	68%	None	Real-Life Trial, Low-Stakes Deceit datasets
[33]	Several ML methods	73%	LIME	Crowdsourced dataset (statements on planned activities)
[34]	SVM	87%	None	Ott Deceptive Opinion Spam dataset
[36]	Neural Network	83%	None	Real-Life Trial dataset

VI. CONCLUSION AND FUTURE WORKS

In addressing the complex challenge of deception detection, this study bridges the gap between the significant capabilities of ML classifiers and their black-box nature. By harnessing the power of model-agnostic XAI techniques within a methodological framework, this research advances the understanding of the interplay between deception visual cues and artificial intelligence.

The methodology, centered around the real-life trial dataset, underscored the dataset's value in reflecting genuine high-stakes environments. Through meticulous preprocessing, model selection, and hyperparameter tuning, this study has underscored the importance of a visual deception detection accuracy.

Among the various ML models evaluated, the Multi-Layer Perceptron (MLP) model emerged as the standout performer, demonstrating a high accuracy of 88% and an exceptional ability to utilize facial and body expressions as predictive cues. This finding validates the chosen methodological approach and highlights the potential of specific non-verbal

cues, such as eyebrow 'Frowning' and 'Raising,' in indicating deceptive behavior.

The application of model-agnostic XAI techniques has been instrumental in peeling back the layers of AI-driven decisions, offering insightful interpretations of how individual features and their interactions contribute to the detection of deception. This level of analysis, facilitated by permutation importance, PDP, and SHAP, has illuminated the path toward models that are accurate and interpretable.

Limitations in this study to be addressed in the future include: the lack of automation in extracting visual features from subjects and the exclusive focus on visual cues, without incorporating other types of cues.

Future research in deception detection could benefit from the incorporation of advanced deep learning techniques to discern subtle changes in deception cues, potentially enhancing accuracy significantly. Employing deep transfer learning with pre-trained models could further refine this process, leveraging existing knowledge bases for improved performance. Additionally, a more meticulous selection of cues, alongside the integration of a broader spectrum of physiological features, could augment the models' results. Expanding the dataset to include a larger, more diverse group of volunteers, providing multiple instances in more controlled environments, will be crucial in generating authentic deceptive cues.

ACKNOWLEDGMENT

This study is based on Suhaib Salah's master's thesis. AI was used to improve the English of the text.

REFERENCES

[1] S. Salah, T. Khater, E. Almajali, W. Khan, and A. Hussain, "Deception detection deep learning comprehensive system utilizing explainable AI," in *Proc. 16th Int. Conf. Develop. eSyst. Eng. (DeSE)*, Dec. 2023, pp. 713–720.

[2] K. Crockett, J. O'Shea, and W. Khan, "Automated deception detection of males and females from non-verbal facial micro-gestures," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2020, pp. 1–7.

[3] W. Khan, K. Crockett, J. O'Shea, A. Hussain, and B. M. Khan, "Deception in the eyes of deceiver: A computer vision and machine learning based automated deception detection," *Expert Syst. Appl.*, vol. 169, May 2021, Art. no. 114341.

[4] M. Shibili, C. V. Sreeshma, V. V. Prasad, C. Nikhilbinoy, and K. Neethu, "Design and development of automatic lie detector using Arduino," in *Proc. 3rd Int. Conf. Artif. Intell. Smart Energy (ICAIS)*, Feb. 2023, pp. 6–11.

[5] S. V. Fernandes and M. S. Ullah, "Use of machine learning for deception detection from spectral and cepstral features of speech signals," *IEEE Access*, vol. 9, pp. 78925–78935, 2021.

[6] M. Karnati, A. Seal, A. Yazidi, and O. Krejcar, "LieNet: A deep convolution neural network framework for detecting deception," *IEEE Trans. Cognit. Develop. Syst.*, vol. 14, no. 3, pp. 971–984, Sep. 2022.

[7] A. Wielgopolan and K. K. Imbir, "Cognitive load and deception detection performance," *Cognit. Sci.*, vol. 47, no. 7, Jul. 2023, Art. no. e13321.

[8] T. R. Levine, K. B. Serota, F. Carey, and D. Messer, "Teenagers lie a lot: A further investigation into the prevalence of lying," *Commun. Res. Rep.*, vol. 30, no. 3, pp. 211–220, Jul. 2013.

[9] T. R. Levine, "Truth-default theory and the psychology of lying and deception detection," *Current Opinion Psychol.*, vol. 47, Oct. 2022, Art. no. 101380.

[10] C. F. Bond, T. R. Levine, and M. Hartwig, "New findings in non-verbal lie detection," in *Detecting Deception: Current Challenges Cognit. Approaches*. Wiley, 2014, pp. 37–58.

[11] M. G. Frank and E. Svetieva, "Microexpressions and deception," in *Understanding Facial Expressions in Communication*. Springer, 2015, pp. 227–242.

[12] D. Matsumoto, H. C. Hwang, A. M. Fullenkamp, and C. Laurent, "Human deception detection from whole body motion analysis," *Air Force Res. Lab.*, vol. 1, pp. 1–55, Dec. 2015.

[13] K. Fukuda, "Eye blinks: New indices for the detection of deception," *Int. J. Psychophysiol.*, vol. 40, no. 3, pp. 239–245, Apr. 2001.

[14] N. Vogler and L. Pearl, "Using linguistically defined specific details to detect deception across domains," *Natural Lang. Eng.*, vol. 26, no. 3, pp. 349–373, May 2020.

[15] J. E. Driskell, E. Salas, and T. Driskell, "Social indicators of deception," *Hum. Factors, J. Hum. Factors Ergonom. Soc.*, vol. 54, no. 4, pp. 577–588, Aug. 2012.

[16] A. Ravindran, G. G. Krishna, Sagara, and S. Sarath, "A comparative analysis of machine learning algorithms in detecting deceptive behaviour in humans using thermal images," in *Proc. Int. Conf. Commun. Signal Process. (ICCSP)*, Apr. 2019, pp. 0310–0314.

[17] D. Kopev, A. Ali, I. Koychev, and P. Nakov, "Detecting deception in political debates using acoustic and textual features," in *Proc. IEEE Autom. Speech Recognit. Understand. Workshop (ASRU)*, Dec. 2019, pp. 652–659.

[18] M. Abouelenien, R. Mihalcea, and M. Burzo, "Analyzing thermal and visual clues of deception for a non-contact deception detection approach," in *Proc. 9th ACM Int. Conf. Pervasive Technol. Rel. Assistive Environments*, Jun. 2016, pp. 1–4.

[19] S. Islam, P. Saha, T. Chowdhury, A. Sorowar, and R. Rab, "Non-invasive deception detection in videos using machine learning techniques," in *Proc. 5th Int. Conf. Electr. Eng. Inf. Commun. Technol. (ICEEICT)*, Nov. 2021, pp. 1–6.

[20] P. Cunningham, M. Cord, and S. J. Delany, "Supervised learning," in *Machine Learning Techniques for Multimedia: Case Studies on Organization and Retrieval*. Cham, Switzerland: Springer, 2008, pp. 21–49.

[21] Z. Ghahramani, "Unsupervised learning," in *Summer School on Machine Learning*. Cham, Switzerland: Springer, 2003, pp. 72–112.

[22] O. Chapelle, B. Scholkopf, and A. Zien, Eds., "Semi-supervised learning (Chapelle, O. Et al., Eds.; 2006) [Book reviews]," *IEEE Trans. Neural Netw.*, vol. 20, no. 3, p. 542, Mar. 2009.

[23] D. Seland, "Garbage in garbage out," *Quality*, vol. 57, no. 5, p. 6, 2018.

[24] L. Gianfagna and A. D. Cecco, *Explainable AI With Python*, vol. 4. Cham, Switzerland: Springer, 2021.

[25] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, Oct. 2001.

[26] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001.

[27] S. Lundberg and S. Lee, "A unified approach to interpreting model predictions," in *Proc. Adv. Neural Inf. Process. Syst.*, Jan. 2017, pp. 4765–4774.

[28] N. Vance, J. Speth, S. Khan, A. Czajka, K. W. Bowyer, D. Wright, and P. Flynn, "Deception detection and remote physiological monitoring: A dataset and baseline experimental results," *IEEE Trans. Biometrics, Behav., Identity Sci.*, vol. 4, no. 4, pp. 522–532, Oct. 2022.

[29] L. M. Ngô, W. Wang, B. Mandira, S. Karaoglu, H. Bouma, H. Dibeklioglu, and T. Gevers, "Identity unbiased deception detection by 2D-to-3D face reconstruction," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021, pp. 145–154.

[30] S. Yildirim, M. S. Chimeumanu, and Z. A. Rana, "The influence of micro-expressions on deception detection," *Multimedia Tools Appl.*, vol. 82, no. 19, pp. 29115–29133, Aug. 2023.

[31] M. Kamboj, C. Hessler, P. Asnani, K. Riani, and M. Abouelenien, "Multimodal political deception detection," *IEEE MultimediaMag.*, vol. 28, no. 1, pp. 94–102, Jan. 2021.

[32] R. Kadakia, P. Kalkotwar, P. Jhaveri, R. Patanwadia, and K. Srivastava, "Analysis of micro expressions using XAI," in *Proc. 3rd Int. Conf. Comput., Anal. Netw. (ICAN)*, Nov. 2022, pp. 1–7.

[33] L. Ilias, F. Soldner, and B. Kleinberg, "Explainable verbal deception detection using transformers," 2022, *arXiv:2210.03080*.

[34] V. Lai and C. Tan, "On human predictions with explanations and predictions of machine learning models: A case study on deception detection," in *Proc. Conf. Fairness, Accountability, Transparency*, Jan. 2019, pp. 29–38.

[35] D. Barsever, S. Singh, and E. Neftci, "Building a better lie detector with BERT: The difference between truth and lies," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2020, pp. 1–7.

[36] M. U. Sen, V. Pérez-Rosas, B. Yanikoglu, M. Abouelenien, M. Burzo, and R. Mihalcea, "Multimodal deception detection using real-life trial data," *IEEE Trans. Affect. Comput.*, vol. 13, no. 1, pp. 306–319, Jan. 2022.

[37] F. M. Talaat, "Explainable enhanced recurrent neural network for lie detection using voice stress analysis," *Multimedia Tools Appl.*, vol. 83, no. 11, pp. 32277–32299, Sep. 2023.

[38] P. Kodavade, S. Bhandigare, A. Kadam, N. Redekar, and K. P. Kamble, "Lie detection using thermal imaging feature extraction from periorbital tissue and cutaneous muscle," in *Innovations in Computer Science and Engineering*. Cham, Switzerland: Springer, 2021, pp. 643–650.

[39] A. Stathopoulos, L. Han, N. E. Dunbar, J. K. Burgoon, and D. Metaxas, "Deception detection in videos using robust facial features," in *Proc. Future Technol. Conf. (FTC)*, vol. 3, Oct. 2020, pp. 668–682.

[40] E. H. Neiterman, M. Bitan, and A. Azaria, "Multilingual deception detection by autonomous agents," in *Proc. Companion Web Conf.*, Apr. 2020, pp. 480–484.

[41] M. Ding, A. Zhao, Z. Lu, T. Xiang, and J.-R. Wen, "Face-focused cross-stream network for deception detection in videos," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 7794–7803.

[42] H.-C. Chou, Y.-W. Liu, and C.-C. Lee, "Joint learning of conversational temporal dynamics and acoustic features for speech deception detection in dialog games," in *Proc. Asia–Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC)*, Nov. 2019, pp. 1044–1050.

[43] R. Loconte, R. Russo, P. Capuozzo, P. Pietrini, and G. Sartori, "Verbal lie detection using large language models," *Sci. Rep.*, vol. 13, no. 1, p. 22849, Dec. 2023.

[44] H. Alaskar, "Hybrid metaheuristics with deep learning enabled automated deception detection and classification of facial expressions," *Comput., Mater. Continua*, vol. 75, no. 3, pp. 5433–5449, 2023.

[45] B. Biçer and H. Dibeklioğlu, "Automatic deceit detection through multimodal analysis of high-stake court-trials," *IEEE Trans. Affect. Comput.*, vol. 15, no. 1, pp. 342–356, Jan. 2024.

[46] N. Mansbach and A. Azaria, "Meta learning based deception detection from speech," *Appl. Sci.*, vol. 13, no. 1, p. 626, Jan. 2023.

[47] H. Zhang, Y. Ding, L. Cao, X. Wang, and L. Feng, "Fine-grained question-level deception detection via graph-based learning and cross-modal fusion," *IEEE Trans. Inf. Forensics Security*, vol. 17, pp. 2452–2467, 2022.

[48] H. Tao, H. Yu, M. Liu, H. Fu, C. Zhu, and Y. Xie, "A semi-supervised high-quality pseudo labels algorithm based on multi-constraint optimization for speech deception detection," *Comput. Speech Lang.*, vol. 85, Apr. 2024, Art. no. 101586.

[49] A. Gallardo-Antolín and J. M. Montero, "Detecting deception from gaze and speech using a multimodal attention LSTM-based framework," *Appl. Sci.*, vol. 11, no. 14, p. 6393, Jul. 2021.

[50] M. M. Rahman, A. Shome, S. Chellappan, and A. B. M. A. A. Islam, "How smart your smartphone is in lie detection?" in *Proc. 16th EAI Int. Conf. Mobile Ubiquitous Syst., Comput., Netw. Services*, Nov. 2019, pp. 338–347.

[51] H. Tao, P. Lei, M. Wang, J. Wang, and H. Fu, "Speech deception detection algorithm based on SVM and acoustic features," in *Proc. IEEE 7th Int. Conf. Comput. Sci. Netw. Technol. (ICCSNT)*, Oct. 2019, pp. 31–33.

[52] D. Avola, L. Cinque, G. L. Foresti, and D. Pannone, "Automatic deception detection in RGB videos using facial action units," in *Proc. 13th Int. Conf. Distrib. Smart Cameras*, Sep. 2019, pp. 1–6.

[53] M. Monaro, S. Maldera, C. Scarpazza, G. Sartori, and N. Navarin, "Detecting deception through facial expressions in a dataset of videotaped interviews: A comparison between human judges and machine learning models," *Comput. Hum. Behav.*, vol. 127, Feb. 2022, Art. no. 107063.

[54] D. Pasquali, J. Gonzalez-Billandon, A. M. Aroyo, G. Sandini, A. Sciutti, and F. Rea, "Detecting lies is a child (robot)'s play: Gaze-based lie detection in HRI," *Int. J. Social Robot.*, vol. 15, no. 4, pp. 583–598, Apr. 2023.

[55] S. Chebbi and S. B. Jebara, "Deception detection using multimodal fusion approaches," *Multimedia Tools Appl.*, vol. 82, no. 9, pp. 13073–13102, Apr. 2023.

[56] X. Chen, S. Ita Levitan, M. Levine, M. Mandic, and J. Hirschberg, "Acoustic-prosodic and lexical cues to deception and trust: Deciphering how people detect lies," *Trans. Assoc. Comput. Linguistics*, vol. 8, pp. 199–214, Dec. 2020.

[57] A. Derakhshan, M. Mikaeili, T. Gedeon, and A. M. Nasrabadi, "Identifying the optimal features in multimodal deception detection," *Multimodal Technol. Interact.*, vol. 4, no. 2, p. 25, Jun. 2020.

[58] V. Pérez-Rosas, M. Abouelenien, R. Mihalcea, and M. Burzo, "Deception detection using real-life trial data," in *Proc. ACM Int. Conf. Multimodal Interact.*, Nov. 2015, pp. 59–66.

[59] C. Bai, M. Bolonkin, J. Burgoon, C. Chen, N. Dunbar, B. Singh, V. S. Subrahmanian, and Z. Wu, "Automatic long-term deception detection in group interaction videos," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2019, pp. 1600–1605.

[60] S. Venkatesh, R. Ramachandra, and P. Bours, "Robust algorithm for multimodal deception detection," in *Proc. IEEE Conf. Multimedia Inf. Process. Retr. (MIPR)*, Mar. 2019, pp. 534–537.

[61] V. Gupta, M. Agarwal, M. Arora, T. Chakraborty, R. Singh, and M. Vatsa, "Bag-of-lies: A multimodal dataset for deception detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 83–90.

[62] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, "Finding deceptive opinion spam by any stretch of the imagination," 2011, *arXiv:1107.4557*.

[63] A. Wawer and J. Sarzyńska-Wawer, "Detecting deceptive utterances using deep pre-trained neural networks," *Appl. Sci.*, vol. 12, no. 12, p. 5878, Jun. 2022.

[64] B. Kleinberg and B. Verschuere, "How humans impair automated deception detection performance," *Acta Psycholog.*, vol. 213, Feb. 2021, Art. no. 103250.

[65] M. Korobov and K. Lopuhin, "Welcome to ELI5's documentation!" 2024. Accessed: Jun. 20, 2024. [Online]. Available: https://eli5.readthedocs.io/en/latest/

SUHAIB SALAH received the bachelor's and master's degrees (Hons.) from the University of Sharjah, United Arab Emirates. He is currently pursuing the Ph.D. degree in computer engineering with Khalifa University. His master's thesis was on the application of explainable artificial intelligence in high-stakes deception detection. During the master's studies, he was a Teaching Assistant in his department.

HAGAR ELBATANOUNY received the Bachelor of Science degree in computer engineering from the University of Sharjah, United Arab Emirates, in 2022, where she is currently pursuing the Master of Science degree in biomedical engineering. Since 2022, she has been with the Department of Electrical Engineering, University of Sharjah, as a Teaching Assistant. She has experience with a range of projects, such as non-invasive blood glucose prediction methods, handwriting recognition, freezing of gait prediction and detection, and deception detection. Her areas of interests include machine learning applications, wearable technology, and signal processing.

ABRAR SOBUH received the bachelor's and master's degrees (Hons.) in electrical and electronics engineering from the University of Sharjah, in 2020 and 2023, respectively. Her research focuses on reliable modeling, simulation, identification, control, and state estimation for dynamic systems with uncertainty. This includes developing innovative methods to improve the reliability and accuracy of these systems, which are crucial in various engineering applications.

EQAB ALMAJALI (Member, IEEE) received the B.Sc. degree (Hons.) from Mu'tah University, Jordan, and the M.A.Sc. and Ph.D. degrees (Hons.) in electrical engineering from the University of Ottawa, Ottawa, ON, Canada, in 2010 and 2014, respectively. He has been an Assistant Professor with the Electrical Engineering Department, University of Sharjah, since August 2017. Before that, he was a Postdoctoral Fellow with the Electronics Department, Carleton University, Canada. He is the author of more than 50 technical publications and two book chapters. His current research interests include frequency selective surfaces, millimeterwave MIMO antennas, reconfigurable antennas, RF passive and active sensors, THz antennas, and wireless power transfer. He received the prestigious Canadian National Science and Engineering Research Council (NSERC) Postdoctoral Fellowship for his research excellence, in 2014; and the NSERC-PGS Scholarship during his Ph.D. studies, in 2012.

WASIQ KHAN (Senior Member, IEEE) received the B.Sc. degree in mathematics, physics, and geography and the M.Sc. degree in computer science from Pakistan, and the M.Sc. degree in artificial intelligence for board games and the Ph.D. degree in speech analysis and intelligent reasoning from Bradford University, U.K. He received a Postgraduate Certificate in teaching and learning in higher education (PGCHEP). He is currently a Senior Academician in artificial intelligence and data sciences with the Department of Computer Science, Liverpool John Moores University, U.K. He is also a Visiting Professor in artificial intelligence with the University of Anbar, Iraq. He has been publishing the research outcomes in high impact journals, peer reviewed conferences, news blogs and media, scientific festivals, and public events. He is an active reviewer for top ranked journals (e.g., IEEE TRANSACTIONS) and government funding bodies.

HAYA ALASKAR received the M.Sc. degree in applied artificial intelligence from the University of Exeter, in 2009, and the Ph.D. degree in computer science from Liverpool John Moores University, in 2014. She is currently an Assistant Professor with the College of Computer Science and Engineering, Prince Sattam Bin Abdulaziz University, Saudi Arabia. She has several publications concentrated on using machine learning in various medical data, such as signals and images. Her research interests include artificial intelligence applications and data science.

ADEL BINBUSAYYIS (Member, IEEE) is currently an Assistant Professor with the College of Engineering and Computer Science, Prince Sattam Bin Abdulaziz University, where he is a specialist in cybersecurity and technology transfer. He is also the Vice-Dean of e-learning with the Deanship of Information Technology and Distance Learning, Prince Sattam Bin Abdulaziz University. He is also an Advisor of Vice Rector with Prince Sattam Bin Abdulaziz University, where he is responsible for monitoring the performance executions of the university strategic goals.

TAIMUR HASSAN (Senior Member, IEEE) received the B.S. degree in computer engineering from Bahria University, Islamabad, Pakistan, in 2013, the M.S. degree in computer engineering from the University of Engineering and Technology (UET), Taxila, Pakistan, in 2015, and the Ph.D. degree in computer engineering from the National University of Sciences and Technology (NUST), Islamabad, in 2019. He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, Abu Dhabi University, United Arab Emirates. Prior to that, he was a Postdoctoral Fellow with the Khalifa University Center for Autonomous Robotic Systems (KUCARS) and the Center for Cyber-Physical Systems (C2PS), Department of Electrical Engineering and Computer Science, Khalifa University, Abu Dhabi, United Arab Emirates. He has worked on many local and foreign-funded research projects as a principal investigator, a co-principal investigator, and a lead scientist/engineer. His research interests include robotic vision, medical imaging, deep learning, signal processing, and computer vision. He was a recipient of various national and international awards.

JAWAD YOUSAF (Senior Member, IEEE) received the M.S. and Ph.D. degrees in electronics and electrical engineering from Sungkyunkwan University, Suwon, South Korea, in 2016 and 2019, respectively. He is currently an Associate Professor with the Electrical and Computer Engineering Department, Abu Dhabi University, United Arab Emirates. He was a Brain of Korea (BK)-Postdoctoral Fellow with the EMC Laboratory, Sungkyunkwan University, Suwon, South Korea, from March 2019 to July 2019. Also, he was a Senior RF Researcher with Pakistan Space and Upper Atmosphere Research Commission (SUPARCO: National Space Agency of Pakistan), from 2009 to 2013. His research work has resulted in over 120 publications in leading peer-reviewed international technical journals and refereed international and national conferences.

ABIR HUSSAIN (Senior Member, IEEE) received the Ph.D. degree from The University of Manchester (UMIST), U.K., in 2000. Her Ph.D. thesis titled "Polynomial Neural Networks for Image and Signal Processing." She is a Visiting Professor of machine learning with Liverpool John Moores University, U.K. She has published several refereed research papers in conferences and journals in the research areas of neural networks, signal prediction, telecommunication fraud detection, and image compression. She has worked with higher-order and recurrent neural networks and their applications to e-health and medical image compression techniques. She has developed with her research students several recurrent neural network architectures. She is a Ph.D. supervisor and an external examiner for research degrees, including Ph.D. and M.Phil. She is one of the initiators and chairs of the Development in e-Systems Engineering (DeSE) series.