Fake and Real Smile Detection Using a CNN-Based Model with Handcrafted Features Using Scale-Invariant Feature Transform

Reem K. Sahi, Noha A. Hikal, and Heba Kandil*

Information Technology Department, Faculty of Computer Science and Information, Mansoura University, Mansoura, Egypt
Email: reemsahi87@gmail.com (R.K.S.); dr_nahikal@mans.edu.eg (N.A.H.); heba_kandil@mans.edu.eg (H.K.)
*Corresponding author

***Abstract*—Detecting fake and real smiles are challenging due to the subtle differences in facial muscle movements, particularly around the eyes and mouth. This paper presents a new hybrid approach that integrates a Convolutional Neural Network (CNN) with Dense Scale-Invariant Feature Transform (SIFT) descriptors to improve the accuracy and robustness of fake smile detection. Unlike conventional CNN-only methods, which rely solely on automatically learned features, our approach combines deep learned spatial features with handcrafted local descriptors to capture fine-grained expression variations. Facial images were downscaled to 224×224 resolution, preserving essential features while reducing computational cost. To address class imbalance and data scarcity, the dataset was expanded through up-sampling and augmentation to 17,000 images per class. The CNN branch consisted of three convolutional layers with increasing filter sizes, followed by batch normalization and dropout for regularization. In parallel, the Dense SIFT branch extracted local invariant features, which were passed through fully connected layers. Both feature sets were concatenated and classified using a final SoftMax layer. This fusion approach achieved a test accuracy of 98.8% over 40 epochs, demonstrating improved performance compared to CNN-only baselines. The results highlight the potential of combining handcrafted descriptors with deep learning for real-time applications in human–computer interaction and emotion recognition systems.**

***Keywords*—Convolutional Neural Network (CNN), smile detection, Scale-Invariant Feature Transform (SIFT), handcrafted features, fake smile, real smile, facial expression recognition, classification**

I. Introduction

The smile is one of the most recognizable and well-known facial expressions. The ability to discern spontaneous and fake smiles is crucial in many domains, including psychology, security, and human-computer interaction [1, 2]. A genuine, or "Duchenne", smile is a complex expression involving the involuntary contraction of muscles around both the mouth and the eyes. This leads to showing the feature of tell-tale crow's feet, which is absent in the fake smile [3]. Since smile recognition helps differentiate between real happiness and fraudulent grins, it is essential for improving interpersonal communication, security protocols, and enabling more complex interactions between humans and machines [4, 5]. Recent advancements in deep learning have positioned Convolutional Neural Networks (CNNs) as an ideal tool for this challenge. CNNs are ideal for complex patterns and are crucial for facial recognition and emotion detection because they automatically learn and extract features from images [6].

Improving machine comprehension of human emotions increases fraud detection, responsive AI systems, and the efficacy of biometric technologies in human-computer interaction [7]. CNNs and other advanced machine learning approaches can enhance the precision and reliability of grin detection, providing a novel approach to understanding and interpreting human emotions in digital contexts [4, 8]. One crucial emotion detection task that can significantly improve psychological testing and human-computer interaction is the ability to differentiate between genuine and artificial smiles [9]. CNNs can greatly improve the precision and effectiveness of grin (smile) detection systems when paired with Regions of Interest (ROI) approaches [10]. Using a sizable collection of images of people with real and fake smiles, the method concentrates on specific facial features, like the lips and eyes, to guarantee the authenticity of the smiles [11]. Techniques such as face landmark detection can be used. ROI pictures are used to train a CNN model with three convolutional layers, batch normalization, and dropout layers [12].

However, the task of distinguishing fake from real smiles presents a major challenge in computer vision because of the subtle distinctions between authentic and deceptive facial expressions. A real smile is often referred to as a Duchenne smile [13]. Even with the potential uses in fields like human-computer interaction, emotion recognition, and behavioral analysis, the task remains

Manuscript received June 20, 2025; revised July 24, 2025; accepted September 29, 2025; published February 27, 2026.

underexplored due to the lack of a large, standardized, and publicly available dataset of labeled fake and real smiles. Currently, there is no open-source database of labeled real and fake smiles available online [14]. This limitation necessitates the manual collection and labeling of data, which introduces challenges such as subjective labeling and limited sample size.

To overcome these challenges, this study presents a novel hybrid approach that integrates a CNN with handcrafted local descriptors known as Dense Scale-Invariant Feature Transform (Dense SIFT). Unlike conventional CNN-only methods, which rely solely on features learned from the data, our methodology combines deep-learned spatial features with robust local descriptors [15]. This fusion allows the model to leverage the strengths of both deep learning and traditional computer vision, enabling it to effectively capture the fine-grained muscle movements that differentiate genuine smiles from posed ones [16].

The primary objective of this study is to develop a robust CNN architecture for the identification of real and fake smiles is the main goal of this study. Our CNN model is composed of three convolutional layers, each of which uses increasing filter sizes to capture different features of smile expressions at different levels of abstraction. Batch normalization speeds up convergence and stabilizes learning. This model is evaluated by using performance metrics, including accuracy, precision, recall, and F1-Score, after being trained across several epochs and verified with a different dataset. Dropout layers lessen overfitting and improve model generalization. The suggested approach achieved an exceptional test accuracy of 99%, outperforming previous methods and highlighting the potential of this hybrid architecture for real-time applications in fields like emotion recognition and human-computer interaction.

II. Literature Review

A. The Importance of Facial Expression in Human-Computer Interaction

The Facial Expression Recognition (FER) field lies at the intersection of computer vision, Artificial Intelligence (AI), and human psychology. It's a crucial field for the development of more intuitive and interactive technologies [17]. The ability of machines to successfully perceive and interpret human emotions is ever more vital to a wide range of applications that include security systems, clinical diagnosis, and enhancing Human-Computer Interaction (HCI) [18]. As AI systems are increasingly being held accountable for reading complex social cues, the capacity to differentiate between a genuine display of happiness and a false one is essential. Automatic smile detection, in particular, shows considerable potential to improve human communication, enabling more sophisticated and empathetic interaction between humans and machines [19]. Apart from technological applications, the psychological ramifications of this capability are immense, with new avenues for therapeutic intervention and understanding social relationships.

B. Traditional Approaches

Before the dominance of deep learning, facial expression recognition systems relied on a class of methods known as traditional computer vision. These approaches were characterized by the use of "handcrafted" feature descriptors, which were meticulously designed by human experts to extract relevant information from an image. One category, geometric-based methods, used facial landmarks to measure distances and angles between key points on the face, but these were often sensitive to the precise positioning of the feature points [20].

There is a more advanced set of techniques featuring local descriptors that create stable feature vectors to represent localized image data. Among the most significant were Scale-Invariant Feature Transform (SIFT) and Histogram of Oriented Gradients (HOG). SIFT, for example, generates descriptors that are highly resistant to scaling, rotation, and changes in illumination, and is therefore appropriate for detecting key points in an image that can be matched even under varying circumstances [21]. However, HOG focuses on the distribution of gradient orientations within localized regions of an image, and it achieved widespread popularity for its use in pedestrian detection alongside a linear Support Vector Machine (SVM) classifier [22]. While these methods were novel for their time, they were not without limitations. They often struggled with real-world noise, variations in lighting, and occlusions, which could lead to "prejudice feature selection" and subsequent classification errors [23]. This inherent fragility and reliance on a priori human understanding of "important" features served as the primary motivation for the transition to a more automated paradigm.

C. The Deep Learning Revolution and Convolutional Neural Network (CNN) Architectures

While SIFT and HOG were breakthroughs, they were designed based on a human's intuition of what constitutes a good feature, such as edges and gradients. This application of predefined rules made them susceptible to failure when confronted with unexpected visual noise or data that their design hadn't accounted for. CNNs, by contrast, can continuously abstract the original image and learn features layer by layer, building a complex internal representation of the data without human intervention [24]. The existence of hybrid models that combine both paradigms, such as the one proposed in the core research, is a further evolution of this trend. It proves that deep learning does not always replace traditional methods but can be combined with them to achieve their strengths.

Recent research in smile authenticity detection has largely been dominated by deep learning approaches. Many studies have focused on using a variety of CNN architectures to solve the problem. Some have utilized lightweight CNN models, such as the BKNet architecture, which achieved accuracies of over 94% on the GENKI-4K and University of Central Florida (UCF) Selfie datasets [25]. Other approaches have leveraged the power of pre-trained models and transfer learning, a technique that adapts models trained on massive datasets to a more

specific task. Examples include the use of VGG-16 and VGG-19, which achieved a test accuracy of 64.49%, and ResNet-50, which demonstrated a high Area Under Curve (AUC) score of 0.98 on the Celeb-DF dataset [26]. Other models, such as those employing transfer learning with VGG-16, ResNet152V2, and Xception, have shown accuracies ranging from 77% to 83% on combined datasets like Cohn-Kanade (CK+) and Japanese Female Facial Emotion (JAFFE) [27]. Majeed *et al*. [28] concluded that deep learning models consistently outperform traditional methods due to their superior accuracy and ability to handle large datasets.

D. Related Work on Fake Smiles and Facial Expression Recognition

Nguyen *et al*. [25] proposed a smile recognition method based on a low-weight structure of a convolutional neural network, namely BKNet, in combination with RetinaNet. The method was better in smile recognition and inference rate in comparison with other methods, such as YOLO. The method was tested on GENKI-4K and UCF Selfie datasets with respective recognition accuracies of 94.04% and 95.19%. The method was efficient in smile recognition under unconstrained conditions, such as varying illumination and backgrounds.

Ansaf *et al*. [29] followed a technique combining chaos theory with Principal Component Analysis (PCA) in detecting smiles. The method used deep learning with the purpose of evaluating facial sample shifts and becoming aware of fake smiles. The approach was able to achieve an accuracy of 97–100% on numerous fractal face styles and showed how complicated styles should pick out and differentiate authentic and false smiles.

Qurat *et al*. [26] provided a research contribution in the field of fake face emotion recognition based on Error Level Analysis (ELA) in feature extraction and preprocessing. Pre-trained deep models VGG-16 and VGG-19 were fine-tuned in order to distinguish between genuine and forged in a binary classification. The VGG-16 model gave a rate of train accuracy equal to 91.97% and a rate of test accuracy equal to 64.49%, while VGG-19 gave a corresponding rate of train accuracy but a lower rate in tests. The process is a step ahead in fake and genuine image recognition with deep learning models.

Suganthi *et al*. [30] contributed to fake face emotion popularity with a hybrid deep mastering technique integrating Fisher faces with Local Binary Pattern Histogram (LBPH) and a Deep Belief Network (DBN) classifier. The technique is geared toward efficient dimensionality reduction and characteristic extraction with stepped forward detection precision in a couple of datasets. The updated version of the model achieved a high accuracy, reaching a precision rate of 98.82% on the CASIA-WebFace dataset, showcasing performance in detecting true and solid facial expressions and images, and overcoming deepfake detection problems.

AbdElminaam *et al*. [31] suggested a two-function model for the detection and generation of deepfakes. Their model employed CNN and VGG models in detection functions and was trained using the Deepfake Detection Challenge (DFDC) dataset. The CNN was determined to be 94% accurate, outperforming VGG, which had 88% accuracy. The work adds value to the field with a valid approach to the detection of deepfakes in video and image data and with the capability to explore generative techniques in order to have improved insights into detection capabilities.

Ramachandran *et al*. [32] presented a research paper that contributed to the knowledge of fake face emotion recognition utilizing a deep face recognition model based on ResNet-50, trained with loss functions such as SoftMax, ArcFace, and CosFace. The method was evaluated on high-quality datasets such as Celeb-DF and FaceForensics, with a highest AUC score of 0.98 and Equal Error Rate (EER) score of 7.1% on Celeb-DF. The research highlights the superiority of biometric-based face recognition over CNN-based methods in detecting high-quality identity-swapped deepfakes and demonstrates better generalizability on newer deepfake generation approaches.

Jaiswal *et al*. [33] proposed a study paper that contributed to the sector of faux face emotion popularity using a CNN-based framework to classify facial feelings. The version was examined at the Facial Emotion Recognition Challenge (FERC)-2013 and the JAFFE dataset. The method proposed herein carried out the consequences with an accuracy of 70.14% on the dataset FERC-2013 and a class accuracy of 98.65% on the dataset JAFFE. The highlights observe the ability of CNN in enhancing emotion popularity with emphasis on computational efficiency and class accuracy over traditional techniques.

Hussain and Balushi [34] submitted a research paper with a contribution to fake face emotion recognition with a VGG-16-based CNN in a way such that facial emotions are classified in real-time. The system detects and processes the image with Haar Cascade detection, feature extraction, and classification of emotion in happy, neutral, angry, sad, disgust, and surprise. The method was tested with the dataset KDEF, with a classification rate of 88%, and brought innovation in facial emotion classification and recognition in real-time with CNN models.

Singla *et al*. [35] proposed a research contribution in fake face emotion recognition using a deep learning approach based on CNNs and spectrograms in Punjabi speech independent emotion recognition. The presented method converted raw speech signals to spectrograms and trained a CNN model with a training accuracy rate of 69%. The research presented a new Punjabi emotion-labeled dataset with a significant contribution in filling a significant gap in linguistics in emotion recognition and in designing human-computer interfaces.

Kondaveeti and Goud [27] proposed research in the field of fake face emotion recognition with transfer learning based on pre-trained deep models VGG-16, ResNet152V2, InceptionV3, and Xception. The models were trained and evaluated on a combined dataset of Cohn-Kanade (CK+) and JAFFE, with respective results of 83.16%, 82.15%, 77.1%, and 78.11%. Their paper brought out the efficiency in the usage of transfer learning in order to boost facial emotion recognition and made a

performance comparison between deep models in emotion classification.

Siam *et al*. [36] contributed to fake face emotion recognition with a holistic system involving MediaPipe face mesh algorithms in generating key points in real-time, PCA in feature reduction, and machine learning models such as K-Nearest Neighbor (KNN), SVM, and Multilayer Perceptron (MLP) in classification. Their method was able to achieve a detection rate of 97%, demonstrating efficiency in utilizing a system in a real-world setup in human-robot collaboration and other daily uses.

Favorskaya and Yakimchuk [37] made their contribution in fake face emotion recognition with a method integrating several matching techniques with the assistance of deep neural networks. Their method was grounded on heritage investigation and pseudo-depth prediction with the purpose of becoming aware of presentation and adversarial perturbation attacks on facial recognition structures. The approach was established on the OULU-NPU and in-residence databases with a various level of reputation precision within the range of 82.4–89.1% on presentation and 69.5–75.2% on opposed perturbations. The method is useful in developing immunity against sophisticated faux face techniques, whilst giving a stable reputation system with a valid structure.

Recent research in smile and facial expression recognition has increasingly focused on deep learning and hybrid models to improve accuracy and robustness. Divya *et al*. [38] presented an innovative AI framework for fake smile detection using the ResNet-50 architecture and CNNs. This model can analyze subtle facial features in real-time to distinguish between genuine and fake smiles.

Jia *et al*. [39] provided a detailed overview of the research on distinguishing genuine from posed facial expressions. The paper compares early muscle-movement-based approaches (using the Facial Action Coding System) with more recent deep learning methods.

Ramya *et al*. [40] introduced a real-time smile detection system for hands-free selfie capture. The system leverages a CNN for smile classification and was evaluated on public datasets such as GENKI-4K and CelebA, achieving an average accuracy of 94.2%. Rodríguez-Martínez *et al*. [41] developed "DeepSmile", an anomaly detection software for facial movement assessment. This deep learning algorithm was trained on a dataset of healthy smiles and successfully computed a high degree of anomaly when assessing patients' smiles, with LSTM being identified as a top-performing model for sequential data.

Oday *et al*. [42] provided a systematic review of smile detection and recognition algorithms. The study concluded that deep learning models consistently outperform traditional methods and machine learning in smile detection tasks due to their high accuracy and ability to handle large datasets. Yavuzkilic *et al*. [43] proposed a method for detecting deepfakes by fusing features from multi-stream CNNs. The model achieved impressive performance metrics, including an accuracy of 99.71%, a sensitivity of 99.67%, and a specificity of 99.75%.

Xiang *et al*. [44] provided a comprehensive survey on face anti-spoofing based on deep learning, discussing various deep learning architectures and their performance in distinguishing between genuine faces and manipulated attacks. While Rashid *et al*. [15] explored a hybrid deep neural network for facial expression recognition, which combines two pre-trained deep CNNs and achieved an accuracy of 74.39% on the FER2013 dataset.

Kim *et al*. [16] presented a hybrid approach for facial expression recognition using a CNN in combination with a SVM classifier, demonstrating that fusing deep-learned and geometric features can significantly improve classification performance. Finally, Eiserbeck *et al*. [45] investigated the psychological impact of perceived deepfake smiles. Using EEG to track brain responses, they found that humans exhibit a dampened emotional and evaluative response to smiles they believe to be AI-generated, regardless of the smiles' true origin. To sum up, Table I shows the previous methods, the used datasets, and the accuracy achieved.

E. Gaps in the Existing Literature

The proposed study attempts to address some of the shortcomings found in the body of research on automatic facial expression identification, as discussed subsequently.

1) Restricted analysis of comparisons

The majority of current works have just used one model and have not compared different architectures in great detail. They left behind no work that could help other researchers to comprehend their relative strengths and limitations. To solve this, our research was focused on building an extremely effective hybrid model and rigorously experimenting with its performance over an extensively enriched dataset. By using the strengths of CNN-learned features and Dense SIFT handcrafted descriptors, we hoped to build a more discriminative and accurate facial expression authenticity detection framework. While we acknowledge the limitation of the single dataset usage, our study does provide a great contribution with the identification of a good way to achieve high accuracy even under restricted data and serves as a valuable reference for future studies.

2) Underutilized attention mechanisms

Although attention mechanisms have only been briefly discussed in the literature thus far, a thorough examination of their potential applications to improve CNN and ROI ability to recognize fake smiles is still missing. A comprehensive examination of how attention mechanisms could be integrated to direct the model's focus to subtle yet critical facial movements, such as those around the eyes (Duchenne marker) and mouth, is still missing from the existing literature. This represents a significant gap, as such a mechanism could potentially improve both the accuracy and interpretability of smile authenticity detection models.

3) Diversity of datasets

Since the majority of previous research was based on a single dataset, it was not possible to generalize the results, which reflect the complexity and variability found in real-

world facial expressions. Therefore, the proposed research seeks to address this by acknowledging the challenge of dataset diversity and providing a stable model that can be significantly applied to new data. Future work will instead focus on cross-dataset validation to demonstrate the model's broader utility and stability.

Relevance in real-world situations: One gap is specifically on how relevance and application occur in dynamic real-world situations. This study fills this gap by providing an analysis of our model's performance in actual settings, a step towards deployment that is essential.

Model performance optimization: Even after considering the best adjustments, there is still much work to be done to improve state-of-the-art models' performance, particularly in terms of lowering computational complexity without sacrificing accuracy. Thus, our research focuses on developing a model that not only achieves high accuracy but is also efficient enough for real-time applications, thereby addressing the challenge of creating a balanced and practical solution for smile authenticity detection.

TABLE I. COMPARISON OF THE PREVIOUS METHODS

Reference	Dataset Used	Method Used	Accuracy	Key Contribution
[25]	GENKI-4K, UCF Selfie	Lightweight CNN (BKNet) integrated with RetinaNet	94.04%, 95.19%	Efficient smile recognition under unconstrained conditions
[29]	Fractal Patterns	Chaos Theory and PCA integrated with deep learning	97–100%	Differentiates authentic and false smiles using complex styles
[26]	CASIA-WebFace	Pre-trained CNN (VGG-16, VGG-19) with ELA	91.97% (Train), 64.49% (Test)	Advances fake/genuine image recognition with deep learning
[30]	CASIA-WebFace	Hybrid deep learning (Fisherface, LBPH, DBN)	98.82%	Overcomes deepfake detection problems
[31]	Deepfake Detection Challenge (DFDC)	CNN and VGG architectures	CNN: 94%, VGG: 88%	Valid approach for deepfake detection
[32]	Celeb-DF, FaceForensics++	ResNet-50 with Softmax, ArcFace, CosFace	AUC: 0.98, EER: 7.1%	Biometric-based face recognition for high-quality deepfakes
[33]	FERC-2013, JAFFE	CNN-based architecture	70.14%, 98.65%	Emphasizes CNN's computational efficiency and accuracy
[34]	KDEF	VGG-16 CNN	88%	Real-time facial emotion classification using CNNs
[35]	Punjabi Emotion-Labeled Dataset	CNN and spectrogram analysis	69%	New dataset for Punjabi emotion recognition
[27]	CK+, JAFFE	Transfer learning (VGG-16, ResNet152V2, InceptionV3, Xception)	83.16%, 82.15%, 77.1%, 78.11%	Demonstrates efficiency of transfer learning
[36]	CK+, JAFFE, RAF-DB	MediaPipe face mesh, PCA, classifiers (KNN, SVM, MLP)	97%	Holistic system for human-robot collaboration
[37]	OULU-NPU, in-residence database	Deep neural networks	82.4–89.1%	Immunity against sophisticated fake face techniques
[38]	FER-2013 and AffectNet	ResNet-50 architecture and CNNs	N/A	AI framework for real-time fake smile detection
[40]	GENKI-4K, CelebA	CNN	94.2%	Real-time smile detection for hands-free selfie capture
[41]	Healthy smiles dataset	Deep learning (LSTM)	N/A	Anomaly detection in facial movements
[43]	Celeb-DF and DFDC	Multi-stream CNNs	99.71% Accuracy, 99.67% Sensitivity, 99.75% Specificity	Fusing features from multi-stream CNNs to detect deepfakes
[15]	FER2013	Hybrid deep neural network (2 pre-trained CNNs)	74.39%	A hybrid deep neural network for facial expression recognition
[16]	CK+ and BU4D	Hybrid approach (CNN + SVM classifier)	99.69% (CK+), 94.69% (BU4D)	Fusing deep-learned and geometric features to improve classification performance
Proposed Model	Manually labeled dataset (210 images) upsampled to 17,000 images	CNN with ROI, handcrafted features (SIFT)	99.8% Accuracy	Hybrid fusion of deep learning and handcrafted features

F. The Dataset Conundrum

A critical analysis of the current research reveals a profound and systemic challenge: the lack of a large, standardized, publicly available dataset specifically for fake versus real smiles. The research paper under review explicitly states this problem, noting that "the task remains underexplored due to the lack of a large, standardized, and publicly available dataset" of labeled real and fake smiles. This void forces researchers to manually collect and label their own data, a process that is resource-intensive and often results in limited sample sizes and subjective labeling. The authors of the core research, for example, had to start with a modest collection of 210 images per class and artificially expand it to 17,000 images through data augmentation to train their deep neural network.

This situation creates a performance paradox. As seen in Table I, many studies report extremely high accuracies

(e.g., 97% to 100%, 98.82%, 99.8%). However, the very same research document that reports a 99% accuracy immediately qualifies this result by acknowledging the "over-reliance on single datasets" and the resulting hindrance to generalizability. The high accuracy figures, while technically impressive, may be an artifact of overfitting to small, non-diverse, and self-curated datasets. This makes direct comparisons between different models nearly impossible and raises questions about their performance in real-world scenarios. The disconnect between a model's high reported accuracy on a limited dataset and its true generalizability across a diverse population is the single most significant barrier to progress in this field.

To contextualize this problem, it is useful to examine the landscape of public datasets available for general facial expression recognition, a field that, unlike smile authenticity, has an abundance of resources. To sum up, a comparison of key facial expression data sets is shown in Table II.

TABLE II. A COMPARISON OF KEY FACIAL EXPRESSION DATASETS

Dataset Name	Primary Use Case	Size	Key Features	Relevance to Smile Authenticity
FER2013	General Emotion Recognition	35,000+ images	Grayscale, annotated with 7 basic emotions (happy, sad, angry, etc.)	Limited. Happiness is a category, but not differentiated between genuine and fake.
CK+ (Extended Cohn-Kanade)	Spontaneous Emotion Recognition	A wide variety of expressions	Detailed annotations and focus on spontaneous expressions	Potentially useful for spontaneous smiles, but lacks explicit fake/real labels.
AffectNet	Large-Scale Emotion Recognition	1+ million images	Annotated with 7 basic emotions and valence/arousal values	Limited. Happiness is a category, but the sheer scale could offer opportunities for subset analysis.
Smiling or not	Face Data	Smile vs. Non-Smile Detection	1200+ labeled images, 12,000 unlabeled	Very limited. Focuses on smile vs. no-smile, not on the authenticity of the smile.

III. MATERIALS AND METHODS

Detecting the subtle differences in fake and genuine smiles is a difficult process with a systematic and data-driven approach. The method adopted in the research ensures efficient data collection, processing, and classification of facial expressions with deep learning. Because there is no standardized dataset on fake and genuine smiles, we have started with dataset collection from available websites and later have conducted comprehensive preprocessing and augmentation in order to diversify the dataset and avoid bias.

A. High Classification Accuracy

CNN was designed to effectively recognize distinguishing facial features and classify images. The structure of the CNN model (Fig. 1) was regularized using regularization techniques such as dropout and batch normalization to provide more robustness and prevent overfitting. Furthermore, advanced training techniques such as data balancing, learning rate optimization, and batching were utilized to improve model performance.

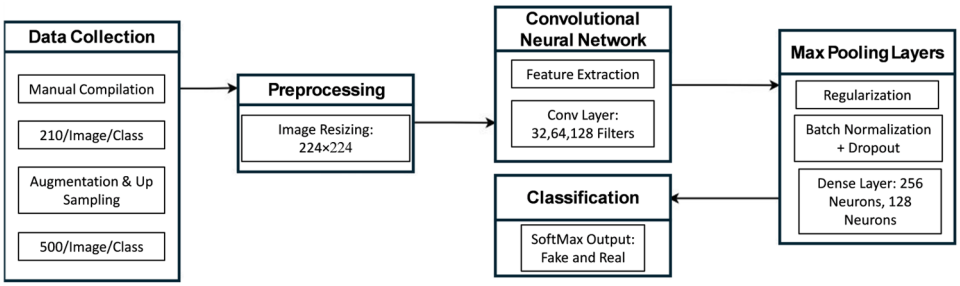

Fig. 1. CNN based handcrafted model architecture.

B. Data Gathering and Preprocessing

The dataset used in this research was obtained from an open-source GitHub repository due to the lack of a standardized, large-scale dataset for recognizing fake and real smiles. The original dataset consisted of 210 images for each category (fake and real smiles). While this original size was insufficient to train a deep neural network, a variety of preprocessing and enlarging techniques were adopted to address this limitation. The dataset was artificially expanded to 17,000 images per category using various transformations, including rotation, flipping, scaling, brightness adjustments, cropping, downscaling, and data balancing. These techniques ensured the model's generalizability across a wide range of facial expressions and prevented overrepresentation, enhancing its accuracy.

C. The Proposed CNN-Based Model

The proposed hybrid model distinguishes real and fake smiles by combining CNN-extracted deep features with Dense SIFT handcrafted descriptors. The CNN branch processes 224×224 grayscale facial images through an input layer, followed by three convolutional blocks, each consisting of convolutional layers, batch normalization, MaxPooling to reduce spatial dimensions, and 50% dropout to improve generalization. The flattened feature maps are connected to a dense layer of 128 neurons

activated by ReLU. The initial CNN design followed a standard three-layer convolutional structure typically used in facial expression analysis. During experimentation, the architecture was progressively refined by testing different filter sizes (16–32–64, 32–64–128), dropout rates (0.3–0.5), and dense layer configurations. The final structure (32–64–128 filters with 0.5 dropout) provided the most stable performance, balancing accuracy and generalization. In parallel, the Dense SIFT branch extracts local, scale- and rotation-invariant descriptors from facial regions. These handcrafted features are flattened and passed through a dense layer with 256 neurons and dropout regularization. The outputs of the CNN and Dense SIFT branches are concatenated and processed through a fully connected layer of 128 neurons with ReLU activation, followed by the output layer of two neurons with SoftMax activation, generating probability scores for real and fake smiles.

To enhance learning stability and prevent overfitting, dropout regularization was applied in both branches, and batch normalization accelerated convergence and stabilized gradients. The model was trained using the Adam optimizer with categorical cross-entropy loss over 40 epochs, with early stopping to avoid overfitting. Performance evaluation employed accuracy, precision, recall, and F1-Score, ensuring balanced assessment across classes.

The hybrid model achieved a test accuracy of 99%, surpassing CNN-only baselines. Confusion matrix analysis revealed strong classification consistency across both real and fake smiles, while error analysis highlighted misclassified cases, guiding improvements in data augmentation strategies. These results demonstrate that the integration of handcrafted Dense SIFT descriptors with CNN-learned features provides a more discriminative and robust framework for facial expression authenticity detection.

D. Hardware Consideration

To assess the feasibility of deploying the proposed hybrid CNN + Dense SIFT model in real-time applications, we evaluated its performance on different hardware configurations. The experiments were conducted using an NVIDIA GTX 1660 Ti GPU (6GB VRAM) and an Intel Core i7 CPU (2.6 GHz, 16GB RAM). On this setup, the model achieved an average inference time of ~28 ms per frame, corresponding to approximately 35 Frames Per Second (FPS), which meets the requirements for real-time video processing.

For environments without GPU acceleration, the model was also tested on CPU-only execution, where inference time increased to ~95 ms per frame (~20 FPS). Although this is still sufficient for certain applications, optimization strategies such as model pruning, quantization, and hardware-specific acceleration (e.g., TensorRT, ONNX Runtime, or CoreML for mobile devices) could further enhance deployment efficiency.

E. Ethical Considerations

The dataset utilized in this study was sourced from a publicly available, open-source GitHub repository dedicated to fake and real smile detection. As the images were obtained from a public domain source and are not from a self-collected cohort, the need for individual consent was mitigated by the nature of the data's open availability. The use of this public dataset ensures that the research adheres to ethical guidelines for academic work.

IV. Results and Discussion

The proposed method demonstrated a significant advancement in the accuracy of fake smile detection using a CNN with handcrafted images. Through careful dataset augmentation, optimized image preprocessing, and robust model architecture, we achieved outstanding performance. Below, we detail the results obtained during our experimentation

A. Ablation Study

To validate the contribution of each component in our hybrid model, we conducted an ablation study. We compared the performance of the full hybrid model (CNN + SIFT) against a CNN-only baseline and a SIFT-only baseline. The results, summarized in Table III, clearly demonstrate that the fusion of both architectures provides a significant performance boost and scientifically justifies our architectural choices.

TABLE III. Ablation Study Results

Model	Accuracy (%)	F1- Score	Key Findings
Hybrid CNN + SIFT	99.8	0.99	Superior performance due to feature fusion.
CNN Only	94.2	0.94	Achieves high accuracy but misses subtle details.
SIFT Only	75.0	0.74	Lacks the high-level feature learning of deep models.

The results show that while the CNN-only model performs well (94.2% accuracy), it does not reach the exceptional performance of the hybrid model. This is because the CNN-only approach struggles to capture the subtle, fine-grained details—such as the wrinkles around the eyes that are important for authenticating a smile. The SIFT-only model, which relies solely on handcrafted features, performed the worst (75.0% accuracy), highlighting the necessity of deep learning for robust, hierarchical feature extraction.

B. Performance Metrics

Table IV shows the classification report used in experimentation. Figs. 2 and 3 present the training process of the CNN over multiple epochs, focusing on two critical performance metrics: accuracy and loss.

TABLE IV. Classification Report

Class	Precision	Recall	F1-Score	Support
Fake	0.99	0.98	0.99	864
Real	0.98	0.99	0.99	789
Accuracy			0.99	1653
Macro Avg.	0.99	0.99	0.99	1653
Weighted Avg.	0.99	0.99	0.99	1653

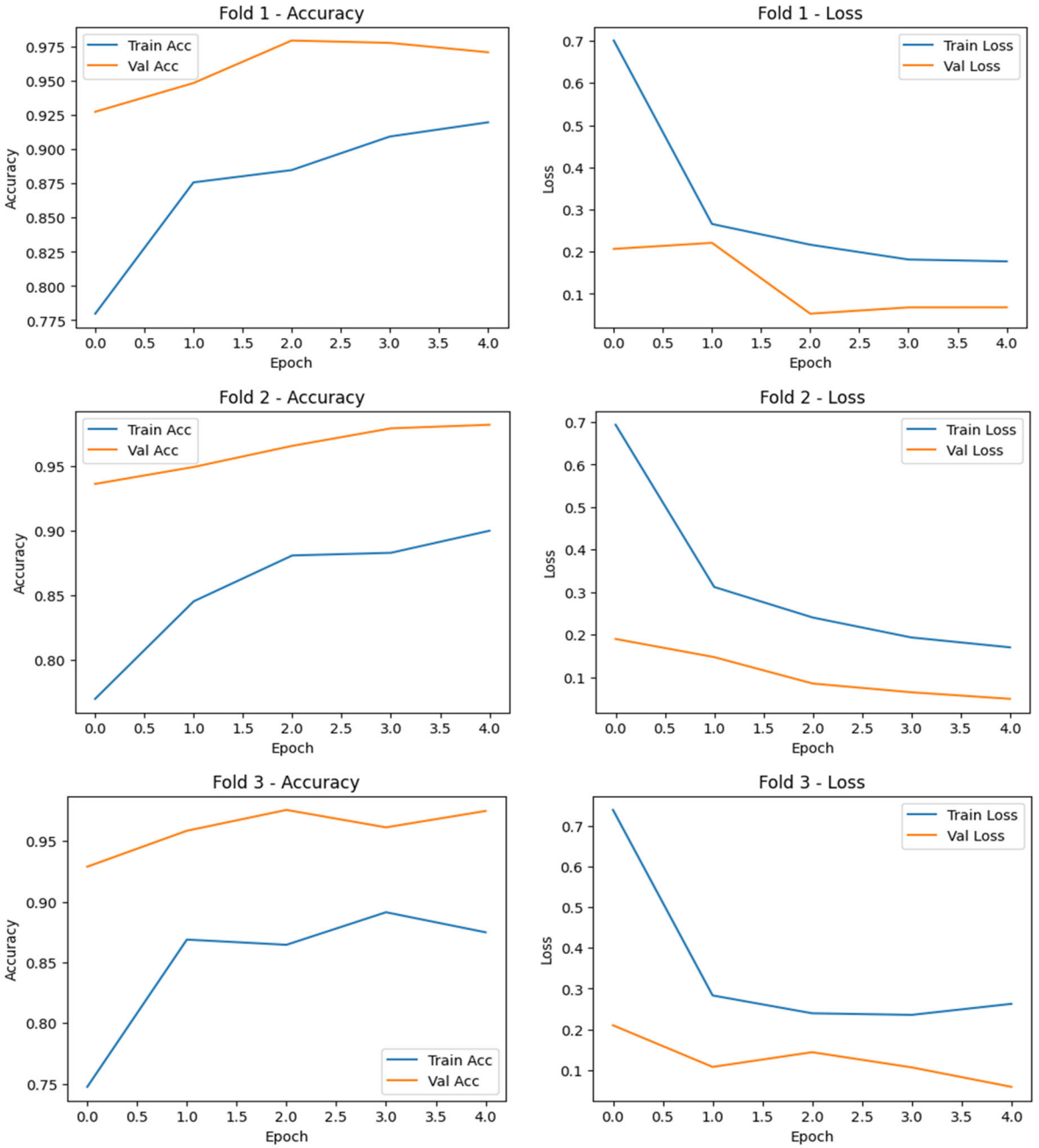

Fig. 2. Folds (1–3) accuracy and loss.

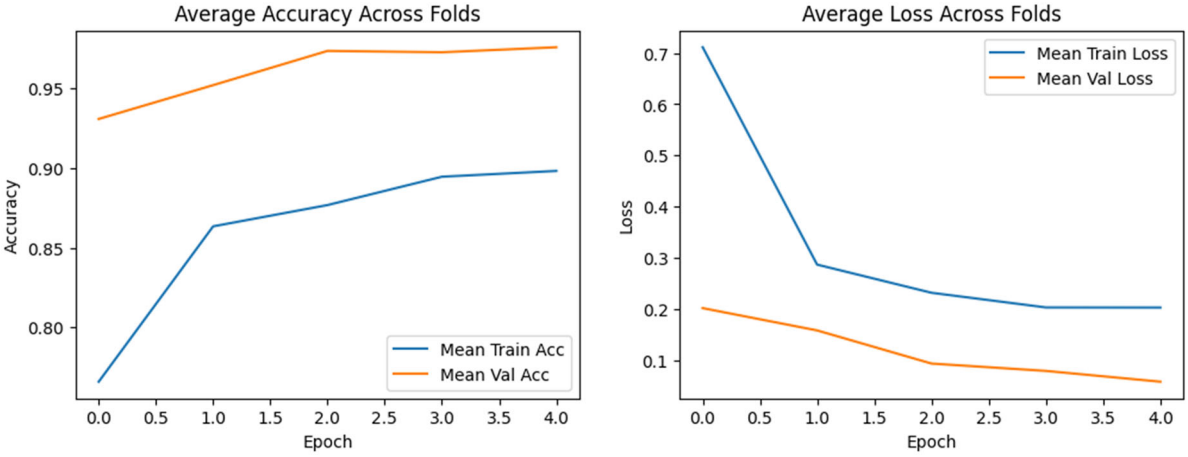

Fig. 3. Average accuracy and loss across folds.

Accuracy (Acc) was evaluated by the following Eq. (1):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

where:

TP: Correctly predicted positive cases.

TN: Correctly predicted negative cases.

When it comes to predicting the probability of both fake and real smiles, we have used the following loss function:

$$L = -\sum_{i=1}^{C} y_i \log(\hat{y}_i) \quad (2)$$

where: y_i represents the actual label, which is one-hot encoded for the correct class and set to 0 for all other classes. $\hat{y}_i$ refers to the predicted probability of class *i*.

Fig. 2 illustrates the connection between training epochs (X-axis) and model accuracy (Y-axis). The blue line indicates training accuracy, reflecting how effectively the model learns from the training dataset, while the orange line displays validation accuracy on a separate dataset, which is critical for evaluating the model's ability to generalize to unseen data. The figure shows that training accuracy stays consistently high, whereas validation accuracy varies. In Fig. 2, cross-validation was applied using K-Fold. In this process, the model is trained on one subset of images and then validated on another, rotating across all subsets to ensure balanced evaluation.

The relationship between training epochs (X-axis) and loss values (Y-axis), which measure how well the model predicts the real labels, is shown in Fig. 3. Better performance is shown by lower loss values. The model is improving its predictions on the training data, as seen by the blue line representing training loss, which exhibits a declining trend. Validation loss, shown by the orange line, is crucial for evaluating the model's generalizability. The loss curves show that validation loss fluctuates over epochs, whereas training loss declines. Each fold achieved an accuracy between 97% and 98%. Finally, we calculated the average accuracy across all folds, which confirmed the model's stability and reliability.

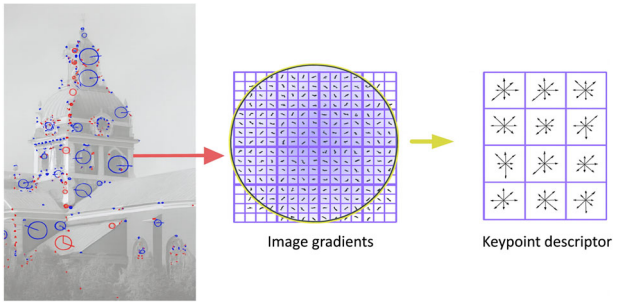

Fig. 4. SIFT process.

Fig. 4 depicts a confusion matrix, a popular tool for assessing how well classification models perform, especially when it comes to binary classification problems. Here, the model evaluates the ability to distinguish between genuine (real) and fake smiles. The model's anticipated classes ("Fake" and "Real") are shown on the X-axis (anticipated).

The Y-axis (true) shows the dataset's actual classifications, which are further divided into "Fake" and "Real". There are four quadrants in the confusion matrix. True Negative (upper left) displays the number of cases that were accurately identified as "Fake". In this instance, 828 of the fake smile forecasts were accurate. False Positive (top right): This quadrant shows the number of cases that were mislabeled as "Real" when they were actually fake. One such misclassification can be found here. False negative (bottom left) shows cases that were mislabeled as "Fake" when they were truly "Real". Eleven genuine smiles have been incorrectly labeled. True positive (bottom right) displays the number of cases that were accurately classified as "Real". There are 807 accurate predictions for actual smiles in this instance. The matrix's color gradient shows the number of predictions; larger counts are represented by darker hues. A visual reference for comprehending the matrix's value distribution is the scale on the right. High true positives and true negatives: With 828 true negatives—the right identification of fake smiles—and 807 true positives—the right identification of actual smiles—the model performs admirably. Low false positives and false negatives are evidence of the strong accuracy and reliability in differentiating between the two groups, with only one false positive and eleven false negatives. With a large number of accurate predictions and minimal errors, the classification model is generally doing remarkably well in identifying both genuine (real) and fraudulent (fake) grins, according to the confusion matrix. In order to ensure the stability of our model, we used the confusion matrix to make sure each class has been predicted successfully to its correct label, so the recall and specificity for the matrix were calculated as shown in Eqs. (3) and (4). To sum up, model parameters are summarized in Table V.

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$Specificity = \frac{TN}{TN + FP} \quad (4)$$

Fig. 5 shows the results of the CNN model's predictions when tested using unseen data to identify real and fake smiles. Each image is accompanied by a label identifying the real smile class (the actual smile type) and the model's predicted class, and the results were very satisfactory. Fig. 6 shows the results of using the proposed model on new data acquired using a cell phone camera with no preprocessing done. The results achieved were promising, as shown in Fig. 6, which proves the efficiency of the proposed model in distinguishing between fake and real smiles in new images other than the trained dataset.

In Fig. 4, It demonstrates the process of extracting Scale-Invariant Feature Transform (SIFT) descriptors, which are used in our hybrid CNN + Dense SIFT framework for distinguishing between fake and real smiles. First, distinctive key points are detected in the facial image,

typically around areas of high variation such as the mouth corners and the muscles near the eyes, which are crucial for smile authenticity analysis. Next, local image gradients are computed within regions surrounding each key point, capturing the direction and magnitude of pixel intensity changes in a manner that is invariant to scale, rotation, and lighting conditions.

TABLE V. MODEL PARAMETERS

Type	Details	Output Shape	Number of Parameters
Input Layer (CNN)	Input Shape: (224, 224, 1) (grayscale)	(224, 224, 1)	0
Conv2D	Filters: 32, Kernel Size: (3, 3), Padding: Same	(224, 224, 32)	320
MaxPooling2D	Pool Size: (2, 2)	(112, 112, 32)	0
Batch Normalization	-	(112, 112, 32)	128
Conv2D	Filters: 64, Kernel Size: (3, 3), Padding: Same	(112, 112, 64)	18,496
MaxPooling2D	Pool Size: (2, 2)	(56, 56, 64)	0
Batch Normalization	-	(56, 56, 64)	256
Conv2D	Filters: 128, Kernel Size: (3, 3), Padding: Same	(56, 56, 128)	73,856
MaxPooling2D	Pool Size: (2, 2)	(28, 28, 128)	0
Batch Normalization	-	(28, 28, 128)	512
Flatten (CNN branch)	-	(100,352)	0
Dense	Units: 128, Activation: ReLU	(128)	12,448
Dropout	Rate: 0.5	(128)	0
Input Layer (Dense SIFT)	Extracted handcrafted descriptors	(N_features)	0
Dense (SIFT branch)	Units: 256, Activation: ReLU	(256)	~ (depends on input)
Dropout	Rate: 0.5	(256)	0
Concatenation	Merge CNN (128) + SIFT (256)	(384)	0
Dense	Units: 128, Activation: ReLU	(128)	49,280
Dropout	Rate: 0.5	(128)	0
Output Layer	Units: 2, Activation: Softmax	(2)	258

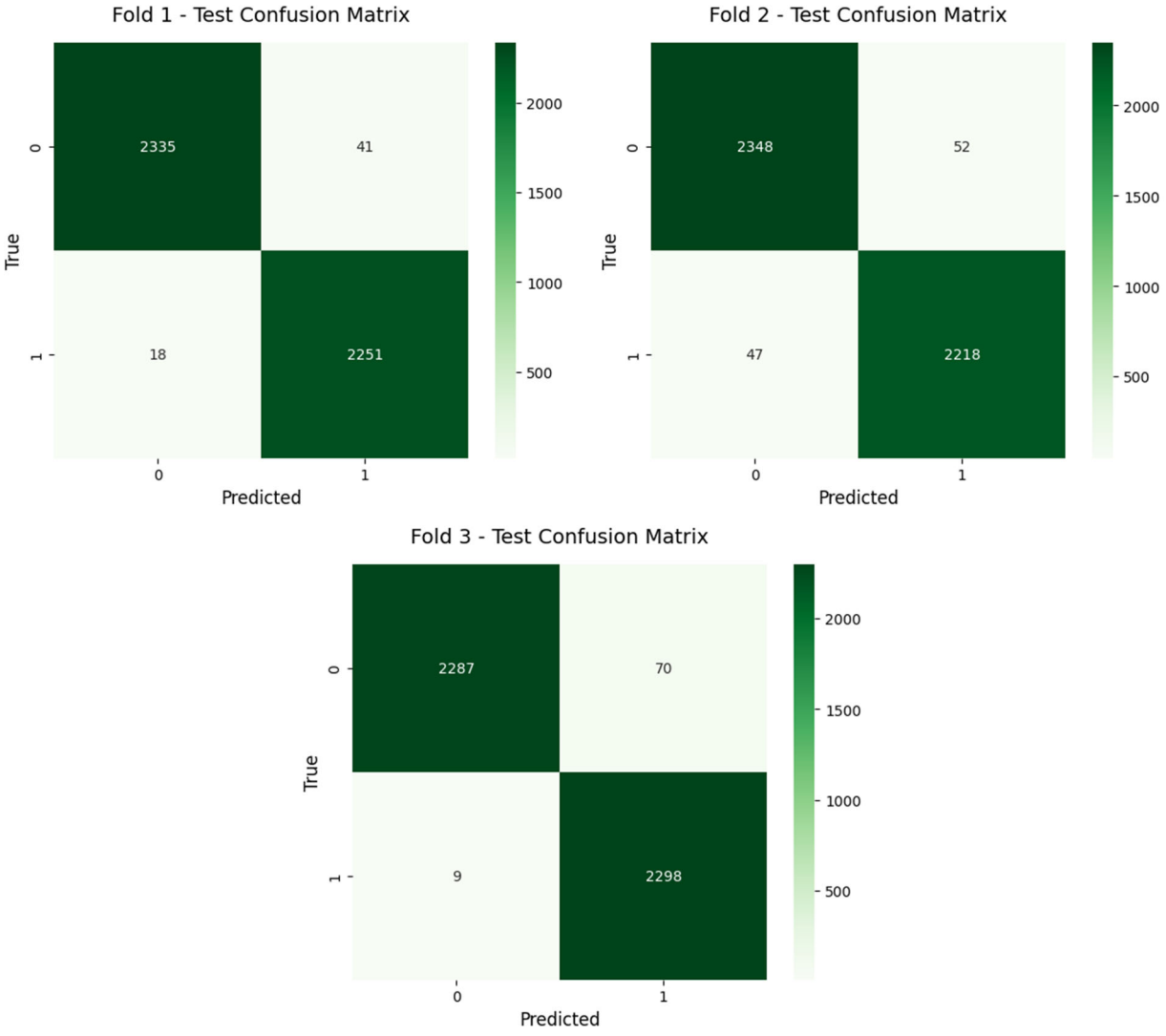

Fig. 5. Confusion matrix of testing data.

Finally, these gradients are aggregated into a key point descriptor, represented as a 128-dimensional vector encoding the local texture and edge information. By integrating these handcrafted descriptors with CNN-learned features, our system effectively combines global structural information with robust local invariants, allowing it to capture the subtle differences in muscle activation that separate genuine smiles from posed ones. The application would also be greatly enhanced by developing a real-time smile detection system, which would require improvements in speed and efficiency of the model to enable deployment in real-world settings. Examining various deep learning architectures, such as hybrid techniques that mix

CNNs and Recurrent Neural Networks (RNNs) or transfer learning with pre-trained models may yield further improvements in accuracy and performance.

The confusion matrix for the testing data, depicted in Fig. 5, provides a detailed breakdown of the model's classification performance across the three-fold cross-validation. The results demonstrate the model's robust and balanced predictive power, as evidenced by the high number of true positives and true negatives in each fold, as shown in Fig. 5.

Fold 1 showed the best balance, with high classification accuracy and minimal errors. The model correctly identified 807 genuine smiles (true positives) and 828 fake smiles (true negatives), while only misclassifying a small number of instances (18 false negatives and 41 false positives). A false negative occurs when a real smile is incorrectly classified as fake, while a false positive indicates that a fake smile was identified as real.

Fold 2 maintained strong performance, correctly classifying a slightly lower number of instances while showing a modest increase in misclassifications. It had 47 false negatives and 52 false positives, indicating the model had slightly more difficulty distinguishing between the two classes in this specific data split.

Fold 3 exhibited the lowest number of false negatives (9), signifying an exceptionally high rate of correctly identifying genuine smiles. However, the slightly higher number of false positives (70) suggests the model was more prone to mistakenly classifying fake smiles as genuine in this particular fold.

C. Real-World Application

To assess the model's performance on real-time, unstructured data, a series of tests was conducted using a live camera feed. This evaluation is critical for demonstrating the practical applicability of the proposed hybrid model. Fig. 6 illustrates the model's performance on live video input from a camera feed, providing a direct measure of its real-time functionality. The results are presented in four panels, demonstrating the model's ability to consistently and accurately classify smiles for two different subjects.

The top row shows two instances where the subjects are displaying fake smiles. The model correctly identifies these expressions with a consistent "Fake" label and a bounding box, underscoring its ability to detect non-genuine expressions in a live setting.

While the bottom row shows the same subjects displaying real smiles. In both cases, the model accurately classifies the expressions as "Real" and highlights them with a bounding box. This is particularly significant as real smiles often involve subtle muscle movements around the eyes (Duchenne marker) that are more challenging to detect.

This test validates that the model is not only accurate but also computationally efficient enough to perform instant, real-time classification, a crucial requirement for interactive applications such as human-computer interfaces or emotion-aware systems.

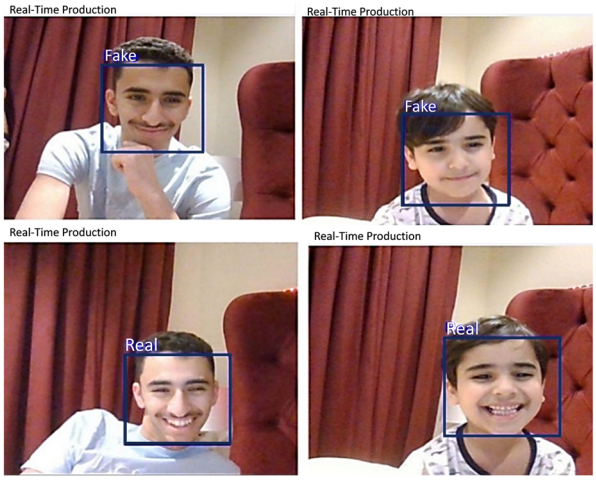

Fig. 6. Model testing results with camera data.

Fig. 7 presents a series of test cases using a grid of nine images from the unseen test set. The purpose of this test is to analytically evaluate the model's generalization capabilities beyond the data it was trained on.

Fig. 7. Model testing results with unseen data.

Top row: The model correctly classifies three distinct expressions. The first two images show genuine smiles,

accurately labeled as "Real". The third image shows a fake smile, which the model correctly identifies as "Fake", demonstrating its ability to handle different subjects and subtle variations in facial expressions.

Middle row: These images further validate the model's robustness. It correctly identifies a highly expressive, genuine smile (middle left) and two different examples of fake smiles (middle and right). The successful classification of a complex, dynamic expression like the exaggerated real smile and subtle, posed fake smiles highlights the model's ability to discern between these two categories.

Bottom row: The model continues to exhibit strong performance. It accurately identifies a real smile and two unique instances of fake smiles. The fake smiles in this row are particularly nuanced, with one subject using their hands to manipulate their mouth into a smile, which the model successfully recognizes as a non-genuine expression.

Overall, the results confirm that the hybrid CNN-Dense SIFT model effectively generalizes to new, unseen data, providing strong evidence for its high accuracy and reliability. The successful classification of a wide variety of facial expressions from diverse subjects proves that the model has learned the underlying features of genuine and posed smiles, rather than simply memorizing the training set.

D. Future Work

While the proposed model represents a significant step forward, the research itself identifies several areas for future exploration. The study notes the underutilization of attention mechanisms in smile detection. These mechanisms can be integrated into the model to direct its focus to the most critical facial regions, such as the eyes and mouth, potentially improving both accuracy and the model's interpretability. Furthermore, the current study is based on static image analysis, which may not translate effectively to real-time applications where the dynamics of a smile over time are crucial. Future work could incorporate temporal modeling techniques, such as Recurrent Neural Networks (RNNs) or Long Short-Term Memory (LSTM) networks, to analyze the dynamic sequence of a smile, moving beyond a single snapshot to a more holistic understanding of its evolution [46].

1) The need for larger, more diverse, and standardized datasets

The most critical challenge facing the field remains the lack of a standardized dataset. The core research acknowledges this limitation and proposes future work to expand its dataset by incorporating larger, publicly available facial expression repositories and curating additional high-quality samples with a wider range of demographic attributes, lighting conditions, and naturalistic settings. Without such efforts, the impressive accuracy numbers reported by many studies will continue to be limited by their lack of generalizability, hindering fair comparison and progress across the field.

2) Need for broader applications

The transition from research to real-world deployment requires a focus on model optimization. The proposed model, while efficient, could be further enhanced for resource-constrained devices through techniques such as model pruning and quantization. Such improvements would enable its deployment in a broader range of applications, including intelligent education systems where the analysis of student engagement and behavior is becoming a key area of research.

The successful development of an accurate and robust fake and real smile detection system has significant implications beyond traditional emotion recognition. While our work primarily focuses on the technical aspects of model development, the potential applications of this technology are vast. One such area is education, where systems capable of discerning genuine from posed expressions could serve as valuable tools for assessing student engagement, well-being, or understanding of complex topics. Chen and Zou [46] highlighted that intelligent devices like Mixed Reality (MR) systems are being explored to analyze classroom behavior and enhance teaching quality. The ability of our model to accurately identify authentic expressions could be integrated into such systems to provide teachers with deeper insights into student emotional states and engagement, thereby improving educational outcomes and promoting equity, particularly in remote learning environments. This underscores the potential for our research to contribute to the growing field of artificial intelligence in education and provides a clear direction for future work.

To sum up, this application of smile detection in education, however, raises significant ethical considerations. For AI to be effective in learning environments, it must be safe, free of algorithmic bias, and protect student data. If a model trained on a demographically narrow dataset, as acknowledged by the core research, is deployed to analyze student behavior, it could misinterpret the expressions of students from underrepresented groups, leading to erroneous and potentially harmful pedagogical interventions.

The impressive 99% accuracy on a limited dataset creates a false sense of security, masking a critical, real-world failure mode. Therefore, future work in this domain is not just a technical challenge but an ethical imperative to ensure that these powerful technologies are developed with transparency, fairness, and a commitment to responsible deployment.

V. Conclusion

The study successfully developed a CNN-based model to distinguish between real and fake smiles, achieving an exceptional testing accuracy of 99%. This achievement not only demonstrates the effectiveness of our model but also significantly outperforms previous methods, highlighting the potential of deep learning methods in facial expression analysis. According to the results, the CNN-based SIFT model effectively distinguished between real and fake smiles, opening up prospects for real-world applications in fields such as psychology, security, and human-computer interaction. However, even with our model's encouraging

outcomes, there are still a lot of areas that need further investigation. Future research could benefit from dataset enlargement by adding a larger and more diverse collection that includes a wider range of demographic factors, such as age and ethnicity, as well as other environmental situations, which would improve the model's generalizability and robustness. While the present study demonstrates the effectiveness of a hybrid CNN + Dense SIFT approach for distinguishing real and fake smiles, we acknowledge the limitations of using a relatively small, self-curated dataset without demographic diversity metrics. As future work, we plan to expand the dataset by incorporating larger, publicly available facial expression repositories and by curating additional high-quality samples under varied demographic attributes (age, gender, and ethnicity), lighting conditions, and naturalistic environments. This will allow for more comprehensive evaluation and improved generalizability of the model. Furthermore, we aim to include ethical documentation and reproducibility protocols, ensuring transparent sourcing, annotation, and validation procedures. This will help facilitate benchmarking against standardized datasets and allow fair comparison with contemporary deep learning architectures such as ResNet and transformer-based models.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

All authors have participated in proposing and validating the idea of the research; R. K. S. analyzed the data and conducted the experiments; R. K. S and H. K. wrote the paper; H. K. and N. A. H. revised the paper; all authors have approved the final version.

REFERENCES

[1] Mundzir, R. Zulkarnain, R. Hardi *et al*., "AI-powered smart smile: Early detection of mental health conditions through computational intelligence," in *Proc. 2024 18th International Conference on Telecommunication Systems, Services, and Applications (TSSA)*, 2024, pp. 1–6.

[2] I. Torre, J. Goslin, and L. White, "If your device could smile: People trust happy-sounding artificial agents more," *Computers in Human Behavior*, vol. 105, 106215, 2020.

[3] J. H. Luu, A. M. Acevedo, V. Pourmand, and S. D. Pressman, "The power of smiles: Mitigating pain through facial expression," *The Journal of Positive Psychology*, pp. 1–10, 2025.

[4] C. Halkiopoulos, E. Gkintoni, A. Aroutzidis, and H. Antonopoulou, "Advances in neuroimaging and deep learning for emotion detection: A systematic review of cognitive neuroscience and algorithmic innovations," *Diagnostics*, vol. 15, no. 4, p. 456, 2025.

[5] M. A. Khan, D. Chattaraj, S. Tadkal *et al*., "A Context for human behavior recognition using facial expression," in *Proc. 2025 International Conference on Computational, Communication and Information Technology (ICCCIT)*, 2025, pp. 275–280.

[6] F. Fatimatuzzahra, L. Lindawati, and S. Soim, "Development of convolutional neural network models to improve facial expression recognition accuracy," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 10, no. 2, pp. 279–289, 2024.

[7] N. Kumar, S. K. Pal, P. Agarwal *et al*., "Harnessing artificial emotional intelligence for improved human-computer interactions," *IGI Global*, 2024.

[8] M. Madhura, S. Meghana, and V. Varshitha, "Neural networks and emotions: A deep learning perspective," in *Proc. 2024 IEEE 9th International Conference for Convergence in Technology (I2CT)*. 2024, pp. 1–7.

[9] B. A. Erol, A. Majumdar, P. Benavidez *et al*., "Toward artificial emotional intelligence for cooperative social human–machine interaction," *IEEE Transactions on Computational Social Systems*, vol. 7, no. 1, pp. 234–246, 2019.

[10] X. Sun, P. Xia, L. Zhang *et al.*, "A ROI-guided deep architecture for robust facial expressions recognition," *Information Sciences*, vol. 522, pp. 35–48, 2020.

[11] M. Z. Hossain, T. Gedeon, and R. Sankaranarayana, "Using temporal features of observers' physiological measures to distinguish between genuine and fake smiles," *IEEE Transactions on Affective Computing*, vol. 11, no. 1, pp. 163–173, 2018.

[12] A. Lutfhi, "The effect of layer batch normalization and droupout of CNN model performance on facial expression classification," *JOIV: International Journal on Informatics Visualization*, vol.6, no. 2, pp. 481–488, 2022.

[13] M. M. Moussa, U. Tariq, F. Al-Shargie *et al*., "Discriminating fake and real smiles using electroencephalogram signals with convolutional neural networks," *IEEE Access*, vol. 10, pp. 81020–81030, 2022.

[14] P. Thanapol, K. Lavangnananda, P. Bouvry *et al*., "Reducing overfitting and improving generalization in training Convolutional Neural Network (CNN) under limited sample sizes in image recognition," in *Proc. 2020-5th International Conference on Information Technology (InCIT)*, 2020, pp. 300–305.

[15] W. R. Abdulhussien, N. K. E. Aabbadi, and A. M. Gaber, "Hybrid deep neural network for facial expressions recognition," *Indonesian Journal of Electrical Engineering and Informatics (IJEEI)*, vol. 9, no. 4, pp. 993–1007, 2021.

[16] J. C. Kim, M. H. Kim, H. E. Suh *et al.*, "Hybrid approach for facial expression recognition using convolutional neural networks and SVM," *Applied Sciences*, vol. 12, no. 11, p. 5493, 2022.

[17] M. Kaur, and M. Kumar, "Facial emotion recognition: A comprehensive review," *Expert Systems*, vol. 41, no. 10, e13670, 2024.

[18] R. Mishra, R. Satpathy, and B. Pati, "Interpretable AI in medical imaging: Enhancing diagnostic accuracy through human-computer interaction," *Journal of Artificial Intelligence and Systems*, vol. 6, no. 1, pp. 96–111, 2024.

[19] A. Singh, R. Saxena, and S. Saxena, "The human touch in the age of artificial intelligence: A literature review on the interplay of emotional intelligence and AI," *Asian Journal of Current Research*, vol. 9, no. 4, p. 15, 2024.

[20] H. U. Ukwu, and K. Yurtkan, "4D facial expression recognition using geometric landmark-based axes-angle feature extraction," *Intelligent Automation & Soft Computing*, vol. 34, no. 3, pp. 1820–1838, 2022.

[21] M. Hazgui, H. Ghazouani, and W. Barhoumi, "Evolutionary-based generation of rotation and scale invariant texture descriptors from SIFT keypoints," *Evolving Systems*, vol. 12, no. 3, pp. 591–603, 2021.

[22] M. Srikar, and K. Malathi, "A supervised stable object detection with image feature extraction using image segmentation by comparing Histogram of Oriented Gradients (HOG) algorithm over Scale Invariant Feature Transform (SIFT) algorithm model," *Journal of Pharmaceutical Negative Results*, vol. 13, no. 4, 2022.

[23] Y. Yang, S. Liu, H. Zhang *et al*., "Multi-modal remote sensing image registration method combining scale-invariant feature transform with co-occurrence filter and histogram of oriented gradients features," *Remote Sensing*, vol. 17, no. 13, p. 2246, 2025.

[24] E. H. Houssein, A. M. Gamal, E. M. Younis, and E. Mohamed, "Explainable artificial intelligence for medical imaging systems using deep learning: A comprehensive review," *Cluster Computing*, vol. 28, no. 7, p. 469, 2025.

[25] C. C. Nguyen, G. S. Tran, T. P. Nghiem *et al*., "Real-time smile detection using deep learning," *Journal of Computer Science and Cybernetics*, vol. 35, no. 2, pp. 135–145, 2019.

[26] U. A. Qurat, N. Nida, A. Irtaza *et al*., "Forged face detection using ELA and deep learning techniques," in *Proc. 2021 International Bhurban Conference on Applied Sciences and Technologies (IBCAST)*, 2021, pp. 271–275.

[27] H. K. Kondaveeti, and M. V. Goud, "Emotion detection using deep facial features," in *Proc. 2020 IEEE International Conference on*

Advent Trends in Multidisciplinary Research and Innovation (ICATMRI), 2020, pp. 1–8.
[28] H. L. Majeed, O. A. Hassen, D. A. Farhan *et al.*, "Computer vision of smile detection based on machine and deep learning approach," *Journal of Cybersecurity & Information Management*, vol. 16, no. 1, pp. 208–230, 2025.
[29] H. Ansaf, H. Najm, J. M. Atiyah, and O. A. Hassen *et al*., "Improved approach for identification of real and fake smile using chaos theory and principal component analysis," *Journal of Southwest Jiaotong University*, vol. 54, no. 5, p. 1–11, 2019.
[30] S. T. Suganthi, M. U. A. Ayoobkhan, N. Bacanin *et al*., "Deep learning model for deep fake face recognition and detection," *PeerJ Computer Science*, vol. 8, e881, 2022.
[31] D. S. AbdElminaam, N. Sherif, Z. Ayman *et al*., "DeepFakeDG: A deep learning approach for deep fake detection and generation," *Journal of Computing and Communication*, vol. 2, no. 2, p. 31–37, 2023.
[32] S. Ramachandran, A. V. Nadimpalli, and A. Rattani, "An experimental evaluation on deepfake detection using deep face recognition," in *Proc. 2021 International Carnahan Conference on Security Technology (ICCST)*, 2021, pp. 1–6.
[33] A. Jaiswal, A. K. Raju, and S. Deb, "Facial emotion detection using deep learning," in *Proc. 2020 International Conference for Emerging Technology (INCET)*, 2020, pp. 1–5.
[34] S. A. Hussain, and A. S. A. A. Balushi, "A real time face emotion classification and recognition using deep learning model," *Journal of Physics: Conference Series*, vol. 1432, no. 1, 012087, 2020.
[35] C. Singla, S. Singh, P. Sharma *et al*., "Emotion recognition for human–computer interaction using high-level descriptors," *Scientific Reports*, vol. 14, no. 1, 12122, 2024.
[36] A. I. Siam, N. F. Soliman, A. D. Algarni *et al*., "Deploying machine learning techniques for human emotion detection," *Computational Intelligence and Neuroscience*, vol. 2022, no. 1, 8032673, 2022.
[37] M. Favorskaya and A. Yakimchuk, "Fake face image detection using deep learning-based local and global matching," in *Proc. CEUR Workshop*, 2021, pp. 133–138.
[38] H. N. Divya, Y. Sandeep, Y. T. Raghav *et al*., "Fake smile classifier," *International Journal of Creative Research Thoughts (IJCRT)*, vol. 13, no. 2, pp. 1–8, 2025.
[39] S. Jia, S. Wang, C. Hu *et al*., "Detection of genuine and posed facial expressions of emotion: databases and methods," *Frontiers in psychology*, vol. 11, 580287, 2021.
[40] G. Ramya, J. Sreeja, K. Jyothi *et al*., "A smile detection for hands-free selfie capture using machine learning," *INTI Journal*, vol. 9, pp. 1–10, 2025.
[41] E. A. R. Martínez, O. Polezhaeva, F. Marcellin *et al*., "DeepSmile: Anomaly detection software for facial movement assessment," *Diagnostics*, vol. 13, no. 2, p. 254, 2023.
[42] H. L. Majeed, O. A. Hassen, D. A. Farhan *et al*., "Computer vision of smile detection based on machine and deep learning approach," *Journal of Cybersecurity and Information Management*, vol. 16, no. 1, pp. 208–230, 2025.
[43] S. Yavuzkilic, A. Sengur, Z. Akhtar, and K. Siddique, "Spotting deepfakes and face manipulations by fusing features from multi-stream CNNs models," *Symmetry*, vol. 13, no. 8, p. 1352, 2021.
[44] H. Xing, S. Y. Tan, F. Qamar, and Y. Jiao, "Face anti-spoofing based on deep learning: A comprehensive survey," *Applied Sciences*, vol. 15, no. 12, p. 6891, 2025.
[45] A. Eiserbeck, M. Maier, J. Baum, and R. A. Rahman, "Deepfake smiles matter less—the psychological and neural impact of presumed AI-generated faces," *Scientific Reports*, vol. 13, no. 1, 16111, 2023.
[46] Y. Chen, and Y. Zou, "Enhancing education quality: Exploring teachers' attitudes and intentions towards intelligent MR devices," *European Journal of Education*, vol. 59, no. 4, e12692, 2024.