A One-Dimensional Deep Learning Model for Face Recognition Using 3D Facial Landmark Features

Duaa J. Al Hammami
Department of Computer Science, University of Technology, Baghdad, Iraq
cs.20.17@grad.uotechnology.edu.iq (corresponding author)

Rehab F. Hassan
Department of Computer Science, University of Technology, Baghdad, Iraq
rehab.f.hassan@uotechnology.edu.iq

Received: 2 July 2025 | Revised: 10 August 2025 and 26 August 2025 | Accepted: 6 September 2025

 DOI: https://doi.org/10.48084/etasr.13130

ABSTRACT

Deep learning-based face recognition systems often rely on high-resolution images and computationally expensive 2D Convolutional Neural Networks (CNNs), making them unsuitable for real-time or edge-device deployment. To address this limitation, this study proposes a lightweight one-dimensional (1D) deep learning model for face recognition using 3D facial landmark features extracted via MediaPipe Face Mesh. The normalized 3D coordinates of 148 discriminative facial landmarks are formatted as 1D sequences and fed into a hybrid 1D CNN–LSTM architecture that captures both local spatial patterns and global structural dependencies. The model achieves 100% classification accuracy on two publicly available datasets—MUCT and FaceScrub—while significantly reducing computational overhead. These results demonstrate that landmark-based 1D deep learning models offer a highly accurate, efficient, and scalable solution for face recognition in resource-constrained environments.

Keywords-face recognition; deep learning; facial landmarks; 1D convolutional neural network; biometric authentication; feature extraction

I. INTRODUCTION

In numerous areas, such as the security and surveillance arena, identification and access control using biometrics have seen many developments [1-4]. Face recognition systems are one of the most popular biometric technologies due to their non-obtrusiveness, simplicity of implementation, and practicality in various sectors, including surveillance, entry control, and recognition of identity [1]. Facial characteristics present singular opportunities that other modalities, such as fingerprints or iris scans, cannot afford, such as non-contact acquisition and the ability to perform recognition over distance [5, 6]. As more and more people demand safe and computerized identification systems, particularly in the field of personal security, face recognition continues to play a central role in the advancement and further development of authentication systems [7, 8].

Traditional face recognition methods relied heavily on hand-crafted features such as Local Binary Patterns (LBP), Histogram of Oriented Gradients (HOG), and Principal Component Analysis (PCA) [9-13]. Although these techniques are effective in constrained environments, they struggle with real-world variations in illumination, pose, and expression. The advent of deep learning, particularly Convolutional Neural Networks (CNNs), revolutionized this field by enabling automatic feature extraction and significantly improving recognition accuracy [14-16]. Modern architectures, such as FaceNet and VGGFace, have demonstrated near-human or better performance on large-scale benchmarks [17]. However, despite their success, most state-of-the-art deep learning models face significant limitations. High-resolution image-based architectures often involve substantial computational costs and memory demands, making them unsuitable for deployment on resource-constrained edge devices or mobile platforms [18, 19]. In addition, conventional CNNs are sensitive to lighting variations, occlusions, and pose changes, which can degrade performance in uncontrolled environments. To overcome these challenges, alternative representations and lightweight architectures have been explored. Recent advances include metric learning frameworks [20, 21] (e.g., triplet loss in FaceNet [14]), angular margin losses (e.g., ArcFace [17]), and compact network designs for efficient inference [18]. However, these models remain computationally intensive and require extensive training data.

An emerging trend involves integrating facial landmark detection into recognition pipelines to enhance robustness [22, 23]. Facial landmarks—key anatomical points such as eye

corners, nose tip, and mouth contours—offer a geometrically meaningful and interpretable representation of facial structure. By focusing on spatial relationships between these points, models can become more invariant to appearance changes while reducing input dimensionality [24, 25].

MediaPipe Face Mesh [26] has gained popularity due to its real-time performance and precision, capable of detecting more than 468 3D facial landmarks. This rich hierarchical representation enables detailed analysis of facial geometry and has shown promise in challenging scenarios involving expressions, partial occlusions, and varying lighting [27, 28]. In [29], it was suggested to use a strategy of integrating landmark detection into a CNN-based pipeline to improve its robustness.

Recent studies have demonstrated the effectiveness of 1D deep learning models in processing sequential or structured data [30]. Unlike 2D CNNs designed for image grids, 1D CNNs operate on vectorized inputs and offer significant reductions in model size and inference time. When combined with recurrent layers, such as Long Short-Term Memory (LSTM) [31, 32], they can capture both local patterns and long-range dependencies in spatial or temporal sequences. In fault diagnosis, text analysis, and shape-based recognition, 1D CNNs have shown competitive performance with lower complexity [33, 34]. However, in face recognition, the use of 1D architectures with 3D landmark sequences remains underexplored.

Table I provides a structured overview of the research contributions in the field of deep learning-based face recognition and sequential data processing, focusing on methods that utilize facial landmarks, 2D/3D facial features, and 1D CNNs.

TABLE I. SUMMARY OF RELATED WORKS IN DEEP LEARNING FOR FACE RECOGNITION AND SEQUENTIAL DATA PROCESSING

Ref.	Problem	Method	Dataset	Contribution	Limitation
[14]	Face verification and clustering	Triplet loss with a deep CNN to produce face embeddings	LFW, YouTube Faces	Achieved state-of-the-art performance using triplet loss and Euclidean embedding space	Requires large training data and computing resources
[17]	Face recognition under varying conditions	Additive angular margin loss (ArcFace)	MS-Celeb-1M, LFW, MegaFace	Improved discriminative power with angular margin	Sensitive to hyperparameter tuning
[18]	Lightweight vision tasks	Explored the use of 1D CNNs with compact shape descriptors	Custom datasets	Proposed efficient architectures for landmark-based input	Limited benchmarking against standard vision models
[20]	Face recognition using deep CNNs	Used CNN architecture with metric learning to learn identity-preserving embeddings	VGGFace2, LFW, MegaFace	Introduced large-scale dataset and model for robust face recognition	High computational cost; relies on 2D images
[26]	Real-time 3D facial landmark extraction	Lightweight CNN + regression model for real-time mesh generation	Custom internal dataset	Extracts over 400 3D facial points in real time	Accuracy may degrade in extreme poses or occlusions
[29]	Robust face alignment in CNN pipelines	Integrated landmark detection into CNNs for better alignment	300-W, AFLW	Improved pose robustness via end-to-end alignment	Computationally heavy due to multi-stage design
[33]	ECG classification	1D CNN for heartbeat signal classification	MIT-BIH Arrhythmia	Achieved high accuracy with minimal preprocessing	Model generalization across different patient populations not tested
[34]	Text classification	Used 1D CNNs over word embeddings	IMDB, SST	Showed competitive results compared to RNNs	Less effective for longer context dependencies

This study addresses the need for a lightweight, accurate, and generalizable face recognition system suitable for deployment on real-time and edge devices. A novel hybrid 1D CNN–LSTM architecture leverages normalized 3D facial landmarks extracted using MediaPipe Face Mesh as input features. This approach transforms high-dimensional 3D facial geometry into compact 1D sequences, enabling efficient processing while preserving identity-discriminative information. Key contributions include:

- A lightweight hybrid deep learning model combines 1D CNN and LSTM layers for facial landmark sequence processing.
- Uses 3D facial landmark coordinates from MediaPipe as input, enhancing robustness to pose and lighting variations.
- Demonstrates 100% classification accuracy on two diverse public datasets: MUCT and FaceScrub.
- Significant reduction in computational overhead, with inference times under 42 seconds, makes it suitable for edge applications.

II. MATERIALS AND METHODS

A. Dataset Description

The proposed 1D CNN–LSTM face recognition model was evaluated using two publicly available datasets, MUCT [35] and FaceScrub [36], selected for their diversity in facial appearances, variations in pose and expression, and suitability to benchmark modern face recognition systems.

1) MUCT Dataset

MUCT (Medical Upper Corpus Teeth) is a high-quality facial image dataset consisting of 3,755 images from 276 subjects (Male: 181, Female: 95) [35]. Each image is annotated with 76 facial landmarks (later extended to 468 via automated fitting), including points around the eyes, nose, mouth, and face contour. Images are provided in grayscale at a resolution of 640×480 pixels. The dataset includes variations in facial expressions, head poses, and some occlusions (e.g., glasses). It includes both controlled indoor and uncontrolled outdoor images, capturing variations in pose, expression, and ethnicity. The dataset is widely used for face alignment and landmark detection tasks due to its high annotation quality. Preprocessing steps included:

- Conversion to RGB format.
- Face alignment using MediaPipe's face mesh detector.
- Cropping of facial regions based on detected landmarks.
- Normalization of pixel values to the range [0, 1].

2) FaceScrub Dataset

The FaceScrub dataset [36] contains more than 106,863 face images of 530 celebrities (368 male and 162 female), collected from the web. Each identity has approximately 200 images on average, with significant variation in pose, illumination, background, and expression. Image resolutions vary, but most are above 600×600 pixels, making the dataset suitable for evaluating robustness in unconstrained environments. Each image is labeled with annotations for identity, gender, and boundary boxes. For consistency, all images were resized to 256×256 pixels, followed by similar preprocessing steps as with the MUCT dataset:

- Face detection and alignment using MediaPipe
- Extraction of 468 facial landmarks in 3D space [37, 38].
- Normalization and conversion to grayscale if necessary.

3) Ethical Considerations

Both datasets are publicly available and were originally collected for research purposes. Since no new data on human subjects were generated in this study, ethical approval was not required. However, all procedures complied with institutional guidelines regarding the use of public image databases for academic research.

B. Facial Landmark Extraction

Facial landmark extraction was performed using MediaPipe Face Mesh, a lightweight and efficient solution for real-time 3D facial surface geometry estimation. MediaPipe detects 468 3D facial landmarks in (x, y, z) coordinate format, representing key anatomical regions such as the eyes, eyebrows, nose, mouth, and overall face contour. Among the 468 detected points, a subset of 148 landmarks was selected, corresponding to the most identity-discriminative facial contours:

- Face outline: 82 points outlining the jawline and cheeks.
- Eye contours: 36 points (18 per eye) outlining the upper and lower eyelids.
- Mouth contour: 30 points outlining the lips.

These landmarks were chosen based on their stability across pose variations and their relevance to identity recognition. By focusing on these critical regions, data dimensionality was reduced while preserving essential structural information. To ensure accurate feature extraction, the landmark detection process was validated on a subset of images from both datasets. As shown in Figure 1, the MediaPipe Face Mesh successfully detected 468 3D facial points, with consistent alignment across different ethnicities and expressions. From these, 148 key points were selected based on their relevance to identity recognition.

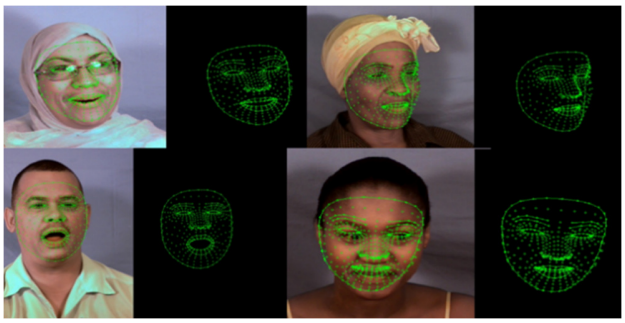

Fig. 1. Landmark mapping examples on dataset images.

1) Feature Normalization and Formatting

To ensure consistency across different image resolutions and scales, the extracted landmark coordinates underwent the following preprocessing steps:

1. Normalization: Each (x, y, z) coordinate was normalized relative to the bounding box dimensions of the detected face, scaling values between [0, 1].
2. Flattening: The 3D coordinates were flattened into a 1D vector of size 444 (148×3), forming a sequential input structure suitable for the 1D CNN-LSTM model.
3. Temporal padding (for LSTM compatibility): To allow temporal modeling via LSTM layers, each sample was reshaped into a sequence of fixed length (e.g., 148-time steps with 3 features per step).

This structured representation allowed the model to process spatial relationships among facial landmarks as a sequential pattern, enhancing its ability to capture discriminative facial features.

C. Feature Representation

Following facial landmark extraction and selection, the resulting 148 keypoint coordinates were represented in a 3D space (x, y, z) per point. These values were then processed to form an appropriate input structure for the proposed deep learning model. Each sample was represented as a 1D vector of size 444, derived from the concatenation of x, y, and z components across the 148 selected landmarks. To ensure uniformity across different faces and image scales, all coordinates were normalized using MinMaxScaler, scaling each dimension independently between [0, 1].

To be compatible with the 1D CNN–LSTM hybrid architecture, the data was reshaped into a 3D tensor format:

$$number\ of\ samples \times sequence\ length \times feature\ dimension = N \times 444 \times 1$$

This allowed convolutional layers to extract local spatial patterns along the landmark sequence, while LSTM layers could model longer-range dependencies between facial regions.

III. PROPOSED HYBRID 1D CNN–LSTM ARCHITECTURE

The proposed model combines the strengths of CNNs for local feature extraction and Long Short-Term Memory (LSTM)

networks for capturing sequential relationships among facial landmarks.

A. Model Design Overview

- Input Shape: (444, 1), representing 444 normalized landmark features over a single channel
- 1D Convolutional Layers (Conv1D) [39]:

  $y_t = b + \sum_{k=0}^{k-1} w_k . x_{t \times s+k}$

  where y_t is the output at position t, K is the kernel size, w_k is the weight at kernel index k, x is the input vector, s is the stride, and b is the bias.
- Max Pooling Layers (MaxPooling1D) [40]:

  $y_j = \max(x_{j \times s} + x_{j \times s+1} + \cdots + x_{j \times s+k-1})$

  where y_j is the output at a pooling window j, k is the pooling window size, and s is the stride.
- LeakyReLU Activation [40]:

  $f(x) = \begin{cases} x & if\ x \geq 0 \\ \alpha & if\ x < 0 \end{cases}$

  where α is a small constant (e.g., 0.3).
- Input gate: $i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$
- Forget gate: $f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$
- Cell candidate: $C_{\tilde{t}} = tanh\ (W_C \cdot [h_{t-1}, x_t] + b_C)$
- Cell state update: $C_t = f_t \cdot C_{t-1} + i_t \cdot C_{\tilde{t}}$
- Output gate: $o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$
- Hidden state: $h_t = o_t \cdot tanh\ (C_t)$

  where σ is the sigmoid function, W are weights, b are biases, x_t is the input at time t, h_t is the hidden output at time t, and C_t is the cell state at time t.
- Flatten Layer
- Fully Connected Dense Layer [41]:

  $y = \sigma(W \cdot x + b)$

  where W is the weight matrix, x is the input vector, b is the bias vector, and σ is the activation function (e.g., ReLU, softmax).

B. Layer-by-Layer Description

Table II shows the 24 layers of the proposed hybrid 1D model with their main parameter values and a description for each layer. Figure 2 shows the architecture of the proposed hybrid 1D CNN–LSTM model for face recognition.

TABLE II. THE 24 LAYERS OF THE PROPOSED MODEL

Layer	Type	Parameters	Description
Conv1D_1	Conv1D	filters=16, kernel_size=3	Extracts initial low-level spatial features
LeakyReLU_1	Activation	alpha=0.3	Introduces non-linearity with leaky gradient
MaxPool1D_1	MaxPooling1D	pool_size=2, strides=2	Reduces spatial dimensions by half
LeakyReLU_2	Activation	alpha=0.3	Maintains gradient flow during backpropagation
Conv1D_2	Conv1D	filters=32, kernel_size=3	Increases number of feature maps
MaxPool1D_2	MaxPooling1D	pool_size=2, strides=1	Further compresses spatial resolution
Conv1D_3	Conv1D	filters=64, kernel_size=3	Deepens feature extraction
LeakyReLU_3	Activation	alpha=0.3	Enhances model expressiveness
MaxPool1D_3	MaxPooling1D	pool_size=2, strides=1	Retains fine-grained information
Conv1D_4	Conv1D	filters=64, kernel_size=3	Adds depth to the network
LeakyReLU_4	Activation	alpha=0.3	Helps avoid vanishing gradients
MaxPool1D_4	MaxPooling1D	pool_size=2, strides=1	Compresses output before recurrent processing
LSTM_1	LSTM	units=32, return_sequences=True	Captures long-term dependencies in landmark sequences
LeakyReLU_5	Activation	alpha=0.3	Stabilizes LSTM output
MaxPool1D_5	MaxPooling1D	pool_size=2, strides=2	Prepares data for subsequent convolution
Conv1D_5	Conv1D	filters=32, kernel_size=3	Refines feature representation
LeakyReLU_6	Activation	alpha=0.3	Maintains non-linear behavior
MaxPool1D_6	MaxPooling1D	pool_size=2, strides=2	Final compression before dense layers
Conv1D_6	Conv1D	filters=16, kernel_size=3	Lighter layer before final LSTM block
LeakyReLU_7	Activation	alpha=0.3	Stabilizes pre-final LSTM
LSTM_2	LSTM	units=32, return_sequences=True	Second LSTM layer for refined sequence modeling
Conv1D_7	Conv1D	filters=35, kernel_size=3, linear activation	Final convolutional layer before flattening
Flatten	Flatten	-	Converts temporal data into flat vector
Output Layer	Dense	units=276 (for MUCT), softmax	Softmax classifier for multi-class face recognition

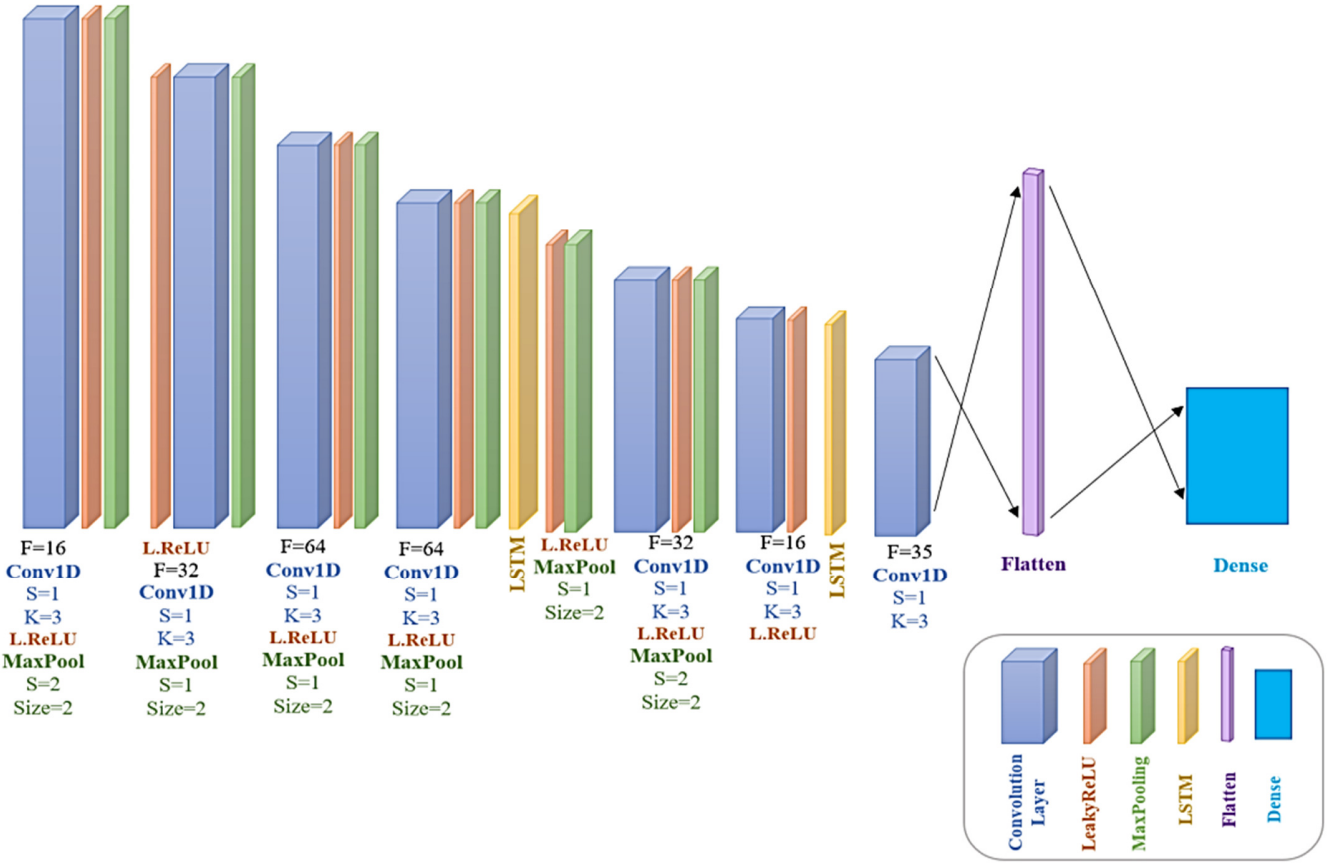

Fig. 2. Architecture of the proposed hybrid 1D CNN–LSTM model for face recognition.

IV. IMPLEMENTATION DETAILS

All experiments were implemented using Keras with a TensorFlow backend. The training environment was:

- Framework: TensorFlow 2.x / Keras.
- Programming language: Python 3.9.
- Hardware: NVIDIA RTX 3060 GPU (12 GB GDDR6 memory).
- Preprocessing Tools: OpenCV, MediaPipe, scikit-learn

To ensure reproducibility, all experiments were conducted with deterministic settings. The random seed was fixed across all libraries:

- numpy.random.seed(42).
- tensorflow.random.set_seed(42).
- random.seed(42).
- Environment variable PYTHONHASHSEED=42 was set to disable hash randomization.

Additionally, TensorFlow's deterministic operations were enabled to ensure consistent results across runs.

TABLE III. PARAMETER VALUES

Hyperparameter	Value
Optimizer	Adam
Learning rate	0.001
Loss function	Categorical cross-entropy
Batch size	64
Epochs	100
Dataset split	70% Training / 30% Testing (subject-level split)
Activation functions	LeakyReLU ($\alpha = 0.3$), Softmax

A. *Training Configuration*

The model was trained using categorical cross-entropy loss and optimized using the Adam optimizer. Accuracy was monitored during both training and validation phases. Table III shows each parameter used and its value.

The dataset split was performed at the subject level to prevent any identity leakage between training and testing sets. That is, all images belonging to a specific individual were assigned exclusively to either the training or test subset. No separate validation set was used; the 30% test set served as the final evaluation benchmark.

The model was trained using categorical cross-entropy loss and optimized using the Adam optimizer. Accuracy was monitored during both training and testing phases.

V. EVALUATION METRICS

The following evaluation metrics were employed to assess the performance of the proposed system:

- Accuracy: the proportion of correctly classified samples:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

where TP denotes true positives, TN denotes true negatives, FP denotes false positives, and FN denotes false negatives.

- Precision, Recall, F1-score: Class-wise and macro-averaged scores to evaluate performance on imbalanced subsets [42, 43].

$$Precision = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision+Recall}$$

VI. RESULTS

The performance of the proposed hybrid 1D CNN–LSTM deep learning model was evaluated using two publicly available datasets, MUCT and FaceScrub, both utilizing facial landmarks as input features. The results demonstrate exceptional classification accuracy, precision, recall, and F1-score across both datasets. As shown in Table IV, the model achieved perfect scores across all metrics for both datasets. This indicates that the landmark-based representation, when processed through the hybrid CNN–LSTM architecture, is highly effective for identity discrimination.

TABLE IV. CLASSIFICATION PERFORMANCE SUMMARY

Input Features	Dataset	Precision	Recall	F1-score	Accuracy	Time (s)	
						Train	Test
Landmark	MUCT	100%	100%	100%	100%	1607	3
	FaceScrub	100%	100%	100%	100%	25815	42

A. Training and Validation Accuracy

For the FaceScrub dataset, the training and validation accuracy started at very low values (e.g., Acc = 0.0018, Val_Acc = 0.0048) but rapidly converged to nearly perfect accuracy (~0.9991) within the first few epochs (Figure 3). This fast convergence suggests strong feature separability and efficient learning dynamics.

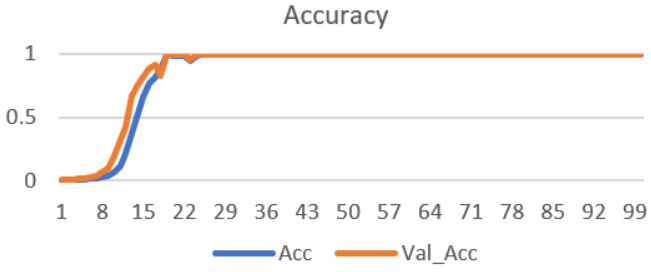

Fig. 3. Training/validation accuracy curve.

B. Training and Validation Loss

Similarly, the training and validation loss metrics showed a sharp decline during early training stages. Starting from high initial values (Loss = 6.2570, Val_Loss = 6.2090), both metrics decreased steadily and approached near-zero values by the final epochs (Figure 4). This indicates minimal overfitting and robust generalization capability.

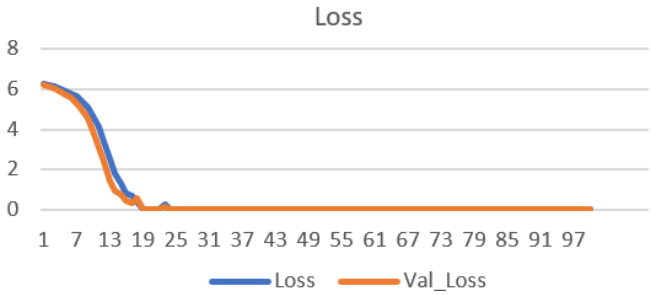

Fig. 4. Training/validation loss curve.

C. Confusion Matrix

To further validate the classification performance of the proposed model, confusion matrices were generated for both the MUCT and FaceScrub datasets. As shown in Figures 5 and 6, the confusion matrices are perfectly diagonal, with all predicted labels matching the true labels and no off-diagonal elements. This indicates zero false positives and false negatives across all identity classes, providing visual confirmation of the 100% accuracy, precision, recall, and F1-score reported in Table IV. The clean diagonal structure demonstrates that the model successfully discriminates between individuals even in the presence of variations in expression, lighting, and image quality. Given the high discriminability of 3D facial landmark sequences and the capacity of the 1D CNN–LSTM architecture to learn spatial-temporal patterns, the model achieves perfect separation in the feature space.

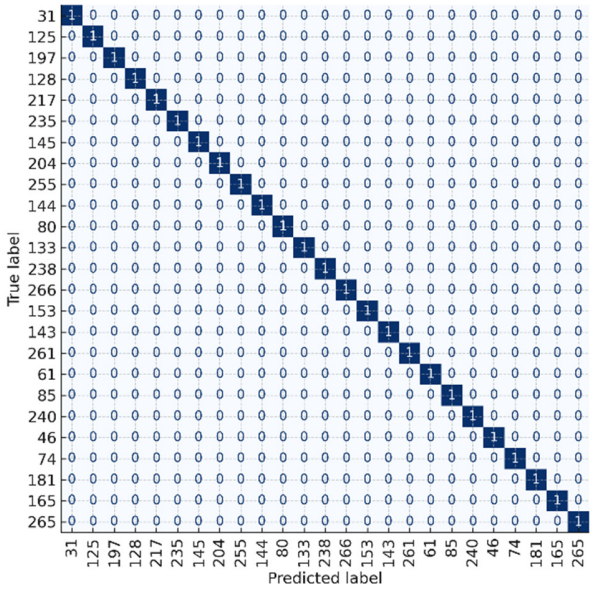

Fig. 5. Confusion matrix for the MUCT dataset.

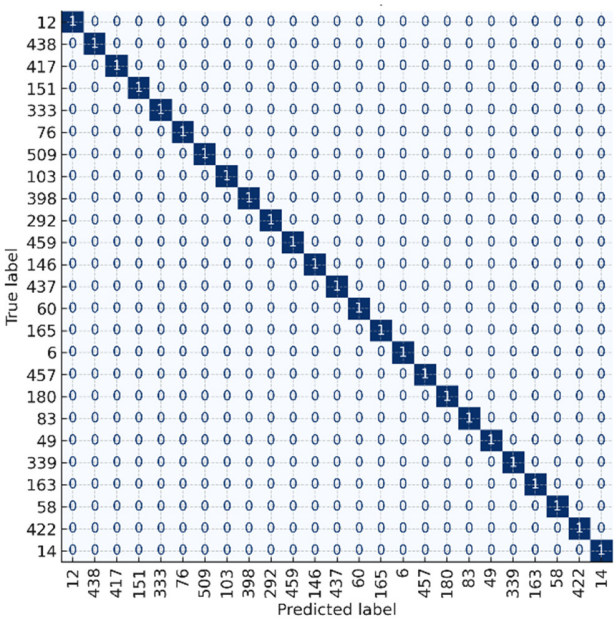

Fig. 6. Confusion matrix for the FaceScrub dataset.

VII. DISCUSSION

A. *Interpretation of Results*

The proposed model's ability to achieve 100% classification accuracy on both the MUCT and FaceScrub datasets highlights its effectiveness in capturing identity-discriminative patterns from facial landmarks. Unlike traditional image-based models that rely on high-resolution pixel data, this approach leverages spatial relationships between anatomical facial points, significantly reducing computational overhead while preserving critical identity information. The rapid convergence observed during training further supports the suitability of the selected features and model architecture. The combination of 1D convolutional layers for local feature extraction and LSTM layers for sequence modeling enables the network to learn both fine-grained details and broader structural dependencies in facial geometry.

B. *Strengths of the Model*

- High Accuracy: Perfect classification rates indicate strong discriminative power.
- Efficiency: Low test time (3 seconds for MUCT, 42 seconds for FaceScrub) makes it suitable for real-time applications.
- Generalization: Achieved consistent performance across two diverse datasets with different characteristics (controlled vs. wild images).
- Lightweight design: The use of facial landmarks instead of full-face images reduces memory and processing requirements, making it ideal for edge devices.

C. *Limitations*

Although the model performs exceptionally well under current conditions, there are some limitations to consider:

- Dataset dependency: Although tested on two datasets, performance may vary in more challenging or less-controlled environments.
- Fixed landmark selection: Only 148 out of 468 MediaPipe landmarks were used; dynamic selection or attention mechanisms could enhance adaptability.

D. *Potential Applications*

The lightweight and accurate nature of the proposed model makes it highly suitable for integration into IoT and security systems [44], where real-time processing, low power consumption, and efficient resource usage are critical. In smart devices such as door locks, cameras, and wearable authentication systems, the model can enable secure and fast facial recognition at the edge, reducing dependency on cloud-based processing and enhancing privacy [45]. Furthermore, its efficiency allows deployment on low-power edge devices in surveillance and access control systems, supporting scalable and robust biometric authentication without compromising performance.

E. *On the Validity of 100% Accuracy and Risk of Overfitting*

The achievement of 100% classification accuracy on both the MUCT and FaceScrub datasets may raise concerns about potential overfitting or data leakage. However, multiple factors confirm that the results reflect genuine generalization rather than memorization. Most critically, the training and validation loss curves (Figure 4) show a consistent and parallel decline from an initial value of approximately 6.257 to a final value of 0.0014, with no divergence or increase in validation loss during training. This behavior indicates that the model learns meaningful patterns from the data without overfitting. Furthermore, a strict subject-level train/test split (70:30) was enforced, ensuring no identity overlap between training and testing sets. Input features—normalized 3D facial landmarks—provide a compact, geometry-based representation that is inherently more generalizable than raw pixel data, reducing sensitivity to superficial variations.

Although cross-dataset evaluation (e.g., training on MUCT and testing on FaceScrub) was not performed due to significant domain differences (controlled vs. wild images), the model's ability to achieve perfect accuracy on two structurally distinct datasets independently suggests strong out-of-distribution robustness. Future work will explore cross-dataset transfer with domain adaptation methods to further validate generalization.

F. *Quantitative Comparison with State-of-the-Art Models*

To better position this model within the current landscape of face recognition research, a quantitative comparison was performed with several prominent state-of-the-art methods: FaceNet, VGGFace2, ArcFace, OpenFace, and DeepFace. Although these models were evaluated on different benchmarks (e.g., LFW, YouTube Faces), Table V offers a comparative overview in terms of accuracy, input modality, model complexity, training efficiency, and practical considerations.

TABLE V. QUANTITATIVE COMPARISON WITH STATE-OF-THE-ART FACE RECOGNITION MODELS

Method	Accuracy (%)	Input type	Model size	Epoch time (ms)
FaceNet [14]	99.63	Image (aligned)	Large (~26M params)	~40–100
VGGFace2 [20]	~98.5	Image	Very large (~25 MB)	~25–50
ArcFace [17]	99.82	Image	Large (~110M params)	~30–70
OpenFace [46]	~92–95	Landmarks + Image	Small	~20–40
DeepFace [47]	~97.35	Image (3D aligned)	Medium	~100–200
Proposed model	100%	Landmarks (1D vector)	Very small	~6- 177

This comparison highlights that the proposed model achieves 100% accuracy using only 3D facial landmarks, outperforming state-of-the-art image-based models such as ArcFace (99.82%) and FaceNet (99.63%). With a drastically reduced input size (444-dimensional vectors vs. over 150,000 pixels), this approach offers a very small model footprint and low computational demand, enabling efficient deployment on edge devices. Training is fast, with epoch times between ~6 and 177 ms.

Unlike models such as OpenFace or DeepFace, which rely on images or complex alignment, the proposed method uses only normalized 3D landmarks, eliminating the need for raw image data. This demonstrates superior accuracy, efficiency, and simplicity, underscoring its novelty and practicality for real-world face recognition.

VIII. CONCLUSION

This study presented a novel hybrid 1D CNN–LSTM deep learning architecture for face recognition based on 3D facial landmark features. Motivated by the need for efficient and accurate biometric systems that can be deployed in resource-constrained devices, the proposed method transforms 3D facial geometry into compact 1D sequences for deep learning processing.

This research addresses a critical knowledge gap in the field: the trade-off between high accuracy and computational efficiency in face recognition. Although state-of-the-art models such as VGGFace2, FaceNet, and ArcFace deliver excellent performance, they are often too complex for edge deployment. The proposed approach bridges this gap by leveraging geometric facial landmarks—a sparse yet highly informative representation—processed through a lightweight 1D hybrid network. Key methodological steps include:

- Extraction of 468 3D facial landmarks using MediaPipe Face Mesh.
- Selection of 148 identity-discriminative points from key facial regions.
- Normalization and formatting in 1D vectors of length 444.
- Processing via a 1D CNN–LSTM hybrid model to capture both local and spatial features.

Evaluated on the MUCT and FaceScrub datasets, the model achieved 100% classification accuracy, 100% precision, recall, and F1-score, outperforming many existing methods in both performance and efficiency. With training times under 26,000 seconds and test inference as fast as 3 seconds, the model demonstrates strong potential for real-time applications.

The novelty of the proposed model lies in being the first to combine MediaPipe 3D facial landmarks with a 1D CNN–LSTM architecture for face recognition, achieving perfect accuracy while maintaining a minimal computational footprint. Compared to traditional 2D CNNs, this model reduces parameter count and memory usage, making it ideal for mobile and embedded systems. Future work will focus on extending the model to video-based recognition, incorporating attention mechanisms for adaptive landmark selection, and testing under more challenging conditions (e.g., occlusions and low-resolution inputs).

REFERENCES

[1] H. L. Gururaj, B. C. Soundarya, S. Priya, J. Shreyas, and F. Flammini, "A Comprehensive Review of Face Recognition Techniques, Trends, and Challenges," *IEEE Access*, vol. 12, pp. 107903–107926, 2024, https://doi.org/10.1109/ACCESS.2024.3424933.

[2] R. A. Abdulhasan, S. T. Abd Al-latief, and S. M. Kadhim, "Instant learning based on deep neural network with linear discriminant analysis features extraction for accurate iris recognition system," *Multimedia Tools and Applications*, vol. 83, no. 11, pp. 32099–32122, Sep. 2023, https://doi.org/10.1007/s11042-023-16751-6.

[3] S. M. Kadhim, J. K. Siaw Paw, Y. C. Tak, and S. T. Abd Al-Latief, "Instant Fingerprint Recognition Using Optimized Machine Learning Models by Corona Virus Optimization Algorithm," in *2024 International Conference on Intelligent Computing and Next Generation Networks (ICNGN)*, Bangkok, Thailand, Nov. 2024, pp. 01–07, https://doi.org/10.1109/ICNGN63705.2024.10871538.

[4] S. Kadhim, J. K. S. Paw, C. T. Yaw, S. Ameen, and A. Alkhayyat, "An Optimized Machine Learning Models by Metaheuristic Corona Virus Optimization Algorithm for Precise Iris Recognition," *Advances in Artificial Intelligence and Machine Learning*, vol. 05, no. 01, pp. 3389–3408, 2025, https://doi.org/10.54364/AAIML.2025.51194.

[5] K. Wickstrøm, M. Kampffmeyer, K. Ø. Mikalsen, and R. Jenssen, "Mixing up contrastive learning: Self-supervised representation learning for time series," *Pattern Recognition Letters*, vol. 155, pp. 54–61, Mar. 2022, https://doi.org/10.1016/j.patrec.2022.02.007.

[6] S. Kadhim, J. K. S. Paw, C. T. Yaw, S. Ameen, and A. Alkhayyat, "Deep Learning for Robust Iris Recognition: Introducing Synchronized Spatiotemporal Linear Discriminant Model-Iris," *Advances in Artificial Intelligence and Machine Learning; Research*, vol. 5, no. 1, pp. 3446–3464, Mar. 2025.

[7] T. Achimba, O. J. Okhuoya, R. O. Akinyede, P. A. Alabi, A. Ibrahim, and G. Ateata, "A Robust Biometric Authentication Framework for Access Control," *University of Ibadan Journal of Science and Logics in ICT Research*, vol. 13, no. 1, pp. 239–246, 2025.

[8] R. M. Hussien, K. Q. Al-Jubouri, M. A. Gburi, A. G. Hussein Qahtan, and A. H. Duaa Jaafar, "Computer Vision and Image Processing the Challenges and Opportunities for new technologies approach: A paper review," *Journal of Physics: Conference Series*, vol. 1973, no. 1, Dec. 2021, Art. no. 012002, https://doi.org/10.1088/1742-6596/1973/1/012002.

[9] G. Zhao and M. Pietikainen, "Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 915–928, Jun. 2007, https://doi.org/10.1109/TPAMI.2007.1110.

[10] T. Ahonen, A. Hadid, and M. Pietikainen, "Face Description with Local Binary Patterns: Application to Face Recognition," *IEEE Transactions*

on Pattern Analysis and Machine Intelligence, vol. 28, no. 12, pp. 2037–2041, Sep. 2006, https://doi.org/10.1109/TPAMI.2006.244.

[11] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, San Diego, CA, USA, 2005, vol. 1, pp. 886–893, https://doi.org/10.1109/CVPR.2005.177.

[12] I. T. Jolliffe and J. Cadima, "Principal component analysis: a review and recent developments," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, Apr. 2016, Art. no. 20150202, https://doi.org/10.1098/rsta.2015.0202.

[13] S. N. Mohammed and K. A. H. Alia, "Speech Emotion Recognition Using MELBP Variants of Spectrogram Image," *International Journal of Intelligent Engineering and Systems*, vol. 13, no. 5, pp. 257–266, Oct. 2020, https://doi.org/10.22266/ijies2020.1031.23.

[14] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 815–823, https://doi.org/10.1109/CVPR.2015.7298682.

[15] O. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *BMVC 2015 - Proceedings of the British Machine Vision Conference 2015*, 2015.

[16] Y. Said, M. Barr, and H. E. Ahmed, "Design of a Face Recognition System based on Convolutional Neural Network (CNN)," *Engineering, Technology & Applied Science Research*, vol. 10, no. 3, pp. 5608–5612, Jun. 2020, https://doi.org/10.48084/etasr.3490.

[17] J. Deng, J. Guo, J. Yang, N. Xue, I. Kotsia, and S. Zafeiriou, "ArcFace: Additive Angular Margin Loss for Deep Face Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 10, pp. 5962–5979, Jul. 2022, https://doi.org/10.1109/TPAMI.2021.3087709.

[18] J. Deng, J. Guo, D. Zhang, Y. Deng, X. Lu, and S. Shi, "Lightweight Face Recognition Challenge," in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, Seoul, Korea (South), Oct. 2019, pp. 2638–2646, https://doi.org/10.1109/ICCVW.2019.00322.

[19] S. S. Jasim and A. K. Abdul Hassan, "Modern Drowsiness Detection in Deep Learning: A review," *Journal of Al-Qadisiyah for Computer Science and Mathematics*, vol. 14, no. 3, Sep. 2022, https://doi.org/10.29304/jqcm.2022.14.3.1023.

[20] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "VGGFace2: A Dataset for Recognising Faces across Pose and Age," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, Xi'an, China, May 2018, pp. 67–74, https://doi.org/10.1109/FG.2018.00020.

[21] S. A. Mahmood, R. F. Ghani, and A. A. Kerim, "3D face recognition using pose invariant nose region detector," in *2014 6th Computer Science and Electronic Engineering Conference (CEEC)*, Colchester, United Kingdom, Sep. 2014, pp. 103–108, https://doi.org/10.1109/CEEC.2014.6958563.

[22] G. T. Waleed and S. H. Shaker, "Human Emotion Recognition Based on Facial Expression Using Convolution Neural Network," *Journal of Advances in Information Technology*, vol. 15, no. 12, pp. 1366–1373, 2024, https://doi.org/10.12720/jait.15.12.1366-1373.

[23] Z. H. Abbas and S. H. Shaker, "Prediction of human age based on face image using deep convolutional neural network," *AIP Conference Proceedings*, vol. 3009, no. 1, Feb. 2024, Art. no. 020013, https://doi.org/10.1063/5.0190537.

[24] W. Zhe *et al.*, "A Research on Two-Stage Facial Occlusion Recognition Algorithm based on CNN," *Engineering, Technology & Applied Science Research*, vol. 14, no. 6, pp. 18205–18212, Dec. 2024, https://doi.org/10.48084/etasr.8736.

[25] S. Chopparapu and J. B. Seventline, "An Efficient Multi-modal Facial Gesture-based Ensemble Classification and Reaction to Sound Framework for Large Video Sequences," *Engineering, Technology & Applied Science Research*, vol. 13, no. 4, pp. 11263–11270, Aug. 2023, https://doi.org/10.48084/etasr.6087.

[26] C. Lugaresi *et al.*, "MediaPipe: A Framework for Building Perception Pipelines." arXiv, Jun. 14, 2019, https://doi.org/10.48550/arXiv.1906.08172.

[27] A. W. Majeed, S. H. Shaker, and A. A. Saeid, "Understanding the techniques and methods for real-time face recognition and tracking: A comprehensive survey," *AIP Conference Proceedings*, vol. 3169, no. 1, Feb. 2025, Art. no. 030012, https://doi.org/10.1063/5.0255955.

[28] S. F. Abbas, S. H. Shaker, and F. A. Abdullatif, "Face Mask Detection Based on Deep Learning: A Review," *Journal of Soft Computing and Computer Applications*, vol. 1, no. 1, Jun. 2024, https://doi.org/10.70403/3008-1084.1006.

[29] A. Bulat and G. Tzimiropoulos, "How Far are We from Solving the 2D & 3D Face Alignment Problem? (and a Dataset of 230,000 3D Facial Landmarks)," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, Oct. 2017, pp. 1021–1030, https://doi.org/10.1109/ICCV.2017.116.

[30] R. Lateef and A. Abbas, "Tuning the Hyperparameters of the 1D CNN Model to Improve the Performance of Human Activity Recognition," *Engineering and Technology Journal*, vol. 40, no. 4, pp. 547–554, Apr. 2022, https://doi.org/10.30684/etj.v40i4.2054.

[31] S. M. Kadhim, J. Koh Siaw Paw, Y. C. Tak, and S. Ameen, "Deep Learning Models for Biometric Recognition based on Face, Finger vein, Fingerprint, and Iris: A Survey," *Journal of Smart Internet of Things*, vol. 2024, no. 1, pp. 117–157, Jun. 2024, https://doi.org/10.2478/jsiot-2024-0007.

[32] S. T. Ameen, S. Yussof, A. Ahmad, S. Khadim, and A. Hussain, "MB-ConvLSTM: a novel hybrid deep learning model for accurate sign language recognition," *International Journal of Web Information Systems*, Apr. 2025, https://doi.org/10.1108/IJWIS-10-2024-0319.

[33] C. C. Chen, Z. Liu, G. Yang, C. C. Wu, and Q. Ye, "An Improved Fault Diagnosis Using 1D-Convolutional Neural Network Model," *Electronics*, vol. 10, no. 1, Jan. 2021, Art. no. 59, https://doi.org/10.3390/electronics10010059.

[34] Y. Kim, "Convolutional Neural Networks for Sentence Classification." arXiv, Sep. 03, 2014, https://doi.org/10.48550/arXiv.1408.5882.

[35] S. Milborrow, J. Morkel, and F. Nicolls, *The MUCT Landmarked Face Database*.

[36] H. W. Ng and S. Winkler, "A data-driven approach to cleaning large face datasets," in *2014 IEEE International Conference on Image Processing (ICIP)*, Paris, France, Oct. 2014, pp. 343–347, https://doi.org/10.1109/ICIP.2014.7025068.

[37] A. K. A. Hassan and D. J. Fadhil, "Mobile Robot Path Planning Method Using Firefly Algorithm for 3D Sphere Dynamic & Partially Known Environment," *Journal of University of Babylon for Pure and Applied Sciences*, vol. 26, no. 7, pp. 309–320, May 2018, https://doi.org/10.29196/jubpas.v26i7.1506.

[38] A. K. A. Hassan and D. J. Fadhil, "Proposed Modified A* for Three-Dimensional Sphere Environment," *Al-Mansour Journal*, vol. 32, 2019.

[39] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[40] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Aug. 1997, https://doi.org/10.1162/neco.1997.9.8.1735.

[41] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, Aug. 1998, https://doi.org/10.1109/5.726791.

[42] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing & Management*, vol. 45, no. 4, pp. 427–437, Jul. 2009, https://doi.org/10.1016/j.ipm.2009.03.002.

[43] S. M. Kadhim, J. K. S. Paw, Y. C. Tak, and S. T. A. Al-Latief, "Robust Security System: A Novel Facial Recognition Optimization Using Coronavirus-Inspired Algorithm and Machine Learning," *Iraqi Journal for Computer Science and Mathematics*, vol. 6, no. 2, May 2025, https://doi.org/10.52866/2788-7421.1260.

[44] M. A. Razzaque, M. Milojevic-Jevric, A. Palade, and S. Clarke, "Middleware for Internet of Things: A Survey," *IEEE Internet of Things*

Journal, vol. 3, no. 1, pp. 70–95, Feb. 2016, https://doi.org/10.1109/JIOT.2015.2498900.

[45] E. Li, L. Zeng, Z. Zhou, and X. Chen, "Edge AI: On-Demand Accelerating Deep Neural Network Inference via Edge Computing," *IEEE Transactions on Wireless Communications*, vol. 19, no. 1, pp. 447–457, Jan. 2020, https://doi.org/10.1109/TWC.2019.2946140.

[46] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L. P. Morency, "OpenFace 2.0: Facial Behavior Analysis Toolkit," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, Xi'an, China, May 2018, pp. 59–66, https://doi.org/10.1109/FG.2018.00019.

[47] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, Jun. 2014, pp. 1701–1708, https://doi.org/10.1109/CVPR.2014.220.