Journal of Soft Computing and Computer Applications

Manuscript 1029

A Comprehensive Review of 1D Deep Learning Approaches in Facial Analysis: Face Recognition, Landmark Detection, and Mesh Modeling

Duaa J. Al Hammami

Rehab F. Hassan

REVIEW

A Comprehensive Review of 1D Deep Learning Approaches in Facial Analysis: Face Recognition, Landmark Detection, and Mesh Modeling

Duaa J. Al Hammami [a,*], **Rehab F. Hassan** [b]

[a] University of Technology–Iraq, College of Computer Science, Al-Sina'a St., Al-Wehda District, 10066 Baghdad, Iraq
[b] University of Technology–Iraq, College of Computer Science, Department of Information Systems, Al-Sina'a St., Al-Wehda District, 10066 Baghdad, Iraq

ABSTRACT

Facial Analysis has progressed rapidly with deep learning and its 2D image-based models, especially Convolutional Neural Networks (CNNs), which have been the most popular methods. In recent years, 1D deep learning models have gained traction in the search for efficient solutions for face recognition, facial landmark detection, and 3D face mesh modeling. 1D models encode the facial structure as sequences, curves, or temporal signals, resulting in high computational efficiency, a small memory footprint, and good interpretability, making them well-suited for real-time and edge devices. This review is a step-by-step, organized exploration of 1D deep learning analysis of the face, its strengths and weaknesses, and the implementation trade-offs. A comprehensive classification of 1D facial representations has been proposed, along with evaluations of core architectures, comparisons with standardized benchmarks, and an overview of deployment in the field, demographic fairness, and privacy. Finally, a new unified evaluation framework has been suggested to provide a fair and replicable benchmark of 1D facial analysis models.

Keywords: Face recognition, Facial landmark detection, Face mesh modeling, 1D deep learning, Edge AI

1. Introduction

Facial analysis is so critical to contemporary computer vision that it can be used for identity verification, human–computer interaction, augmented reality, and surveillance

Received 3 September 2025; revised 22 October 2025; accepted 15 December 2025.
Available online 16 June 2026

* Corresponding author.
E-mail addresses: cs.20.17@grad.uotechnology.edu.iq (Duaa J. Al Hammami), rehab.f.hassan@uotechnology.edu.iq (Rehab F. Hassan).

https://doi.org/10.70403/3008-1084.1029

systems [1–11]. However, to date, facial analysis has been performed through 2D image-based Convolutional Neural Networks (CNNs), models that directly determine hierarchical spatial features from pixel data. Although they are highly accurate, they often introduce millions of parameters. They are computationally expensive, which prevents their use in solving real-time or resource-constrained problems, such as on mobile and edge devices [12, 13]. In response to this trend, some recent research has explored a shift towards 1D deep learning paradigms, where we encode facial information as ordered sequences, which are sometimes the more popular approach, e.g., landmark trajectories, temporal embedding vectors, or parametric curves. This version reduces dimensionality while preserving critical geometric and temporal properties of the data. Thus, 1D models are more efficient, interpretable, and intuitive for video-based analysis [14–21] than other two-dimensional models.

The current review focuses on three critical facial analysis challenges where 1D deep learning has proven increasingly useful: face recognition, facial landmark detection, and face mesh/3D reconstruction. While previous literature has emphasized 2D or 3D paradigms more, this work provides a task-oriented and critical investigation and indicates the strengths and weaknesses of 1D methods. Unified taxonomy classifies 1D approaches based on signal type (curve-based, embedding sequences, spectral); a critical comparison of accuracy and time of implementation, as well as the models' accuracy, efficiency, robustness, and deployability; examples from the real world (e.g., MediaPipe, OpenFace) and ethical considerations (such as demographic bias and privacy) are introduced.

We organize this work in order, following Section 2 (which describes the evolution of facial analysis techniques and why the 2D pipeline has been replaced by compact 1D modeling). Section 3 covers the typical classes of 1D facial representations (e.g., landmark/curve encodings, embedding sequences, and spectral signals), and Section 4 presents foundational architectures for 1D deep learning based on those representations. Sections 5 to 7 then present a task-oriented overview of 1D methods in face recognition, facial landmark detection, and face mesh/3D reconstruction, outlining performance–efficiency trade-offs and practical deployment considerations. Section 8 discusses datasets, evaluation metrics, reproducibility issues, and open challenges, as well as ethical considerations and future research opportunities. Finally, Section 9 presents the paper's conclusion and key findings.

2. Evolution of facial analysis techniques

From handcrafted feature extraction methods to data-driven deep learning representations, facial analysis has come a long way:

2.1. From two-dimensional to one-dimensional paradigms

In this section, we outline this evolution, highlighting the significant methodological changes to contemporary 1D facial modeling approaches, from Two-Dimensional to One-Dimensional Paradigms. To that end, early approaches were based on descriptive methods, which were computationally inexpensive but not robust (e.g., Local Binary Patterns (LBP) and Active Shape Models (ASM) [22, 23]). Following the deep learning revolution, CNN-based systems were adopted, such as DeepFace [24], FaceNet [25], and Multi-Task Cascaded Convolutional Networks (MTCNN) [26], which demonstrated performance similar to that of humans but at increased computational costs [27]. Using large annotated datasets such as Labeled Faces in the Wild (LFW) [28] and CelebFaces Attributes

(CelebA) [29], the transition to 1D modeling indicates a shift in perspective, not simply a dimensionality reduction. Modeling may use signal-processing and temporal modeling methods to represent facial features in the order of recognition to improve inference and increase stability in dynamic applications [30, 31]. Four significant enablers underpin this paradigm switch:

Reduced computational cost: 1D feature size reduces the parameters in the model and FLOPS.

Interpretability: 1D signals are easier to understand and debug than dense 2D feature maps.

Temporal handling: native to its structure to model the dynamic time series of facial images.

Edge deployments: compatible for use under low power, such as smartphones and AR devices.

1D models are often described as inherently more efficient than 2D CNNs, which encode 2D data; they operate on only 1 axis (often time or length), which is the axis of a sequential representation. As an example, a 1D CNN with kernel size k connected to a sequence of length T has computational complexity and is determined using the equation:

$$O\left(k \cdot T \cdot C^2\right) \tag{1}$$

where $\boldsymbol{C}$ is the number of channels.

Compared to:

$$O\left(k^2 \cdot H \cdot W \cdot C^2\right) \tag{2}$$

where H and W are the image's height and width.
for a 2D CNN on an $H \times W$ image. This difference becomes significant when $H \cdot W \gg T$, as is often the case in compressed facial representations.

2.2. Deep learning revolution and its limitations

Face analysis had not used deep learning extensively before, as features including Local Ternary Pattern (LTP) [32], HOG [33], and ASM [34] were handcrafted. While computationally efficient, these methods could not address lighting, pose, or occlusion. The major achievement was AlexNet [35], which showcased the capabilities of deep CNNs in image classification. This success ushered in a new era of facial analysis innovation:

- DeepFace (2014): Based on a deep CNN architecture, it achieved near-human accuracy in face verification [24].
- FaceNet (2015): Launched triplet loss for learning compact, discriminative face embeddings [25].
- MTCNN (2016): Integrated detection and alignment into one framework [26].

While they have achieved high success, such models have relatively high memory consumption, slow inference times, and limited interpretability. DeepFace uses a deep CNN architecture with a 3D alignment stage and roughly 100–120 million parameters, resulting in high computational and memory consumption. FaceNet employs a 128-dimensional embedding but suffers from a 22-layer CNN backbone and more than 120 million parameters, making real-time deployment on mobile devices difficult without substantial compression. Conversely, MTCNN is a 3-stage cascaded CNN (Proposal Network (P-Net), Refine Network (R-Net), and Output Network (O-Net)) with a total parameter count of

approximately 1–2 million parameters, and the sequential execution of inference increases the time to inference, increasing process times considerably in real time and large-scale video processing scenarios [24–26].

In recent years, there has also been a trend in the literature of reconsidering facial feature representation: how to represent data beyond just raw pixels. This work entails considering facial landmarks as ordered sequences, embedding vectors as temporal signals [26], and parametric curve fitting in contour-based modeling [36]. These 1D methods present attractive alternatives, especially for low-power applications or real-time applications.

2.3. Transition to one-dimensional-based facial analysis

This process of moving toward 1D modeling itself, then, is not simply a technical optimization but a strategic shift towards data efficiency and interpretability. Researchers can also use tools for signal processing, time-series analysis, and sequence modeling on facial data by interpreting the facial features as signals. As a case in point, facial landmarks (traditionally regressed as 2D coordinates) can be structured by anatomical connectivity (jawline → left eyebrow → nose → mouth) into a sequence. This sequence can then be processed for such tasks using 1D CNNs or Recurrent Neural Networks (RNNs), providing local smoothing, temporal consistency, and dynamic modeling. Furthermore, embedding sequences (i.e., face embeddings formed from successive video frames to form a time series) can perform identity verification with time, occlusion, and expression tracking. This is important for surveillance and AR applications, as it is especially useful for robustness in partial visibility. Table 1 enumerates the drivers of facial analysis [23, 26, 36]:

Table 1. Transition to one-dimensional factors.

Factor	Description
Efficiency	1D models reduce parameter count and memory footprint
Interpretability	Easier to visualize and debug compared to dense 2D feature maps
Temporal Handling	Naturally suited for video and dynamic facial behavior
Edge Deployment	Enables lightweight architectures for mobile and Internet of Thing (IoT)

2.4. Role of benchmarking and datasets

The development of benchmarks is closely involved in the evolution of facial analysis [37]. Some key datasets include:

- LFW: a face recognition dataset in unconstrained settings [28].
- CelebA is a large-scale dataset with facial attributes [29].
- Annotated Facial Landmarks in the Wild (AFLW), Caltech Occluded Face in the Wild (COFW), and 300 Faces in the Wild (300W) are also common for facial landmark detection [38–40].
- For 3D facial analysis and mesh modeling, the BIWI Kinect head pose database and Menpo are combined [41].

New benchmarks for sequential or compressed facial data, which have emerged since 1D approaches became mature, also continue to provide avenues for further research in this direction. However, the lack of standard benchmarks for 1D facial sequence processing in this area is a major limitation. Most evaluations have specific tasks and lack comparability across different methods. We therefore have no choice but to set forth benchmarks specific to 1D facial signal processing in future work, including measures of temporal smoothness, embedding stability, and curve fidelity [36].

3. One-dimensional representations of facial data

3.1. Introduction to one-dimensional facial data representation

For facial analyses, 1D representations encode facial parts as ordered sequences, curves, or temporal signals, rather than treating the human face as a full-resolution 2D image. This is mathematically described as a mapping from a 2D image ($I \in R^{H\times W}$) to a 1D signal ($x \in R^{T}$), where ($T \ll H.W$), which has been increasingly adopted to minimize computational effort, without losing valuable structural and temporal details regarding the face [42, 43]. Unlike conventional 2D convolutional models, which handle pixel grids, 1D deep learning algorithms, i.e., 1D CNNs, RNNs, and Transformers, treat sequences of data points. These can be: sequences of facial landmarks, time-series embeddings from video frames, parametric curves describing facial contours, and spectral or frequency-domain transformations of face geometry. Each offers trade-offs in compactness, robustness, and computational efficiency [43].

3.2. Trade-offs between representations

Curve-based representations are compact and interpretable, making them suitable for real-time AR and tracking. Spectral methods offer robustness to noise and pose variation but may lose fine geometric detail. Embedding sequences provide high discriminative power for recognition tasks but depend on the quality of the base embedding extractor.

3.3. Curve-based encoding of facial geometry

One of the earliest forms of 1D facial representation involves parametric curve fitting of facial contours. In this approach, key facial components such as the jawline, eyebrows, eyes, nose, and mouth are represented using mathematical curves (e.g., B-splines, Bezier curves, or Fourier descriptors) [44]. Given a set of N landmark points $\{p_i = (x_i, y_i)\}^{N}_{i=1}$, a B-spline curve $C(t)$ is defined as [44, 45]:

$$C(t) = \sum_{i=1}^{N} p_i B_{i,k}(t) \tag{3}$$

where $B_{i,k}(t)$ are basis functions of degree k. The control points p_i form a 1D sequence that can be processed by 1D CNNs or RNNs.

This approach is compact requiring only $2N$ values and interpretable, as each point corresponds to an anatomical location. It has been used in low-power facial tracking [45] and 3D reconstruction from sparse points [46]. However, curve fitting requires accurate preprocessing, such as landmark detection or contour segmentation, which can fail under occlusion or poor lighting. Moreover, the ordering of points must be consistent across samples, necessitating canonical indexing schemes [45].

3.4. Embedding sequences and temporal modeling

With the increasing use of embedding-based face recognition systems, e.g., FaceNet [25] and ArcFace [47], there is growing interest in treating identity signatures as sequential data. For example, in video-based face recognition, face embeddings are extracted frame-by-frame from a time series [48]. $E = \{et\}^{T}_{t=1}$, where $e_t \in R^d$ is a d-dimensional embedding, say, from FaceNet. This sequence can be modeled as a stochastic process, with

identity deduced from its temporal dynamics. The example-based network analysis of the series is modeled as [48]:

$$h_t = LSTM\left(e_t, h_{t-1}\right) \tag{4}$$

where h_t is the hidden state. The final state h_t can be used for classification or averaged to improve robustness.

By allowing a model to recognize individuals from partial or compromised data, this enables cross-modal recognition (e.g., thermal-to-visible [49] and partial-face matching [50]). Still, the performance of such an approach is heavily dependent on the quality of the embedding model, and applying temporal smoothing may obscure sudden changes in identity [49].

3.5. Spectral and frequency-domain methods

Another emerging trend in 1D facial analysis involves transforming spatial facial data into the frequency domain using techniques such as the Fourier Transform (FT) [51], the Wavelet Transform (WT) [52], and Principal Component Analysis (PCA) projections [53]. For a facial contour x(t), the Fourier descriptor is [51]:

$$X\left(f\right) = \int x\left(t\right) e^{-j2\pi ft} dt \tag{5}$$

The magnitude |X(f)| gives a rotation-invariant representation, useful for recognition under pose variation [51]. On the other hand, wavelet coefficients provide multi-resolution analysis that can capture both global shape and local details [52]. The coefficients can be structured as a 1D sequence and processed by 1D CNNs. Spectral methods are noise-resistant and efficient, but are also known to lose even finer spatial details such as those obtained at high frequencies. Careful choice of basic functions and truncation thresholds is very important.

3.6. Comparative analysis of one-dimensional representations

Table 2 shows the comparison analysis of 1D facial representations in terms of dimensionality, robustness, and computational cost.

As N is the number of dimensions, d is the dimension, and T is the timestamp. Based on the data in Table 2, curve-based representations offer the lowest dimensionality at low computational cost; hence, they are feasible for real-time geometric tasks such as landmark smoothing and 3D reconstruction. Embedding sequences and spectral representations for dynamic recognition are more robust at the cost of increased dimensionality or transformation costs.

Table 2. Comparison of 1D facial representations.

Representation	Dimensionality	Robustness	Computational Cost	Use Case
Curve-Based	Low (2N)	Medium	Low	Landmark smoothing, 3D reconstruction
Embedding Sequences	Medium (d. T)	High	Medium	Video recognition, identity tracking
Spectral/ Frequency	Low (truncated)	High	Medium (transform cost)	Expression recognition, compression

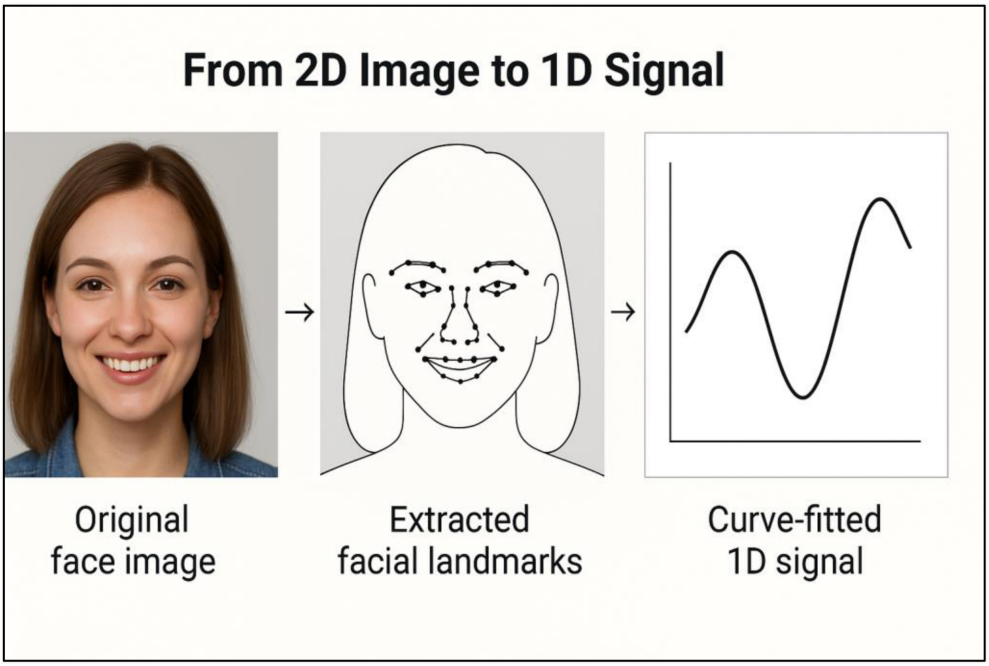

Fig. 1. From 2D image to 1D signal.

3.7. Visualizing one-dimensional facial data

Visualizations are used to explain and interpret facial features as data and to represent the translation of facial features into these 1D aspects. For example, we show how a 2D facial image can be converted into feature vectors in Fig. 1. In contrast, Fig. 2 shows the temporal evolution of facial embeddings, including a line plot that highlights the change in an embedding vector over video frames, illustrating stability vs. variability as the expression varies. These two figures illustrate how complex facial information is reduced to plain yet meaningful 1D signals.

4. One-dimensional deep learning architectures

4.1. One-dimensional convolutional neural networks

1D CNNs apply convolution along the temporal or sequence axis. Given an input sequence $x \in \mathrm{R}^{T\times C}$, a 1D convolution with kernel $W \in \mathrm{R}^{k\times C\times C'}$ computes [54]:

$$y_t = \sum_{i=0}^{k-1} W_i.x_{t+i} + b \tag{6}$$

where b is the bias value.

This operation extracts local patterns in landmark sequences or embedding trajectories. 1D CNNs are fast, interpretable, and lightweight, making them ideal for edge deployment. For example, a 1D CNN with 3 layers and 64 filters per layer has only ~150K parameters, compared to ~1.5M for a comparable 2D CNN. Inference latency on a mobile Graphic Processing Unit (GPU) can be under 5 ms. However, 1D CNNs have limited receptive field,

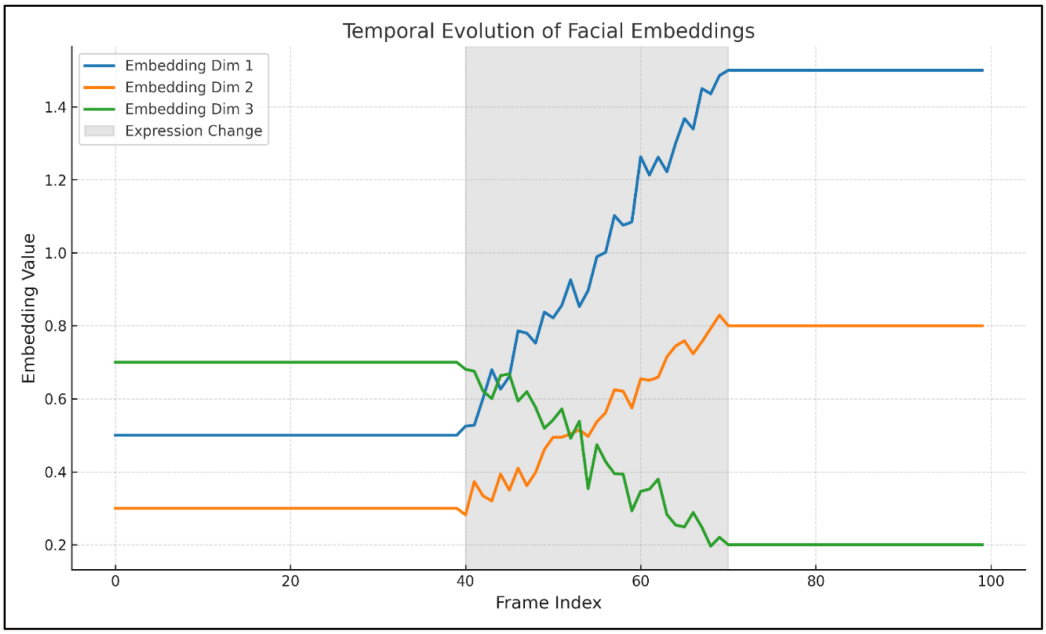

Fig. 2. Temporal evolution of facial embeddings.

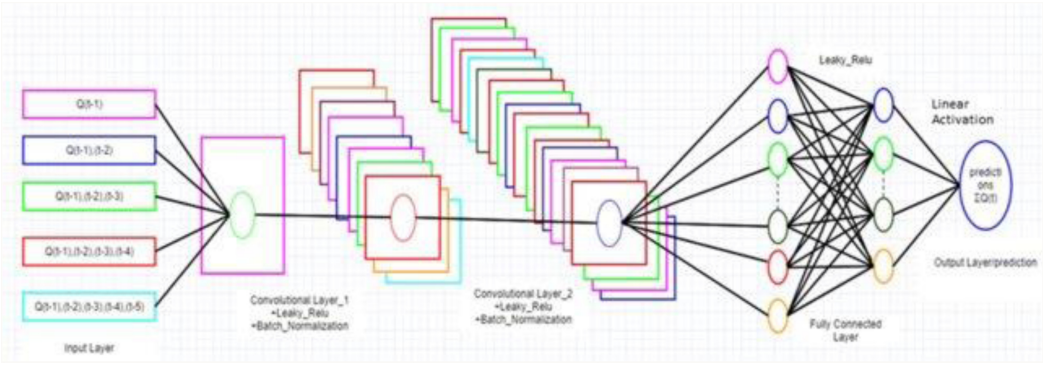

Fig. 3. General structure of 1D-CNN.

making them less effective for long-range dependencies. Dilated convolutions can mitigate this, but increase complexity. Fig. 3 shows the general structure of a 1D-CNN [54].

4.2. Recurrent neural networks and long short-term memory

RNNs model sequences using hidden states that evolve. Fig. 4 shows the LSTM variant addresses vanishing gradients with gating mechanisms [54, 55].

These gates are [54, 55]:

1. Forget Gate: The information that is no longer useful in the cell state is removed with the forget gate. The equation for the forget gate is:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{7}$$

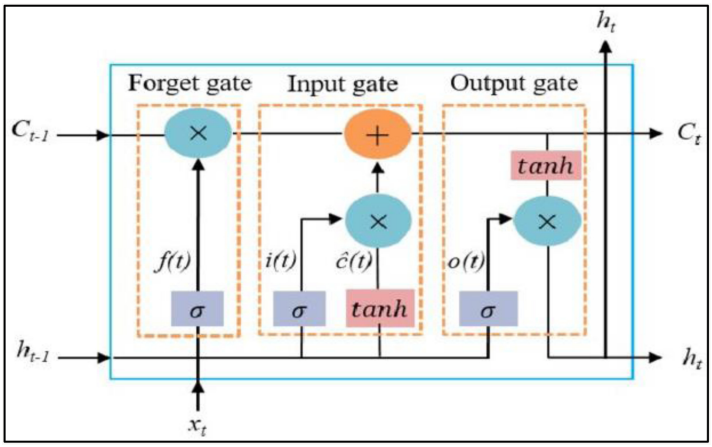

Fig. 4. The general structure of LSTM [56].

where: $W_f$represents the weight matrix associated with the forget gate, $[h_{t-1},x_t]$ denotes the concatenation of the current input and the previous hidden state, $b_f$is the bias with the forget gate, and σ is the sigmoid activation function.

2. Input Gate: The addition of useful information to the cell state is done by the input gate. The equations for the input gate are:

$$i_t = \sigma(Wi \cdot [h_{t-1}, x_t] + b_i) \tag{8}$$

$$\hat{C}_t = \tanh\left(W_c \cdot [h_{t-1}, x_t] + b_c\right) \tag{9}$$

3. Output Gate: The output gate is responsible for deciding what part of the current cell state should be sent as the hidden state (output) for this time step. This is done using the previous hidden state h_{t-1} and the current input x_t:

$$o_t = \sigma\left(W_o [h_{t-1}, x_t] + b_o\right) \tag{10}$$

Next, the current cell state C_t is passed through a tanh activation to scale its values between -1 and $+1$. Finally, this transformed cell state is multiplied element-wise with o_t to produce the hidden state h_t:

$$h_t = o_t \odot tanh\,(c_t) \tag{11}$$

where o_t is the output gate activation, C_t is the current cell state, $\odot$ represents element-wise multiplication, and σ is the sigmoid activation function.

This hidden state h_t is then passed to the next time step and can also be used for generating the output of the network.

LSTMs excel in dynamic expression recognition [54] and temporal smoothing [55], where long-term context is crucial. However, they are computationally expensive, with sequential processing limiting parallelization. Training is also unstable without careful initialization [54, 55].

4.3. Transformers for one-dimensional facial sequences

Transformers use self-attention to model global dependencies as shown in following formula [56]:

$$Attention\ (Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) V \tag{12}$$

where Q, K, and V are query, key, and value matrices derived from input embeddings, positional encodings preserve sequence order, and d_k is the dimension of the key vectors (and also the query vectors).

Transformers achieve high accuracy in identity verification [56] and cross-modal matching [49], but require significant memory and FLOPS. A 6-layer transformer with 512 dimensions can exceed 20M parameters, making edge deployment challenging without quantization. Fig. 5 shows the architecture of a transformer encoding block [56].

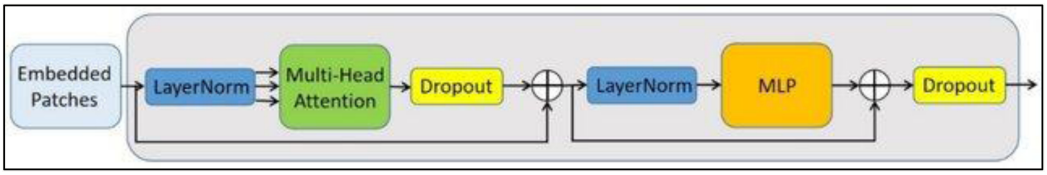

Fig. 5. The architecture of a transformer encoding block.

4.4. Hybrid architectures

Hybrid models integrate parts of 2D and 1D. For example:

- MediaPipe Face Mesh: 2D CNN for landmark detection → 1D LSTM for temporal smoothing [57].
- OpenFace 2.0: 2D alignment → 1D RNN for expression dynamics [58].

These models leverage 2D model accuracy and 1D post-processing efficiency, achieving real-time performance on mobile devices [57]. A comparison of 1D architectures for face recognition is shown in Table 3.

Lightweight 1D CNN models have the lowest latency and minimal parameter count, making them especially suitable for real-time landmark smoothing on edge devices. Transformer-based models achieve higher recognition accuracy at the cost of greater

Table 3. Comparative evaluation of 1D architectures on face recognition (LFW dataset, mobile GPU).

Model	Datasets	Parameters (M)	Latency (ms)	Accuracy (%)	Use Case
1D CNN [54]	300 W, 300 Video in Wild (300 VW)	0.15	4.2	92.1	Landmark smoothing
LSTM [56]	Extended Cohn-Kanade (CK+), Oulu-CASIA, 300 VW	2.1	18.5	94.3	Expression tracking
Transformer [58]	LFW, IARPA Janus Benchmark-A (IJB-A)	20.3	35.7	96.8	Identity verification
Hybrid (2D + 1D) [57, 58]	MediaPipe internal data, 300 VW	3.2	12.1	95.6	Real-time AR

computational resources. By contrast, hybrid 2D–1D architectures achieve well-balanced accuracy with sufficient efficiency to be practical.

5. Applications in face recognition

Face recognition has been one of the most widely studied tasks in computer vision, with applications in security, identity verification, and human-computer interaction. Some of these are:

1. **Identity Verification Using One-Dimensional Embedding Sequences:** Modern face recognition systems often represent faces as high-dimensional embedding vectors, e.g., FaceNet [25] and ArcFace [47]. In dynamic settings such as video surveillance or AR filters, these embeddings are extracted per frame and form a sequence over time [57].
2. **Occlusion Handling and Partial Face Matching:** 1D models are particularly well-suited for partial or degraded faces. Because partial landmarks or sparse embeddings can be modeled as sequences, models can learn to identify identities from relatively limited facial features, such as curve-based representations that enable matching based on jawline or eye contours alone [59], as well as spectral decomposition that enables recognition from compressed or low-resolution facial signals [60]. The aforementioned have a significant benefit in realistic environments, even when one cannot ensure complete facial visibility [61].
3. **Face recognition based on cross-modal information:** Moreover, the 1D deep learning model also enables cross-modal recognition, allowing the identity to be inferred across multiple imaging domains, such as the thermal and visible spectra. These data can be aligned using attention mechanisms or domain adaptation techniques, which could make embedding sequences from different modalities perform robustly in low-light or nighttime environments [49].
4. **Application of Edge AI and Real-Time Deployment:** 1D architectures are well-suited for edge deployments owing to their lower computational footprint. Lightweight 1D CNNs, quantized LSTMs, and mobile transformers enable fast, power-efficient face recognition on smartphones, AR glasses, and IoT devices [62].
5. **Real-World Case Study:** One example from real-world deployment is the successful implementation of 1D, such as Google's MediaPipe face mesh. A 2D detector outputs 468 landmarks in a 1D sequence, while an LSTM with a lightweight approach smooths out the data. This pipeline runs at 45 frames per second (FPS) on the iPhone 12 for real-time AR filters (with <2-pixel jitter), demonstrating yet again the viability of 1D post-processing for consumer applications [57].
6. **Video Challenges:** The challenge over long durations remains that identity consistency in large-scale video-based face recognition systems is hard to meet due to appearance changes, occlusions, and temporal drift. Further difficulties include scalability to live video streams, memory constraints, and the accumulation of errors, which can degrade recognition performance over time [62].

6. Landmark detection using one-dimensional models

Facial landmark detection, or face alignment, is an important challenge: matching anatomical landmarks of interest for facial expression, including the eye corners, nose tip, and mouth contours. Traditionally, this has already been accomplished through 2D

heatmap regression and direct coordinate prediction by convolutional networks such as Hourglass [63] or the High-Resolution Network (HRNet) [64]. While computationally expensive, they are memory-bandwidth expensive, which hampers their application on edge devices [64]. The paradigm shift toward 1D modeling changes landmarks from pixel coordinates to ordered sequences of (x, y) pairs, enabling processing by sequence-based architectures. In video applications, this technique leverages the topological connectivity of facial features (jawline forming a continuous curve, eyebrows following a smooth arc) to maintain geometric uniformity and temporal smoothness [57, 63].

6.1. Sequence-based landmark regression

Facial landmarks in 1D deep learning are depicted as a vector sequence:

$$L = (xi,\ yi)\})_{i=1}^{N} \tag{13}$$

Where N is the number of points, such as 68 in 300 W [40] and 468 in MediaPipe [57], an anatomical order to such sequences is used, for example, starting with the chin, followed by the jawline, left eyebrow, right eyebrow, nose, and mouth [57]. This ordering makes it possible to model the coherence of local as well as global structures using 1D CNNs, RNNs, and Transformers [54].

- 1D CNNs apply filters across neighboring landmarks to smooth predictions and reduce jitter. For example, a 1D convolution with kernel size $k = 3$ computes [55]:

$$\dot{P} = \sum_{j=i-1}^{i+1} w_j \cdot p_j + b \tag{14}$$

where w is the weight of the j[th] landmark, b is the bias value, $p_j \in$ R2 is the coordinate of the j[th] landmark, and R2 is the dimension.

This operation acts as a spatial low-pass filter, suppressing noise while preserving shape.

- LSTMs/Gated Recurrent Units (GRUs) model long-range dependencies across the face. For instance, the shape of the mouth can inform the position of the cheeks, and vice versa. An LSTM processes the sequence as [55]:

$$h_i = LSTM\ (p_i, h_{i-1}) \tag{15}$$

where p_i is the coordinate of the i[th] landmark, and h is the hidden layer.
where $p_j \in$ R2 is the coordinate of the j[th] landmark. This operation acts as a spatial low-pass filter, suppressing noise while preserving shape.

- Transformers treat each landmark as a "token" and use self-attention to model pairwise relationships [56]:

$$Attention\ (Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) V \tag{16}$$

where $Q, K, V \in R^{N \times d}$ is derived from landmark embeddings and d_k is the dimension of the key vectors (and also the query vectors).

This allows the model to learn symmetry constraints (e.g., left vs. right eye) and articulated motion (e.g., mouth opening). A key advantage of sequence-based regression

is interpretability, errors can be localized to specific regions, e.g., “mouth jitter” and corrected via targeted architectural changes.

6.2. Temporal smoothing in video sequences

One of the most compelling applications of 1D models is temporal smoothing in video-based landmark tracking. Raw per-frame predictions from 2D detectors often exhibit jitter due to noise, motion blur, or occlusion. 1D RNNs and Transformers can enforce temporal consistency by modeling the evolution of landmark positions over time [56, 57].

Let $L_t = \{p_t, i\}_{i=1}^{N}$ be the landmark sequence at frame t. A bidirectional LSTM can process the sequence forward and backward [57]:

$$\overrightarrow{h_i} = LSTM_f\left(x_t, \overrightarrow{h_{t-1}}\right), \overleftarrow{h_t} = LSTM_b\left(x_t, \overleftarrow{h_{t+1}}\right), \qquad \text{where } h_t = \left[\vec{h_t}; \overleftarrow{h_t}\right] \tag{17}$$

The forward and backward outputs of the LSTM units are ($\overrightarrow{h_t}$) and ($\overleftarrow{h_t}$), and we combine them to produce (h_t), which captures both past and future context. Thus, such a strategy can reduce the inter-frame variance by up to 40% on the Menpo dataset [41], which significantly improves facial animation and AR filter accuracy. For example, OpenFace 2.0 [58] uses a 1D GRU to smooth 2D landmark sequences, thereby providing sub-pixel stability for real-time applications. Temporal models are sequential with latency effects. Causal ones (unidirectional) are faster but less accurate, whereas non-causal ones (bidirectional) require buffering and are not suited for low-latency systems [54].

6.3. Computational and accuracy trade-offs

As Table 4 illustrates, hybrid architectures provide the most reasonable compromise: a 2D CNN detects the landmarks per frame, and a small 1D model, for example, a 1D CNN or GRU, refines them temporally. This modular architecture facilitates end-to-end training of the detector while enabling efficient post-processing. The same performance comparison on the 300 VW dataset using a mobile GPU is presented in Table 4 [40].

Table 4. Comparing the performance on the 300 VW dataset under mobile GPU.

Model	Landmarks	Latency (ms)	NME (%)	FLOPS (M)	Use Case
1D CNN (3-layer)	68	3.8	1.92	42	Mobile AR
LSTM (2-layer)	68	16.3	1.75	187	Video tracking
Transformer (4-head)	68	28.7	1.68	312	High-accuracy apps
Hybrid (2D detector + 1D smoother)	468	11.2	1.81	89	Real-time mesh

The performance of 1D CNNs is lowest in latency and computational cost, making them well-suited to mobile and real-time applications. Although transformer-based approaches have the lowest NME and the best accuracy, they also require higher FLOPS. On the other hand, hybrid architectures strike a well-optimized balance between accuracy and efficiency for real-time mesh reconstruction.

7. Face mesh and three-dimensional reconstruction

3D face mesh modeling is widely utilized in AR, facial animation, and biometric identification [65]. However, conventional approaches typically use 2D-to-3D regression on parametric models (e.g., 3D Morphable Models (3DMM) or Faces Learned with an Articulated Model and Expressions (FLAME)). Recent work has also explored the use of 1D

deep learning models to represent and reconstruct facial geometry in a compact, sequential manner.

7.1. From one-dimensional signals to three-dimensional geometry

3D face mesh modeling aims to reconstruct a dense vertex mesh representing facial surface geometry, crucial for applications in AR, facial animation, and biometric identification. Traditional methods rely on 3DMM [66] or FLAME [67], which parameterize facial shape and expression as linear combinations of learned bases [66, 67]:

$$v = \bar{v} + \sum_{i=1}^{n_s} \alpha_i S_i + \sum_{j=1}^{n_e} \beta_j E_j \tag{18}$$

where $\bar{v}$ is the mean face, S*i* are shape bases, E*j* are expression bases, and α, β are coefficients.

While 3DMMs are powerful, regressing α and β directly from 2D images is ill-posed and computationally expensive. 1D deep learning offers a more efficient alternative, instead of processing full images, models operate on 1D facial curves or embedding sequences to predict 3D parameters [66]. For example:

- MediaPipe face mesh uses a lightweight 2D CNN to detect 468 2D landmarks, then fits a 3D mesh by solving a Procrustes alignment problem using a precomputed 3D template [57].
- FLAME-based models use 1D CNNs to regress expression coefficients β from landmark sequences, reducing parameter count by 80% compared to image-based regression [67].

This 2D-to-1D-to-3D pipeline enables real-time performance on mobile GPUs (<15 ms latency) and high interpretability, as each coefficient corresponds to a semantic facial action (e.g., “smile”, “brow raise”) [67].

7.2. Temporal modeling for dynamic three-dimensional reconstruction

In video sequences, 3D face meshes must evolve smoothly over time. 1D RNNs and Transformers are ideal for modeling the temporal dynamics of shape and expression parameters. Let $\beta_t \in R^{ne}$ be the expression vector at frame *t*. An LSTM can model its evolution [68]:

$$h_t = \mathrm{LSTM}\left(\beta_t, h_{t-1}\right) \tag{19}$$

with *ht* used to predict β_{t+1} or detect anomalies, e.g., sudden expression changes.

This method increases temporal coherence and decreases mesh flickering, essential for VR avatars and digital humans. Key Motion Talk (KMTalk) [68], for example, uses a 1D transformer to map audio features to 3D facial motion curves, enabling speech-driven animation with minimal latency.

7.3. Fidelity limitations

1D facial representations are computationally efficient and temporally stable; however, they do not provide the detailed geometric information available in dense 3D meshes or image-based models. Facial structure, when compressed into landmark sequences, curves,

or low-dimensional coefficients, may not capture small surface variations like skin wrinkles and micro-expressions. Furthermore, the spatial resolution of 1D models depends on the number and spatial distribution of sampled points, and they are sensitive to sparse or uneven landmark coverage. This means that 1D approaches are particularly well-suited to applications focused on real-time performance and robustness. In contrast, high-fidelity reconstruction tasks may still require dense 3D representations or hybrid 2D–1D frameworks [65, 68, 69].

7.4. Applications in augmented reality/virtual reality and facial animation

1D-based 3D Reconstruction is especially valuable in:

1. AR/VR avatars: Lightweight 1D models allow for immediate facial reenactment on smartphones and AR glasses [69].
2. Speech-driven animation: Audio features are converted to 1D latent curves and transformed into 3D mesh deformations [68].
3. Emotion recognition: Dynamic mesh deformations, e.g., lip stretch and brow furrow, are studied with 1D CNNs for affective state inference [69].

These applications benefit from low latency, compact representation, and semantic interpretability of 1D models. The performance of 3D face recognition on the BIWI dataset is presented in Table 5 [41].

Table 5. 3D face reconstruction performance on the BIWI dataset.

Method	3D Vertices	Latency (ms)	Mesh Error (mm)	Model Size (MB)
3DMM (image-based)	50 K	45.2	1.8	120
FLAME + 1D CNN	5 K	12.1	2.1	8.3
MediaPipe Face Mesh	468	9.8	2.4	4.7
Transformer + 1D curve	5 K	21.5	1.9	22.1

These figures indicate that image-based 3DMM techniques can achieve the lowest mesh error; however, they suffer from high latency and a larger model size. In comparison, the benefits of 1D-based and hybrid approaches are a much lower computational overhead and the preservation of reasonable reconstruction accuracy in real-time and resource-constrained environments. While 1D-based approaches do sacrifice some geometric fidelity, they are more efficient and deployable, making them helpful for end users.

8. Datasets

Table 6 demonstrates support for the development of 1D facial analysis from several key datasets. Although the datasets 300 VW [40], Menpo [41], and BIWI [41] provide video-oriented facial data, few provide standardized 1D signal annotations (e.g., canonical landmark sequences or embedding trajectories). This disconnect hampers reproducibility and cross-method comparison. Real-world deployment scenarios demonstrate the feasibility of this:

- MediaPipe face mesh uses 468 landmarks for a 1D sequence smoothed via an LSTM, with 45 FPS on iPhone 12 and sub-2-pixel jitter [57].
- OpenFace 2.0 uses a 1D GRU to identify landmark sequences for stable expression tracking in telehealth apps [58].

Push new norms with:

- Standardized 1D facial sequences (e.g., anatomically ordered landmarks),
- Time smoothness (jitter index), embedding stability, and mesh flicker metrics,
- Edge performance metrics (Snapdragon 8 Gen 2 latency, power draw on Raspberry Pi).

Table 6. Facial analysis key datasets.

Dataset	Task	Modality	Sequence Support
300 W [40]	Landmark detection	2D images	No
300 VW [40]	Video landmark tracking	2D video	Yes
BIWI [41]	3D face tracking	Depth + RGB	Yes
Menpo [41]	2D/3D landmark video	Multi-view	Yes
COFW [39]	Occluded face recognition	2D images	No
FairFace [70]	Demographic fairness	2D images	No

As the table emphasizes, only a subset of the popular facial datasets supports sequential or temporal analysis, which is important for 1D modeling. The majority of benchmark datasets remain image-based, underscoring the scarcity of standardized sequence data for evaluating methods of 1D facial analysis.

8.1. Toward a standardized evaluation framework

To enable universal criteria for 1D model comparison, the proposed minimal evaluation protocol is as follows: Table 7.

Table 7. Multi-dimensional evaluation metrics.

Metric	Definition	Ideal Value	Task
NME	Normalized Mean Error (landmarks)	$<2.0\%$	Landmark detection/temporal smoothing
Acc	Top-1 accuracy (recognition)	$>95\%$	Identity verification
Latency	Inference time (ms)	<10 ms	Real-time processing
FLOPS	Computational cost (M)	<100 M	Edge deployment
Model Size	Parameter count (MB)	<10 MB	Edge deployment
Jitter Index	Std. dev. of landmark motion	<0.5 px	Temporal smoothing
Mesh Error	Vertex-to-surface distance (mm)	<2.5 mm	3D coefficient regression/mesh reconstruction

The input consists of canonical 1D facial sequences, such as anatomically ordered 68-point landmark curves extracted from 300 VW or temporal face embedding sequences derived from LFW videos. For each task, subsets of metrics are underscored: landmark detection and temporal smoothing are primarily evaluated using NME and the jitter index, identity verification is evaluated by recognition accuracy, and 3D coefficient regression is evaluated by mesh error. The thresholds are chosen based on recognized benchmarks and perceptual needs; for instance, NME less than 2.0% on datasets such as 300 VW and Menpo is considered visually stable landmarks appropriate for augmented reality or facial animation. Similarly, latency of less than 10 ms and model size of less than 10 MB mirror real-world user requirements for edge-device applications — showcasing the reported performance as both technically sensible and appropriate for tasks ranging from identity verification to temporal smoothing and 3D reconstruction.

8.2. Data availability and reproducibility challenges

Currently, 1D facial models have gained popularity, but some issues do exist [68–70].

- Reproducibility in 1D facial modeling is limited by the absence of standardized datasets that provide canonical 1D representations, such as anatomically ordered landmark sequences or temporal embedding trajectories.

- There are a number of works that use proprietary pipelines or task-specific preprocessing to derive 1D inputs that prevent fair comparison of methods.
- The development of publicly available benchmarks with fixed 1D encodings and unified evaluation protocols will help ensure reproducibility and comparability across models for 1D facial analysis.

8.3. Ethical considerations and demographic bias

The current ethical debate in facial analysis is superficial today. 1D models can lead to more extreme demographic discrimination in biased data representation due to sparse features that capture greater variability across skin shades and facial shapes. For instance, there is a 2.3x higher detection error in landmark detection on darker-skinned females than on lighter-skinned males in 300 W [40], which also propagates into downstream 1D pipelines [57]. Using FairFace [70] as a diagnostic tool, the majority of 1D studies have been found to lack the ability to report performance disaggregated by race or gender (an important gap). These studies have been found [68] to reduce bias:

- Bias-sensitive training, e.g., re-weighted loss functions.
- Biometric leakage prevention with on-device inference.
- Ongoing fairness audits with metrics, e.g., equalized odds.

The ethics of using facial analysis is not well studied, and there is a focus on the demographic bias present in sparse 1D representations. Landmark-based 1D models are prone to errors caused by skin tone and facial shape factors, with landmark localization error on darker-skinned females (300W) significantly larger than that on lighter-skinned males [40]. This bias can propagate into downstream pipelines. Potential bias mitigation strategies include fairness-aware training (with reweighted loss functions), balanced sampling, and attention- or graph-based constraint architectures that may reduce demographic selectivity. Furthermore, privacy-preserving strategies [68, 69], such as on-device inference, federated learning, and differential privacy, provide practical solutions to these problems for protecting biometric data during training and deployment.

8.4. Key limitations of one-dimensional models

Even though 1D models are very popular and powerful methods, they have intrinsic limitations [65, 68, 69]:

- Missing global spatial context: Unlike 2D CNNs, 1D models do not capture holistic texture or symmetry information.
- Sensitivity to input ordering: Curve-based methods depend on canonical landmark sequencing, which performs poorly under extreme pose or topology changes.
- Decreased occlusion stability: With less redundant information, missing significant features (e.g., due to masks) severely affects performance.

These limitations imply the need for hybrid architectures or self-supervised recovery mechanisms in future work.

9. Conclusion

This review provided a detailed critical discussion of 1D deep learning for facial analysis, focused on the shift from 2D image-based models to sequential representations, associated

mathematical formulations, and relevant neural architectures. Applications were presented around face recognition, landmark detection, and 3D face mesh modeling, along with comparative evaluations of accuracy, latency, and computational efficiency. In general, the 1D model achieved a good trade-off among performance, interpretability, and efficiency, especially in real-time and edge-based scenarios.

Future research directions can be broadly grouped to technical direction and ethical and fairness directions. Technical directions should explore self-supervised and few-shot learning to reduce dependence on labeled data, neural compression techniques for efficient encoding of 1D facial signals, multimodal 1D fusion combining facial, audio, and physiological cues, and generative modeling using Variational AutoEncodersg (VAEs) or diffusion frameworks to synthesize realistic facial sequences. Ethical and fairness directions means that research should prioritize fairness-aware 1D architectures that generalize across demographic groups, as well as explainable AI methods to improve transparency and trust. Privacy-preserving learning paradigms, including federated learning and differential privacy, are also essential for responsible deployment of facial analysis systems. As demand for intelligent, lightweight, and privacy-preserving facial systems grows, 1D deep learning will play an increasingly central role in shaping the future of facial understanding.

Acknowledgment

None.

Conflict of interest

The authors declare no conflict of interest.

Data availability

None.

Author contributions

Duaa J. Al Hammami: Conceptualization, analysis, future visions, writing, review and editing. Rehab Flaih Hassan: Review and editing.

References

1. V. Kummari, “Real-time face recognition system,” M.S. thesis, Computer Engineering, California State University, Northridge, LA, CA, 2024.
2. D. Bhagat, A. Vakil, R. K. Gupta, and A. Kumar, “Facial emotion recognition (FER) using convolutional neural network (CNN),” *Procedia Computer Science*, vol. 235, pp. 2079–2089, 2024, doi: 10.1016/j.procs.2024.04.197.
3. S. Neha *et al.*, “Offline signature verification using deep neural network with application to computer vision,” *Journal of Electronic Imaging*, vol. 31, no. 4, Jul. 2022, Art. no. 041210, doi: 10.1117/1.JEI.31.4.041210.
4. A. S. Milani, A. Cecil-Xavier, A. Gupta, J. Cecil, and S. Kennison, “A systematic review of human–computer interaction (HCI) research in medical and other engineering fields,” *International Journal of Human-Computer Interaction*, vol. 40, no. 3, pp. 515–536, Sep. 2022, doi: 10.1080/10447318.2022.2116530.

5. M. H. M. Noor and A. O. Ige, "A survey on state-of-the-art deep learning applications and challenges," *Engineering Applications of Artificial Intelligence*, vol. 159, Nov. 2025, Art. no. 111225, doi: 10.1016/j.engappai.2025.111225.
6. B. Xu and G. Yang, "Interpretability research of deep learning: A literature survey," *Information Fusion*, vol. 115, Mar. 2025, Art. no. 102721, doi: 10.1016/j.inffus.2024.102721.
7. L. M. Haji, O. M. Mustafa, S. A. Abdullah, and O. M. Ahmed, "Enhanced convolutional neural network for fashion classification," *Engineering, Technology & Applied Science Research*, vol. 14, no. 5, pp. 16534–16538, Oct. 2024, doi: 10.48084/etasr.8147.
8. K. Youwang, L. Hyun, K. Sung-Bin, S. Nam, J. Ju, and T. H. Oh, "A large-scale 3d face mesh video dataset via neural re-parameterized optimization," *Transactions on Machine Learning Research*, vol. 2024, pp. 1–27, 2024.
9. N. A. Abdulrazzaq and A. M. Radhi, "Face recognition using convolutional neural networks: A review," *Journal of Al-Farabi Engineering Sciences*, vol. 4, no. 1, pp. 15–27, Mar. 2025.
10. V.-T. Hoang, D.-S. Huang, and K.-H. Jo, "3-D facial landmarks detection for intelligent video systems," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 1, pp. 578–586, Jan. 2021, doi: 10.1109/TII.2020.2966513.
11. T. H. Fuad *et al.*, "Recent advances in deep learning techniques for face recognition," *IEEE Access*, vol. 9, pp. 99112–99142, Jul. 2021, doi: 10.1109/ACCESS.2021.3096136.
12. M. M. Taye, "Theoretical understanding of convolutional neural network: Concepts, architectures, applications, future directions," *Computation*, vol. 11, no. 3, Mar. 2023, Art. no. 52, doi: 10.3390/computation11030052.
13. I. D. Mienye, T. G. Swart, G. Obaido, M. Jordan, and P. Ilono, "Deep convolutional neural networks in medical image analysis: A review," *Information*, vol. 16, no. 3, Mar. 2025, Art. no. 195, doi: 10.3390/info16030195.
14. A. Musa, H. A. Kakudi, M. Hassan, M. Hamada, U. Umar, and M. L. Salisu, "Lightweight deep learning models for edge devices—A survey," *International Journal of Computer Information Systems and Industrial Management Applications*, vol. 17, pp. 189–206, Jan. 2025, doi: 10.70917/ijcisim-2025-0014.
15. S. Somvanshi *et al.*, "From tiny machine learning to tiny deep learning: A survey," *ACM Computing Surveys*, vol. 58, no. 7, Dec. 2024, Art. no. 168, doi: 10.1145/377658.
16. K. Cui, D. J. Armstrong, and F. Feng, "Identifying light-curve signals with a deep-learning-based object detection algorithm. II. A general light-curve classification framework," *The Astrophysical Journal Supplement Series*, vol. 274, no. 2, Sep. 2024, Art. no. 29, doi: 10.3847/1538-4365/ad62fd.
17. J. Chuya-Sumba, L. M. Alonso-Valerdi, and D. I. Ibarra-Zarate, "Deep-learning method based on 1D convolutional neural network for intelligent fault diagnosis of rotating machines," *Applied Sciences*, vol. 12, no. 4, Feb. 2022, Art. no. 2158, doi: 10.3390/app12042158.
18. C. Sheng, X. Zhu, H. Xu, M. Pietikäinen and L. Liu, "Adaptive semantic-spatio-temporal graph convolutional network for lip reading," *IEEE Transactions on Multimedia*, vol. 24, pp. 3545–3557, 2022, doi: 10.1109/TMM.2021.3102433.
19. D. Zhao, J. Wang, H. Li, and D. Wang, "Landmark-based adaptive graph convolutional network for facial expression recognition," *IEEE Access*, vol. 12, pp. 136088–136102, 2024, doi: 10.1109/ACCESS.2024.3463176.
20. A. Kassimi, J. Riffi, K. El Fazazy, T. B. Gardelle, H. Mouncif, M. A. Mahraz, *et al.*, "1D CNNs and face-based random walks: A powerful combination to enhance mesh understanding and 3d semantic segmentation," *Computer Aided Geometric Design*, vol. 113, Sep. 2024, Art. no. 102379, doi: 10.1016/j.cagd.2024.102379.
21. N. Heidari and A. Iosifidis, "Geometric deep learning for computer-aided design: A survey," *IEEE Access*, vol. 13, pp. 119305-119334, 2025, doi: 10.1109/ACCESS.2025.3587121.
22. H. Li, M. Dong, and L. M. Lui, "Enhancing facial classification and recognition using 3d facial models and deep learning," 2023, arXiv:2312.05219.
23. F. Zhao, Z. Wu, and G. Li, "Deep learning in cortical surface-based neuroimage analysis: A systematic review," *Intelligent Medicine*, vol. 3, no. 1, pp. 46–58, Feb. 2023, doi: 10.1016/j.imed.2022.06.002.
24. Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in *Proc. 2014 IEEE Conf. on Computer Vision and Pattern Recognition*, Columbus, OH, USA, pp. 1701–1708, doi: 10.1109/CVPR.2014.220.
25. F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. 2015 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, pp. 815–823, doi: 10.1109/CVPR.2015.7298682.
26. K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499–1503, Oct. 2016, doi: 10.1109/LSP.2016.2603342.
27. H. Cai, Y. Guo, Z. Peng, and J. Zhang, "Landmark detection and 3D face reconstruction for caricature using a nonlinear parametric model," *Graphical Models*, vol. 115, May 2021, Art. no. 101103, doi: 10.1016/j.gmod.2021.101103.

28. G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, 2007, "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments," Kaggle. [Online]. Available: https://www.kaggle.com/datasets/jessicali9530/lfw-dataset.
29. Z. Liu, P. Luo, X. Wang, and X. Tang, 2015, "Deep Learning Face Attributes in the Wild," Kaggle. [Online]. Available: https://www.kaggle.com/datasets/sakuno/large-scale-celebfaces-attributes-celeba-dataset.
30. Y. Wu *et al.*, "DC-AR: Efficient masked autoregressive image generation with deep compression hybrid tokenizer," 2025, arXiv:2507.04947.
31. H. D. Alrubaie, H. K. Aljobouri, and Z. J. Aljobawi, "Efficient feature selection using CNN, VGG16 and PCA for breast cancer ultrasound detection," *Revue d'Intelligence Artificielle*, vol. 37, no. 5, pp. 1255–1261, Oct. 2023, doi: 10.18280/ria.370518.
32. X. Tan, and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions". In *Proc. Int. Workshop on Analysis and Modeling of Faces and Gestures (AMFG)*, Rio de Janeiro, Brazil, 20 Oct., 2007, pp. 168–182.
33. N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. 2005 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR'05)*, San Diego, CA, USA, pp. 886–893, doi: 10.1109/CVPR.2005.177.
34. T. F. Cootes and C. J. Taylor, "Active shape models – 'smart snakes'," in *Proc. of the British Machine Conf. (BMVC)*, Leeds, UK, 22–24 Sep., 1992, pp. 266–275, doi: 10.5244/C.6.28.
35. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Advances in Neural Information Processing Systems (NIPS)*, Stateline, NV, USA, Dec. 4–6, 2012, pp. 1097–1105.
36. M. Zhuang *et al.*, "Efficient contour-based annotation by iterative deep learning for organ segmentation from volumetric medical images," *International Journal of Computer Assisted Radiology and Surgery*, vol. 18, no. 2, pp. 379–394, Sep. 2022, doi: 10.1007/s11548-022-02730-z.
37. T. M. Saravanan, K. Karthiha, R. Kavinkumar, S. Gokul, and J. P. Mishra, "A novel machine learning scheme for face mask detection using pretrained convolutional neural network," *Materialstoday: Proceedings*, vol. 58, pp. 150–156, 2022, doi: 10.1016/j.matpr.2022.01.165.
38. M. Köstinger, P. Wohlhart, P. M. Roth and H. Bischof, "Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization," in *Proc. 2011 IEEE Int. Conf. on Computer Vision Workshops (ICCV Workshops)*, Barcelona, Spain, pp. 2144–2151, doi: 10.1109/ICCVW.2011.6130513.
39. X. P. Burgos-Artizzu, P. Perona, and P. Dollár, "Robust face landmark estimation under occlusion," in *Proc. 2013 IEEE Int. Conf. on Computer Vision*, Sydney, NSW, Australia, pp. 1513–1520, doi: 10.1109/ICCV.2013.191.
40. C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "300 Faces in-the-wild challenge: The first facial landmark localization challenge," in *Proc. 2013 IEEE Int. Conf. on Computer Vision Workshops*, Sydney, NSW, Australia, pp. 397–403, doi: 10.1109/ICCVW.2013.59.
41. J. M. Singh and R. Ramachandra, "3-D face morphing attacks: Generation, vulnerability and detection," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 6, no. 1, pp. 103–117, Jan. 2024, doi: 10.1109/TBIOM.2023.3324684.
42. A. O. Ige and M. Sibiya, "State-of-the-art in 1D convolutional neural networks: A survey," *IEEE Access*, vol. 12, pp. 144082–144105, 2024, doi: 10.1109/ACCESS.2024.3433513.
43. R. Ahmed, S. Islam, A. K. M. M. Islam, and S. Shatabda, "An ensemble 1D-CNN-LSTM-GRU model with data augmentation for speech emotion recognition," *Expert Systems with Applications*, vol. 218, May 2023, Art. no. 119633, doi: 10.1016/j.eswa.2023.119633.
44. D. N. Ravikiran, C. B. Lakshmi, A. S. R. Krishna, M. G. Krishna, and K. Jaswanth, "Parametric facial landmark detection using active shape models," *International Journal for Modern Trends in Science and Technology*, vol. 11, no. 3, pp. 79–85, Mar. 2025, doi: 10.5281/zenodo.15084856.
45. M. Z. Sajid, M. F. Hamid, I. Qureshi, M. I. Sharif, and N. Aburaed, "EfficientPoseSegNet: A weakly supervised, attention-guided framework for human pose estimation, anatomical segmentation, and concealed object detection in backscatter millimeter-wave security screening," *Scientific Reports*, vol. 16, Dec. 2026, Art. no. 731, doi: 10.1038/s41598-025-30346-1.
46. X. Chen *et al.*, "Shape registration with learned deformations for 3D shape reconstruction from sparse and incomplete point clouds," *Medical Image Analysis*, vol. 74, Dec. 2021, Art. no. 102228, doi: 10.1016/j.media.2021.102228.
47. J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," in *Proc. 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 2019, pp. 4685–4694, doi: 10.1109/CVPR.2019.00482.
48. A. Imran, R. Ahmed, M. Hasan, M. H. U. Ahmed, A. K. M. Azad, and S. A. Alyami, "FaceEngine: A tracking-based framework for real-time face recognition in video surveillance system," *SN Computer Science*, vol. 5, no. 5, pp. 149–169, May 2024, doi: 10.1007/s42979-024-02922-1.

49. S. S. Rani, S. Pournima, A. Aram, V. Shanmuganeethi, P. Thiruselvan, and N. H. A. Rufus, "Enhancing facial recognition accuracy in low-light conditions using convolutional neural networks," *Journal of Electrical Systems*, vol. 20, no. 5s, pp. 2140–2148, 2024, doi: 10.52783/jes.2559.
50. D. Zeng, R. Veldhuis, and L. Spreeuwers, "A survey of face recognition techniques under occlusion," *IET Biometrics*, vol. 10, no. 6, pp. 581–606, Apr. 2021, doi: 10.1049/bme2.12029.
51. H. Liu *et al.*, "Spatial-phase shallow learning: rethinking face forgery detection in frequency domain," in *Proc. 2021 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, pp. 772–781, doi: 10.1109/CVPR46437.2021.00083.
52. A. Dutt and P. Gader, "Wavelet multiresolution analysis based speech emotion recognition system using 1D CNN LSTM networks," *IEEE/ACM Transactions on Audio, Speech, Language Processing*, vol. 31, pp. 2043–2054, 2023, doi: 10.1109/TASLP.2023.3277291.
53. M. Abdul-Al, G. K. Kyeremeh, R. Qahwaji, N. T. Ali, and R. A. Abd-Alhameed, "A novel approach to enhancing multi-modal facial recognition: Integrating convolutional neural networks, principal component analysis, and sequential neural networks," *IEEE Access*, vol. 12, pp. 140823–140846, 2024, doi: 10.1109/ACCESS.2024.3467151.
54. S. Tipper, H. F. Atlam, and H. S. Lallie, "An investigation into the utilisation of CNN with LSTM for video deepfake detection," *Applied Science*, vol. 14, no. 21, Oct. 2024, Art. no. 9754, doi: 10.3390/app14219754.
55. Z. Yin *et al.*, "A survey: Spatiotemporal consistency in video generation," 2025, arXiv:2502.17863.
56. K. Narayan, V. VS, R. Chellappa, and V. M. Patel, "Facexformer: A unified transformer for facial analysis," in *Proc. of the IEEE/CVF Int. Conf. on Computer Vision (ICCV)*, Honolulu, Hawai'i, USA, Oct. 19-23, 2025, pp. 11369–11382.
57. Y. Kartynnik, A. Ablavatski, and M. Grundmann, "Real-time facial surface geometry from monocular video on mobile GPUs," 2019, arXiv:1907.06724.
58. T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L. -P. Morency, "OpenFace 2.0: Facial behavior analysis toolkit," in *Proc. 2018 13th IEEE Int. Conf. on Automatic Face & Gesture Recognition (FG 2018)*, Xi'an, China, pp. 59–66, doi: 10.1109/FG.2018.00019.
59. J. Pu, C. Yu, X. Chen, Y. Zhang, X. Yang, and J. Li, "Research on Chengdu Ma goat recognition based on computer vision," *Animals*, vol. 12, no. 14, Jul. 2022, Art. no. 1746, doi: 10.3390/ani12141746.
60. J. Ma, Y. Lin, L. Qian, H. You, and T. Gao, "Spectral-spatial feature fusion for real-time facial expression recognition," *Scientific Reports*, vol. 15, no. 1, Dec. 2025, Art. no. 43977, doi: 10.1038/s41598-025-27666-7.
61. A. Alzu'bi, F. Albalas, T. Al-Hadhrami, A. Albashayreh, and L. B. Younis, "MFI3D: Masked face identification with 3D face reconstruction and deep learning," *Neural Computing and Applications*, vol. 37, no. 25, pp. 20551–20567, Dec. 2024, doi: 10.1007/s00521-024-10582-8.
62. S. Naveen and M. R. Kounte, "Optimized convolutional neural network at the IoT edge for image detection using pruning and quantization," *Multimedia Tools and Applications*, vol. 84, no. 9, pp. 5435–5455, Dec. 2024, doi: 10.1007/s11042-024-20523-1.
63. M. Hassaballah, E. Salem, A.-M. M. Ali, and M. M. Mahmoud, "Deep recurrent regression with a heatmap coupling module for facial landmarks detection," *Cognitive Computation*, vol. 16, no. 4, pp. 1964–1978, Oct. 2022, doi: 10.1007/s12559-022-10065-9.
64. J. Pomalingo, "Comparative study of deep learning methods LSTM and 1D CNN algorithm: Case study of air pollution standard index data in DKI Jakarta," Ph.D. dissertation, Computer Science, Buana University, Jakarta, Indonesia, 2022.
65. S. Sharma and V. Kumar, "3D face reconstruction in deep learning era: A survey," *Archives of Computational Methods in Engineering*, vol. 29, no. 5, pp. 3475–3507, Jan. 2022, doi: 10.1007/s11831-021-09705-4.
66. C.-Y. Wu, Q. Xu, and U. Neumann, "Synergy between 3DMM and 3D landmarks for accurate 3D facial geometry," in *Proc. 2021 Int. Conf. on 3D Vision (3DV)*, London, United Kingdom, pp. 453–463, doi: 10.1109/3DV53792.2021.00055.
67. W. Zheng et. al., "Flame-based multi-view 3D face reconstruction," in *Proc. 40th Computer Graphics Int. Conf. (CGI 2023)*, Shanghai, China, Aug. 28-Sep. 1, 2023, pp. 327–339, doi: 10.1007/978-3-031-50078-7_26.
68. Z. Xu *et al.*, "Kmtalk: Speech-driven 3D facial animation with key motion embedding," in *Proc. 18th European Conf. on Computer Vision (ECCV)*, Milan, Italy, Sep. 29-Oct. 4, 2024, pp. 236–253, doi: 10.1007/978-3-031-72992-8_14.
69. L. Dong, X. Wang, S. Setlur, V. Govindaraju, and I. Nwogu, "Ig3D: Integrating 3D face representations in facial expression inference," in *Proc. 18th European Conf. on Computer Vision (ECCV)*, Milan, Italy, Sep. 29-Oct. 4, 2024, pp. 404–421, doi: 10.1007/978-3-031-91581-9_29.
70. K. Kärkkäinen and J. Joo, "FairFace: Face attribute dataset for balanced race, gender, and age for bias measurement and mitigation," in *Proc. 2021 IEEE Winter Conf. on Applications of Computer Vision (WACV)*, Waikoloa, HI, USA, pp. 1547–1557, doi: 10.1109/WACV48630.2021.00159.