IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, October 2026
Impact Factor: 8.2

Artificial Intelligence-Based Face Recognition Using Deep Learning - A Review of Models, System Design, Applications and Challenges

[1]Prof. Sanghesh Bele and [2]Sakshi R. Haramkar
[1]Assistant Professor, MCA
[2]Student of MCA
Department of Research and PG Studies in Science and Management, MCA Program
Vidyabharati Mahavidyalaya, Amravati, Maharashtra, India
sakshiharamkar5@gmail.com

Abstract: *Face recognition has become one of the most widely deployed applications of artificial intelligence, supporting smartphone unlocking, attendance management, banking and border control. Deep learning has replaced hand-crafted features with representations learned directly from large collections of face images, producing large gains in accuracy under unconstrained conditions. This paper presents a structured review of artificial intelligence-based face recognition using deep learning. It explains the stages of a recognition pipeline, compares the main families of deep models and training objectives, proposes a modular system architecture with a liveness check, and summarizes datasets and evaluation measures. Application areas, including education and financial services, are discussed together with the main challenges: pose and occlusion, spoofing, demographic bias, adversarial attacks and privacy. The paper closes with emerging directions such as self-supervised learning, synthetic training data, federated learning and privacy-preserving matching, and offers practical recommendations for responsible deployment*

Keywords: Face recognition, deep learning, convolutional neural network, biometrics, liveness detection, embedding, artificial intelligence..

I. INTRODUCTION

Recognizing a person from the face is a natural human skill and, for machines, a long-standing goal of computer vision. Unlike fingerprints or iris scans, a face can be captured at a distance, without contact and with an ordinary camera, which makes face recognition attractive for large-scale and convenient identity checks. Early systems relied on hand-designed descriptors and shallow classifiers, and they degraded sharply when lighting, pose or expression changed.

The arrival of deep learning, in which multi-layer neural networks learn features directly from data [1], changed the field. Landmark systems demonstrated accuracy approaching that of humans on benchmark face verification tasks [2], and embedding-based approaches learned a compact numerical signature in which faces of the same person lie close together [3]. Since then, larger datasets, better network designs and improved loss functions have continued to raise performance, and comprehensive surveys document this progress [4].

Despite these advances, deploying face recognition responsibly remains difficult. Systems must cope with masks, poor lighting and aging; resist presentation attacks such as printed photographs; treat people of different demographic groups fairly; and respect privacy law. For students and practitioners in countries such as India, where face-based attendance and electronic identity verification are expanding, a clear and integrated understanding of these issues is valuable.

The objectives of this paper are: (i) to describe the stages of a deep learning based face recognition pipeline; (ii) to compare the principal model families and training objectives; (iii) to propose a modular system architecture suitable for

Copyright to IJARSCT
www.ijarsct.co.in
DOI: 10.48175/IJARSCT-40364

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, October 2026
Impact Factor: 8.2

practical deployment; (iv) to summarize datasets and evaluation measures; and (v) to discuss applications, challenges and future directions.

III. RESEARCH METHODOLOGY

This work is a qualitative, analytical review. The authors studied peer-reviewed publications, benchmark documentation and government reports on face recognition, and then synthesized the material into an original framework. Sources were chosen for relevance to four themes: network architectures, training objectives, evaluation practice and responsible use. No experimental results or figures from other papers have been reproduced. The pipeline diagram, comparison tables and conceptual illustrations are the authors' own work, and statements about relative performance are qualitative rather than measured.

III. FOUNDATIONS

Verification and identification

Face recognition covers two related tasks. In verification, the system decides whether two images show the same person, which is a one-to-one comparison used when unlocking a device. In identification, the system searches a gallery of enrolled people to find the best match for a query image, which is a one-to-many comparison used in attendance or watch-list scenarios. A closed-set identification assumes the person is enrolled, while an open-set identification must also be able to answer that the person is unknown.

Stages of the pipeline

A modern pipeline has four main stages. Face detection locates faces in an image or video frame; cascaded convolutional detectors that also predict facial landmarks are a common choice [5]. Alignment uses the landmarks to rotate and scale each face to a standard pose. Feature extraction passes the aligned face through a deep network to obtain an embedding, which is a vector of typically a few hundred numbers. Finally, matching compares embeddings using a distance or similarity measure and applies a threshold to accept or reject the claim.

Why deep learning

Hand-crafted features capture a fixed set of cues and cannot adapt to the variability of real images. Deep networks instead learn hierarchies of features, from edges and textures in early layers to parts and identity-specific patterns in deeper layers, and they improve as more labeled data become available. They also produce embeddings that generalize to people never seen during training, so new users can be enrolled without retraining the network.

IV. DEEP LEARNING MODELS

Convolutional neural networks

The convolutional neural network (CNN) remains the workhorse of face recognition. Convolution layers share weights across the image and exploit its spatial structure. Early deep face models used stacks of plain convolution layers [6], whereas residual networks introduced skip connections that allow much deeper models to be trained stably [7] and are now a standard backbone.

Metric learning and loss functions

How the network is trained matters as much as its shape. Training with an ordinary softmax classification loss separates the training identities but does not guarantee compact clusters for unseen people. Triplet-based training addresses this directly by pulling an anchor image toward a positive example of the same person and pushing it away from a negative example [3], as illustrated in Fig. 1. Margin-based classification losses go further by adding a required gap between identities in angular space, producing tighter and more separable embeddings [8], [9].

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, October 2026
Impact Factor: 8.2

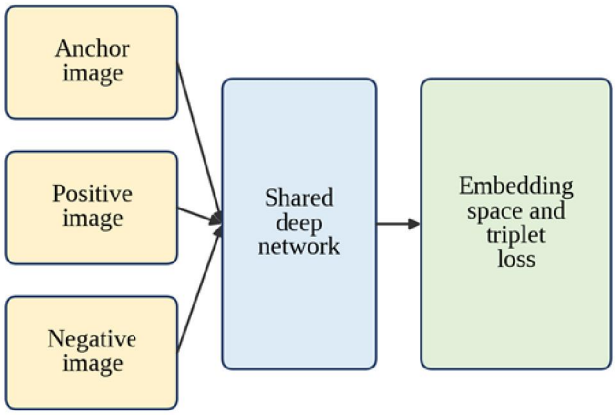


Fig. 1. Triplet-based embedding learning with a shared network.

Transformer-based models

Vision transformers divide an image into patches and use self-attention to relate every patch to every other [10]. Applied to faces, they can capture global relationships such as the consistency of both eyes, but they generally demand more data and computation than CNNs to reach comparable accuracy.

Lightweight models for devices

Applications on phones and embedded cameras need small, fast networks. Designers reduce cost through compact architectures, pruning, quantization and knowledge distillation, in which a small student network imitates a large teacher. The usual trade-off is a modest loss of accuracy on difficult images in exchange for large savings in memory and latency. Table I summarizes the main model families.

TABLE I. COMPARISON OF MODEL FAMILIES

Family	Strength	Limitation
CNN backbones	Mature, efficient	Local receptive field
Margin-based losses	Compact, separable embeddings	Sensitive to label noise
Triplet training	Learns similarity directly	Needs careful sampling
Vision transformers	Global context	Data and compute hungry
Lightweight networks	Fast on-device use	Lower accuracy on hard cases

V. PROPOSED SYSTEM ARCHITECTURE

We propose a modular architecture, shown in Fig. 2, that separates concerns so that each component can be improved independently. An input module receives still images or video frames. A detector finds faces, and an alignment module normalizes them. A liveness check examines the aligned face for signs of a photograph, screen replay or mask before any identity decision is made. An embedding network then produces the face signature, which a matching module compares with enrolled templates stored in a protected database. The decision module applies a threshold tuned for the target application and logs the outcome for audit.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, October 2026
Impact Factor: 8.2

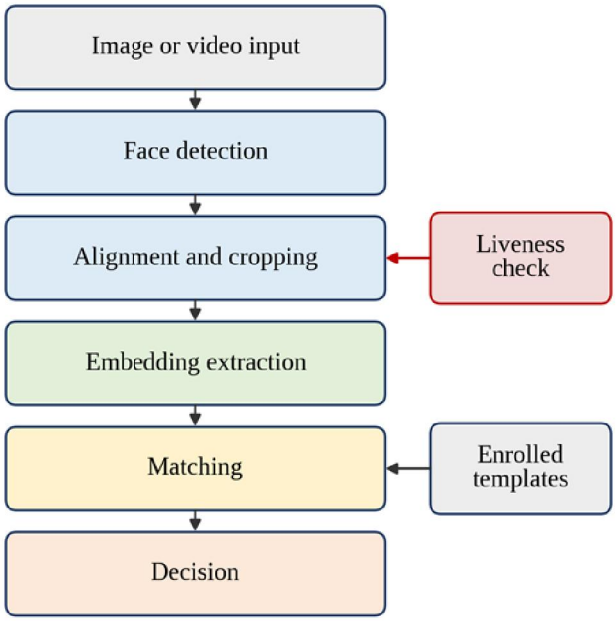


Fig. 2. Proposed modular face recognition pipeline.

Enrollment and template protection

During enrollment, several images of each person are captured under varied conditions and their embeddings are averaged or stored together. Storing embeddings instead of raw photographs reduces exposure, and encrypting the template database and restricting access further limit the damage of a breach. Because a face cannot be changed like a password, protection of templates deserves particular care.

Liveness detection

Presentation attacks are among the simplest ways to defeat a recognition system. Liveness checks may analyze texture and reflection patterns, estimate depth, detect eye blinks or head movement in response to a challenge, or use infrared sensing. Combining several cues is more robust than relying on one.

Threshold selection

The matching threshold controls a trade-off. A strict threshold rejects impostors but also turns away genuine users, whereas a lenient threshold does the opposite. High-security uses such as banking favor strict thresholds, and convenience uses such as classroom attendance may tolerate looser ones. Fig. 3 illustrates this relationship conceptually.

Copyright to IJARSCT
www.ijarsct.co.in
DOI: 10.48175/IJARSCT-40364

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, October 2026
Impact Factor: 8.2

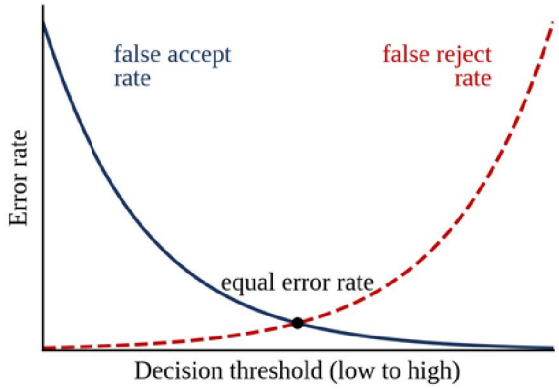


Fig. 3. Conceptual trade-off between false accept and false reject rates.

VI. DATASETS AND EVALUATION

Datasets

Progress depends on large and varied training data. Widely used public resources include Labeled Faces in the Wild, which holds web photographs for unconstrained verification [11], and VGGFace2, which was assembled to cover variation in pose and age [12]. Researchers should check the license and consent conditions of every dataset, because several large collections have been criticized or withdrawn over how images were gathered. Collecting a new dataset locally requires informed consent and a clear retention policy.

Evaluation measures

Verification is judged by error rates at a chosen threshold. The false accept rate (FAR) is the fraction of impostor attempts wrongly accepted, and the false reject rate (FRR) is the fraction of genuine attempts wrongly rejected. The point at which the two are equal is the equal error rate (EER). The receiver operating characteristic (ROC) curve plots the true accept rate against FAR across all thresholds. Identification is judged by the rank-1 identification rate, the proportion of queries whose correct identity is the top match. Table II defines these measures.

TABLE II. COMMON EVALUATION MEASURES

Measure	Definition	Typical use
FAR	Impostors wrongly accepted	Security level
FRR	Genuine users wrongly rejected	User convenience
EER	Threshold where FAR equals FRR	Single-number summary
ROC curve	True accept rate versus FAR	Comparing systems
Rank-1 rate	Correct identity is top match	Identification

Reporting a single accuracy figure on one benchmark can mislead. Results should be given across thresholds, across demographic groups and under realistic conditions such as low light, motion blur and masks.

Copyright to IJARSCT
www.ijarsct.co.in
DOI: 10.48175/IJARSCT-40364

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, October 2026
Impact Factor: 8.2

VII. APPLICATIONS

Security and access control

Offices, laboratories and residential complexes use face-based entry to replace cards and keys. Phones use on-device recognition to unlock and to authorize payments.

Education and attendance

Colleges can record attendance automatically as students enter a classroom, saving lecture time and preventing proxy attendance. Such systems must obtain consent, store templates securely and offer an alternative for students who decline.

Banking and electronic identity verification

Remote account opening increasingly compares a live selfie with a photograph on an identity document. Liveness detection is essential here to resist spoofing, and results are typically combined with other checks rather than used alone.

Healthcare

Hospitals can use face matching to confirm patient identity before treatment or medication, reducing record mix-ups. Related research examines facial analysis as a screening aid for some genetic conditions, though this requires clinical validation and careful ethics review.

Law enforcement and public safety

Police may compare images from cameras with databases of suspects or missing persons. This is the most contested use, because errors can lead to wrongful suspicion and mass surveillance can affect civil liberties. Many experts recommend human review of every match and strict legal limits.

Retail and customer service

Stores may use face analysis to personalize service or to detect known offenders. Transparent notice and opt-in consent are central to public acceptance.

VIII. CHALLENGES

Pose, illumination, aging and occlusion

Extreme angles, harsh light, blur and the passage of years change facial appearance. Face masks and sunglasses hide the features that models rely on, and systems trained mostly on unobstructed faces perform worse when these are present. Data augmentation, occlusion-aware training and the use of the visible eye region help, but do not remove the problem.

Spoofing and adversarial attacks

Beyond printed photographs and video replays, attackers can craft inputs designed to fool a neural network. Small, carefully computed perturbations can change a model's output while looking harmless to humans [13]. Defenses include adversarial training, input checks and the combination of recognition with strong liveness detection.

Demographic bias and fairness

Independent evaluations have found that error rates can vary across age, sex and ethnic groups, with some algorithms performing markedly worse for certain groups [14]. Audits of commercial systems have shown similar disparities in related facial analysis tasks [15]. Remedies include balanced and representative training data, per-group evaluation, threshold calibration and ongoing monitoring after deployment.

Privacy, consent and regulation

Facial images are sensitive biometric data. Laws such as the European General Data Protection Regulation place special conditions on processing biometric data [16], and India's Digital Personal Data Protection Act, 2023, requires lawful purpose and consent for personal data processing [17]. Developers should practice data minimization, clear consent, defined retention periods and the right to correction and deletion.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, October 2026
Impact Factor: 8.2

Computation and deployment cost

Training large models needs many graphics processing units (GPUs), and real-time matching against very large galleries demands efficient indexing. Organizations with limited resources should consider pretrained backbones, model compression and approximate nearest-neighbor search.

IX. FUTURE TRENDS

Self-supervised and synthetic data

Labeling millions of faces is costly and ethically sensitive. Self-supervised methods learn useful representations from unlabeled images, and synthetic faces generated by computer graphics or generative models may reduce reliance on scraped photographs while allowing controlled variation of pose and lighting.

Federated and privacy-preserving learning

In federated learning, devices train a shared model locally and send only updates to a server, so raw faces never leave the device [18]. Combined with encrypted matching and template protection schemes, this can enable collaboration across institutions while limiting exposure of biometric data.

Robust and multimodal recognition

Fusing the face with other signals, such as voice, gait or a device token, can improve accuracy and resist spoofing. Three-dimensional and infrared sensing are also becoming common in consumer devices.

Explainable and accountable systems

Users and regulators increasingly ask why a system produced a particular match. Research on visual explanations, uncertainty estimation and calibrated confidence scores aims to make decisions easier to audit and to challenge.

Efficient on-device intelligence

Dedicated neural processing hardware in phones and cameras allows recognition without sending images to the cloud, which reduces latency and privacy risk. Continued progress in compact architectures will extend this to low-cost devices.

X. RECOMMENDATIONS

For organizations planning to adopt face recognition, the following practices are recommended:

- Define the purpose narrowly and choose verification over identification whenever possible.
- Include liveness detection and set thresholds according to the risk of the application.
- Evaluate accuracy by demographic group and under local conditions before deployment.
- Obtain informed consent, minimize stored data and protect templates with encryption.
- Keep a human in the loop for high-impact decisions and provide a way to appeal.
- Monitor performance after launch and update models as conditions change.

IX. CONCLUSION

Deep learning has made face recognition accurate, convenient and widely available. This paper reviewed the stages of the recognition pipeline, compared the main model families and loss functions, proposed a modular architecture with liveness detection and template protection, summarized evaluation practice, and examined applications, challenges and future directions. Technical accuracy alone is not sufficient: fairness, privacy and accountability are equally necessary for public trust. Future work should validate the proposed architecture empirically on locally collected, consented data and report error rates across thresholds and demographic groups.

ACKNOWLEDGMENT

The authors thank the Department of Research and PG Studies in Science and Management, MCA Program, Vidyabharati Mahavidyalaya, Amravati, for providing the academic environment and resources that supported this work.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 6, October 2026
Impact Factor: 8.2

REFERENCES

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, May 2015.
[2] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2014, pp. 1701–1708.
[3] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2015, pp. 815–823.
[4] M. Wang and W. Deng, "Deep face recognition: A survey," Neurocomputing, vol. 429, pp. 215–244, March 2021.
[5] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," IEEE Signal Process. Lett., vol. 23, no. 10, pp. 1499–1503, October 2016.
[6] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in Proc. Br. Mach. Vis. Conf., 2015, pp. 41.1–41.12.
[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 770–778.
[8] H. Wang et al., "CosFace: Large margin cosine loss for deep face recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 5265–5274.
[9] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2019, pp. 4690–4699.
[10] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," in Proc. Int. Conf. Learn. Represent., 2021.
[11] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Univ. Massachusetts Amherst, Amherst, MA, Tech. Rep. 07-49, October 2007.
[12] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "VGGFace2: A dataset for recognising faces across pose and age," in Proc. IEEE Int. Conf. Autom. Face Gesture Recognit., 2018, pp. 67–74.
[13] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," in Proc. Int. Conf. Learn. Represent., 2015.
[14] P. Grother, M. Ngan, and K. Hanaoka, "Face recognition vendor test (FRVT) part 3: Demographic effects," National Institute of Standards and Technology, Interagency Report 8280, December 2019.
[15] J. Buolamwini and T. Gebru, "Gender shades: Intersectional accuracy disparities in commercial gender classification," in Proc. Conf. Fairness, Accountability Transparency, 2018, pp. 77–91.
[16] Regulation (EU) 2016/679 of the European Parliament and of the Council (General Data Protection Regulation), Official Journal of the European Union, L 119, April 2016.
[17] Digital Personal Data Protection Act, 2023, Government of India, August 2023.
[18] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in Proc. 20th Int. Conf. Artif. Intell. Stat., 2017, pp. 1273–1282.