E-ISSN: 2964-6219, P-ISSN: 2964-6200
Volume 4 Number 2 August 2026.
pp: 304-311. DOI: https://doi.org/10.58477/cj.v4i2.503.

Computer Journal

RESEARCH ARTICLE Open Access

Security Analysis on Face Biometric Authentication Under Different AI-Manipulated Face

Fathimah Hasanti [1*], **Nurul Ilmi** [2], **Desi Nurnaningsih** [3]

[1*] Department of Information Technology, Jakarta Campus Directorate, Universitas Telkom, West Jakarta City, Special Capital Region of Jakarta, Indonesia.

*Correspondence email: fathimahhasanti@telkomuniversity.ac.id

Received: 26 July 2026.
Revised: 14 July 2026.
Accepted: 27 August 2026.
Published: 30 August 2026.

Full list of author information is available at the end of the article.

Abstract

The rapid evolution of Generative Artificial Intelligence (Gen AI) has enabled realistic facial manipulation using AI-generated images, creating serious security threats to face biometric authentication systems. Previous studies have mainly focused on deepfake detection and face recognition improvement, while limited attention has been given to the effects of AI-based facial manipulation on biometric authentication security. This study examines the vulnerability of face biometric authentication to various Gen AI-powered facial manipulation attacks. FaceForensics++ was used to create verification pairs consisting of genuine, impostor, and manipulated faces. ArcFace generated face embeddings, while cosine similarity measured identity similarity between faces. Equal Error Rate (EER) was used to determine the authentication threshold. The threshold was then applied to evaluate DeepFakes, FaceSwap, Face2Face, NeuralTextures, and FaceShifter using False Acceptance Rate (FAR), False Rejection Rate (FRR), EER, and cosine similarity scores. The results indicate different authentication behaviors among manipulation algorithms. Face2Face and NeuralTextures produced higher FAR and similarity scores, indicating greater ability to retain identity information.

Keywords: AI-Manipulated; Authentication; Face Biometric; Facial Manipulation; Security.

Abstrak

Perkembangan pesat Generative Artificial Intelligence (Gen AI) telah memungkinkan manipulasi wajah yang realistis menggunakan gambar hasil olahan AI, sehingga menimbulkan ancaman serius terhadap keamanan sistem autentikasi biometrik wajah. Penelitian sebelumnya terutama berfokus pada deteksi deepfake dan peningkatan pengenalan wajah, sementara perhatian terhadap dampak manipulasi wajah berbasis AI terhadap keamanan autentikasi biometrik masih terbatas. Penelitian ini mengkaji kerentanan autentikasi biometrik wajah terhadap berbagai serangan manipulasi wajah berbasis Gen AI. FaceForensics++ digunakan untuk membuat pasangan verifikasi yang terdiri atas wajah asli, impostor, dan wajah termanipulasi. ArcFace menghasilkan embedding wajah, sedangkan cosine similarity mengukur kemiripan identitas antarwajah. Equal Error Rate (EER) digunakan untuk menentukan ambang autentikasi. Ambang tersebut diterapkan untuk mengevaluasi DeepFakes, FaceSwap, Face2Face, NeuralTextures, dan FaceShifter berdasarkan False Acceptance Rate (FAR), False Rejection Rate (FRR), EER, dan skor cosine similarity. Hasil menunjukkan perbedaan perilaku autentikasi antaralgoritma. Face2Face dan NeuralTextures menghasilkan FAR dan skor kemiripan lebih tinggi, menunjukkan kemampuan lebih besar dalam mempertahankan informasi identitas.

Kata Kunci: Manipulasi AI; Autentikasi; Biometrik Wajah; Manipulasi Wajah; Keamanan.

1. Introduction

The field of Artificial Intelligence (AI) has experienced notable advances in recent years due to continuous developments in computing, GPUs, cloud computing, and big data. These advancements have enabled intelligent algorithms to perform increasingly complex tasks across diverse fields, including answering natural language questions, writing code, analyzing text documents, generating digital art, manipulating images, and automating business operations (Kishri *et al.*, 2025). AI models are expected to become increasingly powerful and accessible as technological capabilities continue to improve. Although substantial evidence demonstrates the benefits of these technological advances for users, they also introduce new cybersecurity challenges (Holstein, 2024). The same AI technologies that improve productivity can be exploited by malicious actors to automate cyberattacks, making AI a double-edged sword that enhances defensive capabilities while enabling increasingly sophisticated attacks. AI can be used to generate convincing phishing content, bypass traditional security mechanisms, and impersonate legitimate users (Khaund, 2025). One of the most threatening applications enabled by AI is deepfake technology. Deepfakes are AI-generated media in which pre-existing audiovisual content is synthesized using deep neural networks, particularly GANs and diffusion models (Khaund, 2025). Recent developments have substantially improved the quality of facial manipulation while reducing the computational resources required for such operations. Modern deepfake tools can produce high-definition facial manipulations that humans may struggle to distinguish from authentic content (Masood *et al.*, 2022; He *et al.*, 2025). In enterprise settings, the use of deepfake technology is becoming a practical challenge for identity verification. Videos and speeches of executives can be created from public interviews, video calls can be manipulated during live sessions, and images can be generated to closely resemble actual staff members (Verma, 2025). This issue is particularly concerning for authentication systems because an increasing number of organizations use face recognition to verify identities (Fitzmaurice & O'Brien, 2025). Several studies have examined deep learning-based forensic techniques for detecting artificial intelligence-generated facial content (Patel & Musale, 2025; Naik *et al.*, 2025; Gore *et al.*, 2025; Castro *et al.*, 2026). Other studies have analyzed the resilience of face recognition algorithms against adversarial and presentation attacks (Vakhshiteh *et al.*, 2021; Lin *et al.*, 2023; Sheikh & Zafar, 2023; Ren *et al.*, 2024). In the area of detection, Sonkusare *et al.* (2022), Tolosana *et al.* (2022), and Tariq *et al.* (2022) primarily focused on detection, whereas Maaref *et al.* (2024) and Zhalgas *et al.* (2025) examined the influence of facial manipulation on authentication. Existing research also tends to evaluate only one or two manipulation techniques rather than multiple techniques within a single framework. For example, Tolosana *et al.* (2022) compared deepfake-related content without comprehensively evaluating different manipulation techniques, whereas Tariq *et al.* (2022) focused on impersonation rather than the broader security implications for authentication systems. Moreover, deepfake detection studies generally use classification metrics such as Accuracy, AUC, and F1 rather than authentication metrics such as FAR, FRR, and EER. Tariq *et al.* (2022) also reported a 78% success rate for targeted attacks against face recognition APIs. Few studies have analyzed the security implications of AI-manipulated faces and their impact on authentication systems. Various AI-based manipulation methods alter facial identity in different ways. For example, FaceSwap changes the facial identity between two individuals, whereas Face2Face manipulates facial expressions. Other techniques may modify facial characteristics through different manipulation mechanisms. These differences may result in different effects on biometric authentication systems. To address this issue, this paper analyzes the security of face biometric authentication under various AI-based facial manipulation scenarios using ArcFace and five facial manipulation methods: DeepFakes, FaceSwap, Face2Face, FaceShifter, and NeuralTextures. These manipulations are evaluated using False Acceptance Rate (FAR), False Rejection Rate (FRR), Equal Error Rate (EER), and cosine similarity scores.

2. Method

This study adopted an experimental approach to evaluate the robustness of a face biometric authentication system against multiple AI-manipulated face attacks. The methodology consisted of four major stages: dataset collection, data preprocessing, embedding extraction, and security testing and evaluation. Figure 1 shows the general workflow of the research process.

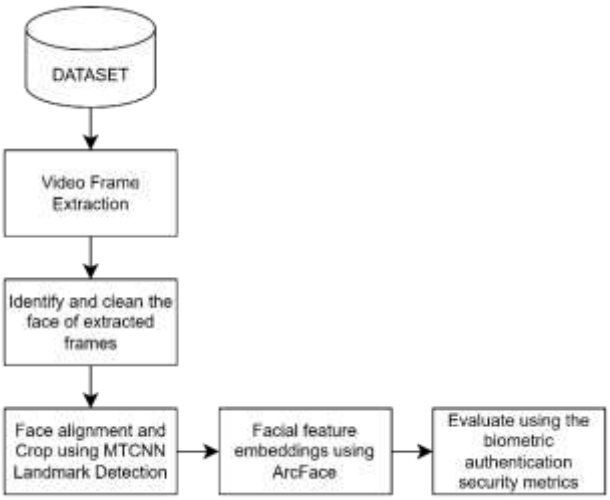

Figure 1. Research workflow

The experiments used facial images from the FaceForensics++ dataset obtained from a publicly available database contributed by users Xdxd_003 and Darshan. The dataset consists of multiple AI-manipulated faces based on real-life facial video footage. Five manipulation methods were selected for evaluation: DeepFakes, FaceSwap, Face2Face, FaceShifter, and NeuralTextures. These manipulation techniques represent different approaches to facial and identity modification, enabling a comparison of their effects on biometric authentication performance. Each manipulated facial image was compared with its corresponding original identity face. Data preprocessing consisted of three main stages: frame extraction, data cleaning, and face alignment. In the frame extraction stage (Sonkusare *et al.*, 2022), one middle frame was extracted from each video in the facial manipulation dataset. For the original non-manipulated video group, one supplementary frame was extracted two seconds after the middle frame in addition to the middle frame. These two frames constituted the Original extra dataset, which was used in the subsequent stage to establish the benchmark performance of the biometric authentication process. The data cleaning process removed duplicate data, blurry images, images in which faces were not optimally detected, and images containing faces that were too small. These issues were addressed using three parameters. A Blur Threshold of 60 was implemented based on Laplacian variance, a Confidence Threshold of 0.95 was set based on the MTCNN detection confidence level (Zhalgas *et al.*, 2025), and a Face Area Threshold of 2,000 pixels was applied. The cleaning process was performed consistently across all manipulation groups. If an identity did not meet the criteria in one group, all corresponding identity pairs in the other manipulation groups were also removed to maintain data consistency. The faces were then aligned using MTCNN Landmark Detection (Zhalgas *et al.*, 2025) by utilizing the left and right eye coordinates to calculate the rotation angle. Cropping and zooming were subsequently performed on the face area to standardize the orientation, position, and scale of the images before embedding extraction. After data preprocessing, facial embedding extraction was performed using the ArcFace method, a widely used face recognition approach (Zhalgas *et al.*, 2025). This process converted each face image into a 512-dimensional numerical vector. The performance evaluation of the biometric authentication system was conducted systematically to establish baseline performance before assessing the effects of different AI-based facial manipulations. The main metric used for matching was the cosine similarity score (Zhalgas *et al.*, 2025). The similarity score equation is formulated as follows (Krantz & Jonker, 2026):

$$Cosine\ Similiarty\ (A,B) = \frac{A \cdot B}{\|A\|\|B\|} \quad (1)$$

In Equation (1), A represents the vector of the reference face, while B represents the vector of the tested face. The symbol "·" indicates vector multiplication, and the denominator represents the length of each vector. The resulting score ranges from -1 to 1, where a score closer to 1 indicates a higher degree of similarity between the two faces. Authentication is considered successful when the similarity score is equal to or greater than a predetermined threshold; otherwise, the authentication is rejected. To determine the system's baseline security threshold, a preliminary evaluation was conducted exclusively using the original datasets without manipulated faces to assess the fundamental recognition capability of the ArcFace model. This phase involved calculating two types of matching scores: genuine scores and impostor scores. Genuine scores were obtained by comparing the ArcFace facial embedding generated from the primary frame of an original identity with another embedding of

the same identity obtained from the extra frame. In contrast, impostor scores were obtained by comparing the original ArcFace embedding of one identity with the original ArcFace embeddings of different identities in the dataset. Once the genuine and impostor similarity scores were obtained from the ArcFace embeddings, the system determined the authentication threshold based on the EER (Zhalgas *et al.*, 2025). The threshold was swept continuously from -1.0 to 1.0 with a step size of 0.001. At each threshold, the system calculated the FAR (Zhalgas *et al.*, 2025), defined as the proportion of impostor scores that met or exceeded the threshold, as mathematically formulated in Equation (2):

$$FAR = \left(\frac{Number\ of\ False\ Acceptances}{Total\ Number\ of\ Unauthorized\ Access\ Attempts} \right) \times 100\% \quad (2)$$

The FRR, which represents the proportion of genuine scores that fell below the threshold (Zhalgas *et al.*, 2025), was calculated as shown in Equation (3):

$$FRR = \left(\frac{Number\ of\ False\ Rejections}{Total\ Number\ of\ Legitimate\ Access\ Attempts} \right) \times 100\% \quad (3)$$

The optimal threshold was determined as the point at which the absolute difference between the FAR and FRR values was minimized. This threshold defined the EER value and was used as the fixed authentication threshold for all subsequent security testing procedures involving the ArcFace model. During the final evaluation stage, the resilience of the ArcFace biometric system was assessed against each AI-manipulated face approach: DeepFakes, FaceSwap, Face2Face, FaceShifter, and NeuralTextures. For each manipulation method, cosine similarity was measured between the embedding of the original face and the embedding of its corresponding manipulated face, with both embeddings extracted using ArcFace. The success of each manipulation was measured using FAR, calculated as the percentage of manipulated faces that produced a similarity score greater than or equal to the fixed baseline threshold established previously. Throughout the evaluation, the baseline FRR and EER values were kept constant because they were determined from the original genuine and impostor scores used to establish the ArcFace authentication threshold. This evaluation therefore focused on the effects of facial manipulation on FAR and the average cosine similarity score.

3. Research Method

This study uses a gaming addiction dataset consisting of 250 player records and 49 attributes. The dataset was obtained through survey-based data collection targeting active gamers across different age groups and gaming platforms. The survey collected information related to gaming behavior, psychological conditions, health indicators, and productivity-related factors associated with gaming addiction severity. The data collection process was conducted in accordance with the applicable research procedures, and informed consent was obtained from the participants. The target variable, addiction_severity, consists of four categories: Mild, Moderate, Severe, and Critical. The predictor variables include daily playtime hours, screen time duration, dopamine dependency index, self-control score, impulsiveness score, sleep hours, stress score, absenteeism days, and productivity drop percentage. These variables represent behavioral, psychological, health-related, and productivity-related aspects relevant to gaming addiction severity. Data preprocessing was performed to prepare the dataset for classification. The preprocessing procedures included data cleaning, missing-value handling, categorical encoding, feature-target separation, and feature scaling. Duplicate records were identified and removed, while potential outliers were examined using the interquartile range (IQR) method. Missing values in numerical features with less than 10% missing observations were handled using median imputation, whereas features with higher missing rates were processed using multiple imputation by chained equations (MICE). The target variable, addiction_severity, was encoded into numerical labels: Mild = 0, Moderate = 1, Severe = 2, and Critical = 3. Categorical predictor variables were encoded using label encoding for ordinal variables and one-hot encoding for nominal variables. The predictor variables were then separated from the target variable. Although Random Forest does not require feature scaling, standardization was applied to the numerical features to maintain a consistent preprocessing procedure for potential comparisons with other algorithms.

The dataset was divided into training and testing subsets using a stratified hold-out approach. Eighty percent of the data were allocated for training and 20% for testing, resulting in 200 training records and 50 testing records. Stratified sampling was applied to preserve the class distribution of the target variable

in both subsets. A random state of 42 was used to ensure reproducibility. Stratified 5-fold cross-validation was applied to the training data during model development. The 200 training records were divided into five folds, with four folds used for training and one fold used for validation in each iteration. This process was repeated until all folds had served as the validation set. Hyperparameter optimization was performed using Grid Search with cross-validation. The parameters evaluated included n_estimators = [100, 200, 300, 500], max_depth = [None, 10, 20, 30], min_samples_split = [2, 5, 10], min_samples_leaf = [1, 2, 4], and max_features = [sqrt, log2]. The best configuration was selected based on macro F1-score across the validation folds and subsequently used to train the final model. Random Forest was used as the primary classification algorithm. The model combines multiple decision trees and determines the final class through majority voting. To address the imbalanced distribution of addiction severity classes, balanced class weighting was applied during model training. The final Random Forest configuration used 300 decision trees (n_estimators = 300). The remaining hyperparameters were determined through the Grid Search procedure described Model performance was evaluated on the testing set using accuracy, precision, recall, and F1-score. Both macro-average and weighted-average scores were considered to provide a more complete assessment of performance under class imbalance. A confusion matrix was also used to examine correct and incorrect predictions for each addiction severity category. Feature importance analysis was conducted to identify the features with the highest importance in the Random Forest model and to determine which variables contributed most to the model's predictions. The experiments were implemented using Python 3.10. The Random Forest classifier and preprocessing procedures were implemented using scikit-learn version 1.3.0. Pandas version 2.1.0 was used for data manipulation, NumPy version 1.25.0 for numerical computation, and Matplotlib version 3.8.0 and Seaborn version 0.12.0 for data visualization. The experiments were conducted on a Windows 11 system equipped with an Intel Core i7 processor and 16 GB of RAM.

4. Results And Discussion

4.1 Result

Following the data cleaning process, the dataset size decreased. From the 1,000 identities across the five data groups, 494 were accepted and 506 were rejected. The details of the final data distribution used for the embedding and evaluation processes are presented in Table 1.

Table 1. Dataset Distribution After Pre-Processing

Explanation	Identity	Images
Initial Total Data	2,952.4	3,172.895
Accepted Data	779.3024	762.065
Rejected Data	43.4713	30.210

The extracted embedding vectors were used to calculate the similarity scores between faces. Tables 2 and 3 present samples of the first five genuine and impostor matching scores, respectively.

Table 2. Genuine Similarity Score Results Sample

Type	Images
Genuine	0.87432736
Genuine	0.91443604
Genuine	0.6326605
Genuine	0.7553873
Genuine	0.82342523

Table 3. Impostor Similarity Score Results Sample

Type	Images
Impostor	0.05934775
Impostor	0.05238277
Impostor	-0.03959059
Impostor	0.023657521
Impostor	-0.058396883

The matching process produced 494 genuine scores and 243,542 impostor scores. The optimal operating point was obtained at a threshold of 0.411. At this threshold, the authentication performance was 0.05% FAR, 0.00% FRR, and 0.02% EER. The threshold of 0.411 was subsequently used as a fixed benchmark for evaluating the AI facial manipulation attacks. The evaluation results for the five AI facial manipulation techniques are presented in Table 4.

Table 4. Evaluation of AI Manipulation Attacks at Threshold 0.411

Manipulation	Treshold	FAR	FRR	EER	Similiarty Score
DeepFakes	0.411	23.68%	0.00%	0.02%	0.3333
FaceSwap	0.411	37.85%	0.00%	0.02%	0.3851
Face2Face	0.411	99.60%	0.00%	0.02%	0.7269
NeuralTextures	0.411	100.00%	0.00%	0.02%	0.8462
FaceShifter	0.411	35.83%	0.00%	0.02%	0.3761

4.2 Discussion

The results demonstrate substantial differences in authentication vulnerability across the five AI facial manipulation techniques. The genuine similarity scores were substantially higher than the impostor scores, providing a basis for establishing the authentication threshold. The threshold of 0.411 was selected as the optimal operating point based on the baseline genuine and impostor score distributions, with 0.05% FAR, 0.00% FRR, and 0.02% EER. The results in Table 4 indicate that NeuralTextures and Face2Face posed the greatest risk to the ArcFace-based biometric authentication system. NeuralTextures produced the highest FAR at 100.00% and the highest similarity score at 0.8462, while Face2Face produced a FAR of 99.60% and a similarity score of 0.7269. Because both similarity scores were substantially above the fixed threshold of 0.411, manipulated faces generated using these techniques were frequently accepted by the authentication system. In contrast, DeepFakes, FaceSwap, and FaceShifter produced considerably lower FAR values of 23.68%, 37.85%, and 35.83%, respectively. Their corresponding similarity scores were also lower, at 0.3333, 0.3851, and 0.3761. These results indicate that these manipulation techniques were less successful in producing manipulated faces that exceeded the established authentication threshold. Nevertheless, their FAR values show that these techniques could still cause false acceptance under the experimental conditions. The different results among the manipulation techniques indicate that facial manipulation does not produce the same level of vulnerability in biometric authentication. In particular, the very high FAR values of Face2Face and NeuralTextures indicate that these techniques can produce manipulated faces whose identity-related characteristics remain sufficiently recognizable by the ArcFace model. This is also reflected in their substantially higher cosine similarity scores compared with the other manipulation techniques. The findings therefore address the objective of evaluating the effects of different AI-manipulated face techniques on biometric authentication robustness. The results show that the security impact varies according to the manipulation approach, with Face2Face and NeuralTextures representing the most critical cases in this experiment. Their FAR values of 99.60% and 100.00%, respectively, indicate a high likelihood of false acceptance when the manipulated faces are evaluated using the fixed authentication threshold. The findings should be interpreted within the scope of the experimental design. The baseline threshold of 0.411 was determined using genuine and impostor comparisons from the original dataset and was subsequently fixed during the manipulation evaluation. Therefore, the FRR and EER values reported in Table 4 represent the baseline authentication characteristics rather than independently recalculated values for each manipulation technique. The primary indicators of manipulation vulnerability in this evaluation are therefore FAR and similarity score. This study has several limitations. The scope was restricted to AI-manipulated faces sourced from the FaceForensics++ dataset and to the ArcFace recognition model for identity verification. Therefore, the FAR and similarity score distributions may differ when other generative datasets or face recognition algorithms are used. In addition, this study evaluated the biometric authentication system separately from the presentation attack detection process; consequently, the effectiveness of combining both processes in practice has not yet been established. Future studies may investigate the integration of specialized deepfake detection algorithms as a mandatory preprocessing step within the biometric authentication pipeline, as proposed by Sonkusare *et al*. (2022) and Tolosana *et al*. (2022).

5. Conclusion

The experimental results demonstrate that the vulnerability of face biometric authentication to AI-manipulated face attacks varies significantly depending on the manipulation technique. Among the five evaluated techniques, NeuralTextures and Face2Face showed the highest vulnerability, with FAR values of 100.00% and 99.60%, respectively, accompanied by the highest similarity scores of 0.8462 and 0.7269. These results indicate that both techniques can produce manipulated faces with identity-related characteristics that remain sufficiently similar to the corresponding original faces to exceed the ArcFace authentication threshold of 0.411. In comparison, DeepFakes, FaceSwap, and FaceShifter produced lower FAR values of 23.68%, 37.85%, and 35.83%, respectively. These findings confirm that different AI facial manipulation techniques have different impacts on biometric authentication security. NeuralTextures and Face2Face represent the most critical manipulation techniques under the experimental conditions because of their high false acceptance rates. This study is limited to the FaceForensics++ dataset and the ArcFace recognition model. Future research should evaluate additional generative datasets and recognition models and investigate the integration of deepfake detection with biometric authentication to improve resistance against AI-manipulated face attacks.

References

Castro, F., Gattulli, V., Impedovo, D., & Monaco, A. (2026). DeepDect: An explainable AI platform for face swapping and face generation DeepFake detection. *Multimedia Tools and Applications, 85*, Article 500. https://doi.org/10.1007/s11042-026-21672-1

Fitzmaurice, D., & O'Brien, O. (2025). Generative AI deepfakes: Real-world threats to facial biometric authentication. In *2025 Cyber Research Conference–Ireland (Cyber-RCI)*. https://doi.org/10.1109/cyber-rci68134.2025.11385234

Gore, S., Shaikh, N., Sandeep, K. S., Jagadish, S., Priya, L., & Priyanka, T. P. (2025). Next-gen biometric security: AI for identity authentication. In *2025 6th International Conference on Electronics and Sustainable Communication Systems (ICESC)* (pp. 2205–2212). IEEE. https://doi.org/10.1109/ICESC65114.2025.11212442

He, S., Lei, Y., Zhang, Z., Sun, Y., Li, S., Zhang, C., & Ye, J. (2025). *Identity deepfake threats to biometric authentication systems: Public and expert perspectives*. arXiv. https://doi.org/10.48550/arXiv.2506.06825

Holstein, D. (2024). *Analysis of attack methods with artificial intelligence* [Bachelor's thesis, Esslingen University]. https://doi.org/10.13140/RG.2.2.23071.06566

Khaund, B. (2025). AI and identity security: The threat of deepfakes and the future of authentication. *World Journal of Advanced Engineering Technology and Sciences, 15*(3), 2094–2101. https://doi.org/10.30574/wjaets.2025.15.3.1135

Al Kishri, W., Yousif, J. H., Al Bahri, M., Zakarya, M., Khan, N., Al Maskari, S. S., & Gurhanli, A. (2025). A comparative study of deepfake facial manipulation technique using generative adversarial networks. *Discover Artificial Intelligence, 5*(1), Article 109. https://doi.org/10.1007/s44163-025-00337-2

Krantz, T., & Jonker, A. (2026, May 14). *What is cosine similarity?* IBM. https://www.ibm.com/think/topics/cosine-similarity

Lin, C., Chen, F., Ng, H., & Lin, W. (2023). Invisible adversarial attacks on deep learning-based face recognition models. *IEEE Access, 11*, 51567–51577. https://doi.org/10.1109/ACCESS.2023.3279488

Maaref, Z., Belhadj, F., Attia, A., Akhtar, Z., Jasser, M. B., Ramly, A. M., & Mohamed, A. W. (2024). A comprehensive review of vulnerabilities and attack strategies in cancelable biometric systems. *Egyptian Informatics Journal, 27*, Article 100511. https://doi.org/10.1016/j.eij.2024.100511

Masood, M., Nawaz, M., Malik, K. M., Javed, A., Irtaza, A., & Malik, H. (2023). Deepfakes generation and detection: State-of-the-art, open challenges, countermeasures, and way forward. *Applied Intelligence, 53*(4), 3974–4026. https://doi.org/10.1007/s10489-022-03766-z

Naik, D., Inamdar, A., & Sonawane, Y. (2025). Fake media forensics: AI-driven forensic analysis of fake multimedia content. *International Journal of Scientific Research in Engineering and Management, 9*(5), 1–9. https://doi.org/10.55041/IJSREM47208

Patel, J., & Musale, V. (2025). Automated AI-powered intrusion detection system (IDS), biometric authentication and deepfake detection. In *Proceedings of the International Conference on Computational Complexity and Intelligent Algorithms* (pp. 663–676). Springer. https://doi.org/10.1007/978-981-96-7473-2_49

Ren, M., Wang, Y., Zhu, Y., Huang, Y., Sun, Z., Li, Q., & Tan, T. (2024). Artificial immune system of secure face recognition against adversarial attacks. *International Journal of Computer Vision, 132*(12), 5718–5740. https://doi.org/10.1007/s11263-024-02153-0

Sheikh, B. U. H., & Zafar, A. (2024). Beyond accuracy and precision: A robust deep learning framework to enhance the resilience of face mask detection models against adversarial attacks. *Evolving Systems, 15*(1), 1–24. https://doi.org/10.1007/s12530-023-09522-z

Sonkusare, M. G., Meshram, H. A., Sah, A., & Prakash, S. (2022). Detection and verification for deepfake bypassed facial feature authentication. In *2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS)* (pp. 646–649). IEEE. https://doi.org/10.1109/ICAIS53314.2022.9742980

Tariq, S., Jeon, S., & Woo, S. S. (2022). Am I a real or fake celebrity? Evaluating face recognition and verification APIs under deepfake impersonation attack. In *Proceedings of the ACM Web Conference 2022* (pp. 512–523). ACM. https://doi.org/10.1145/3485447.3512212

Tolosana, R., Romero-Tapiador, S., Vera-Rodriguez, R., Gonzalez-Sosa, E., & Fierrez, J. (2022). DeepFakes detection across generations: Analysis of facial regions, fusion, and performance evaluation. *Engineering Applications of Artificial Intelligence, 110*, Article 104673. https://doi.org/10.1016/j.engappai.2022.104673

Vakhshiteh, F., Nickabadi, A., & Ramachandra, R. (2021). Adversarial attacks against face recognition: A comprehensive study. *IEEE Access, 9*, 92735–92756. https://doi.org/10.1109/ACCESS.2021.3092646

Verma, S. (2025). Synthetic identities and deepfake attacks: The next frontier in enterprise AI security. *European Modern Studies Journal, 9*(5), 176–189. https://doi.org/10.59573/emsj.9(5).2025.17

Zhalgas, A., Amirgaliyev, B., & Sovet, A. (2025). Robust face recognition under challenging conditions: A comprehensive review of deep learning methods and challenges. *Applied Sciences, 15*(17), Article 9390. https://doi.org/10.3390/app15179390

How Cites

Hasanti, F., Ilmi, N., & Nurnaningsih, D. (2026). Security Analysis on Face Biometric Authentication Under Different AI-Manipulated Face. *Computer Journal*, *4*(2), 304-311. https://doi.org/10.58477/cj.v4i2.503.

Publisher's Note