Journal of Information Technology Management

Print ISSN: 2008-5893

Online ISSN: 2423-5059

F-MIM: Feature-based Masking Iterative Method to Generate the Adversarial Images against the Face Recognition Systems

Khushabu Agrawal*

*Corresponding Author, Department of Computer Engineering & Applications, GLA University, Mathura, 281406, (U.P.) India. E-mail: agkhushboo1996@gmail.com

Charul Bhatnagar

Department of Computer Engineering & Applications, GLA University, Mathura, 281406, (U.P.) India. E-mail: charul@gla.ac.in

Abstract

Numerous face recognition systems employ deep learning techniques to identify individuals in public areas such as shopping malls, airports, and other high-security zones. However, adversarial attacks are susceptible to deep learning-based systems. The adversarial attacks are intentionally generated by the attacker to mislead the systems. These attacks are imperceptible to the human eye. In this paper, we proposed a feature-based masking iterative method (F-MIM) to generate the adversarial images. In this method, we utilize the features of the face to misclassify the models. The proposed approach is based on a black-box attack technique where the attacker does not have the information related to target models. In this black box attack strategy, the face landmark points are modified using the binary masking technique. In the proposed method, we have used the momentum iterative method to increase the transferability of existing attacks. The proposed method is generated using the ArcFace face recognition model that is trained on the Labeled Face in the Wild (LFW) dataset and evaluated the performance of different face recognition models namely ArcFace, MobileFace, MobileNet, CosFace and SphereFace under the dodging and impersonate attack. The F-MIM attack is outperformed in comparison to the existing attacks based on Attack Success Rate evaluation metrics and further improves the transferability.

Keywords: Adversarial attack, Black-box attack, Dodging attack, Face Recognition, Feature based attack.

Journal of Information Technology Management, 2023, Vol. 15, Special Issue, pp. 80- 93
Published by University of Tehran, Faculty of Management
doi: https://doi.org/ 10.22059/jitm.2023.95247
Article Type: Research Paper

Received: July 06, 2023
Received in revised form: August 24, 2023
Accepted: November 09, 2023
Published online: December 24, 2023

Introduction

Recent advancements in deep neural network (DNN) technology have increased the immense success of computer vision applications. Face recognition (FR) is a crucial application of computer vision that is commonly used in real-world applications like passport matching, attendance systems, and security. Face recognition models have two tasks: face identification and face verification (Qiu et al., 2022; Yan et al. 2021; Turk et al. 1991). The first determines the identity of the face images, while the second verify that the predicted face image matches the input image.

According to the present study, DNN is vulnerable to adversarial attacks (Szegedy et al., 2013; Dong et al., 2019; Deb et al., 2020; Biggio et al., 2013). The attackers create these attacks on purpose by adding minimum perturbations so that FR models predict the wrong output and degrade their robustness (Zhong and Deng, 2020a; Massoli et al., 2020; Zhou et al., 2018; Agrawal and Bhatnagar, 2021; Agrawal k. et al.,2023). The attacker tries to mislead the model classifiers by applying the imperceptible attack.

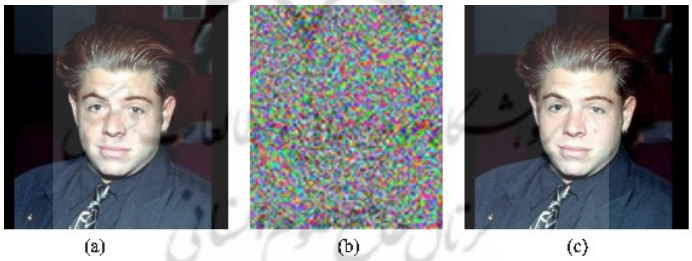

Figure 1. Shows the sample image of adversarial attack on face recognition models (Huang et al., 2008).

In Figure 1, we can see the input image and the generated image are similar but after adding the adversarial perturbation the output image is misclassified by the model. The attackers can fool the model in two ways: 1) the model either misclassifies the identity of a person or is unable to recognize it which is called an untargeted attack. 2) The model classified the identity of an individual person to another target person's identity which is called target identity. We input the pair of images to achieve the target and untargeted attack while generating the adversarial attack. We take a pair of images with similar identities for untargeted attacks and pair of distinct identities for targeted attacks.

However, current research indicates that the majority of attacks against FR models come under the white-box scenario. In this attack, the adversary has internal knowledge of the target model like structure and parameters. White-box attacks are impractical in real-world scenarios where we do not know the detailed information of the target models. In the black-box scenario, the adversary generates the attacks without any knowledge of the target models.

The attack has been divided into two forms based on the target model's knowledge: white-box attack and black-box attack.

White-box attack: To generate the white-box attack, the adversary must understand the model's parameters, architecture, and other characteristics. These attacks are model-specific and are incapable of deceiving other models (Dong et al., 2019; Li et al., 2019; Dong et al., 2020).

Black-box attack: The adversary lacks in-depth knowledge of the target models. These attacks are difficult to generate for the adversary. In addition, black-box attacks enhance transferability by decreasing their dependency on model hyper-parameters and model structure. The adversary can observe the generated output with respect to the given input to the target model (Dong et al., 2019; Li et al., 2019; Dong et al., 2020; Agrawal et al., 2023).

The face recognition model can be fooled in two ways:

Dodging Attack: In the context of adversarial attacks, a dodging attack refers to a specific type of attack strategy that aims to evade or "dodge" detection by the targeted system. It involves manipulating or crafting adversarial examples that can deceive the system into making incorrect decisions or classifications while remaining inconspicuous or undetectable to human observers.

The primary objective of a dodging attack is to create adversarial examples that closely resemble the original input but lead to erroneous or desired outcomes when processed by the target system (Sharif et al., 2016). These attacks are designed to exploit vulnerabilities or limitations in the system's recognition algorithms, decision boundaries, or feature extraction processes. Figure 2. Shows an example of dodging attack.

Figure 2. Shows the sample image of dodging attacks on face recognition models. (a) Original image, (b) Generated image, (c) Generated perturbation mask (Huang et al., 2008).

Impersonate Attack: An impersonate attack refers to a specific type of attack strategy where an adversary attempts to deceive a system by impersonate a specific individual or a target class of individuals. The goal of an impersonate attack is to manipulate the system into misclassifying the adversary as the desired target or granting unauthorized access by exploiting the system's vulnerabilities. Impersonate attacks often focus on bypassing authentication systems or identity verification processes (Sharif et al., 2016). This can include techniques such as spoofing biometric data (e.g., fingerprints, facial images) or manipulating input data to mimic the characteristics of the target individual. Figure 3. Shows the example of impersonate attack.

Figure 3. Shows the sample image of impersonate attack on face recognition models. (a) Original image, (b) Generated image, (c) Target identity (Huang et al., 2008).

The main contributions of this paper are as follows:

•The feature-based attack is introduced using the binary masking technique. In this method, we modify the features of the face by applying the mask. The mask is utilized in such a way that the changes are not easily noticeable to humans but the model is unable to predict the correct output.

• The proposed attack is fused with the Momentum iterative method (MIM) to increase the transferability of the proposed attack. The transferability refers to, the generated attack can also fool the different FR models.

In the paper, the proposed attack is generated using the ArcFace face recognition model that is trained on LFW dataset and evaluated the performance of different face recognition models namely ArcFace, MobileFace, MobileNet, CosFace and SphereFace under the dodging and impersonate attack using the Attack Success Rate evaluation metric.

Literature Review

Face recognition algorithms are applied in several safety-critical areas. It is important to understand the level of vulnerability of deep-face models in real-world scenarios. Adversarial attacks are used to fool face recognition by adding some modifications to the input images that are not visualized by the human eye. In this section, background study of the existing adversarial attacks on face recognition models is discussed in detail.

The researchers (Mahmood et al., 2016) introduced an attack generation method and named it the first gradient-based attack. It limits the amount of perturbation to eye-wear and observed the target and untargeted attacks for the FR system. Further, an adversarial attack was proposed for eyeglasses by employing a generative adversarial network. This proved those face recognition systems are vulnerable and showed that the attacks are physically feasible.

Similarly, the author (Zhou et al., 2018) crafts a cap with the help of infrared LED dot light. To generate this malicious physical attack the infrared dot light has been illuminated on the face. This attack also evades the detection of a face by adjusting the illumination, and size of the light.

The author (Rozsa et al., 2017) presented LOTS attack to target face recognition systems. It focused on perturbing the features of the existing neural networks. The approach is similar to where the internal features are directly modified. Moreover, the researcher also proved the vulnerability of face recognition systems based on geometrical perturbation. It altered the landmark location points of the original image to create the adversarial images. Until now, white-box settings have been the primary target of adversarial attacks.

White-box settings (Sharif et al. 2016; Schroff et al. 2015) have attack success rates that are almost 100% but they are not useful in real-world scenarios. As the attackers have accessibility to target face recognition models. The author (Dong et al. 2019) considered the limitation of the white-box and presented a query adversarial attack based on an evolutionary algorithm. The proposed attack works well in terms of attack success rate. However, query-based approaches in the black-box settings rely on the number of queries that can be easily detected by target models.

In this research, the researcher found that adding random noise and blur or even grid horizontal and vertical lines can also deceive the robustness of the FR model. The author (Goswami et al., 2018) presented the image-level and face-level image distortion attacks and reduce the accuracy of renowned face recognition models.

The author (Yang et al., 2021) proposed a method to fool the model by using another identity. The model predicts the original image as the target image instead of misclassifying it. The author introduced an Attentional Adversarial Attack Generating Network (A3GN) to create adversarial samples that are close to the actual pictures but have the same visual features as the target face. To learn the instance-level correspondences between the faces, they included a conditional variational autoencoder and attention modules to extract the target human's semantic features.

Also, the author (Wei et al., 2022) introduced an approach to generate the sticker by using the heuristic differential algorithm. To successfully generate this region-based sticker,

the algorithm discovered the solution for the sticker's parameter which is utilized to find the new region of the image for aggregation and modification of effective results. The proposed method was evaluated on face recognition and also extended for traffic sign recognition and image retrieval.

Hence, despite considerable advancements in attack generation, there remains a necessity to devise effective attacks that operate without requiring access to the model parameters or architectural details. Such attacks are commonly referred to as black-box attacks. In this paper, we generate the black-box attack by using a model as a surrogate model and then transferring the generated adversarial attack to the target models. Later, we evaluate the robustness of the model and the attack success rate of the proposed attack.

Methodology

Feature-based Masking Iterative Method (F-MIM) is proposed to generate adversarial examples to fool Face Recognition Systems. We have used ArcFace FR model (Deng et al., 2019) to generate the attack and tested the robustness of different FR models namely ArcFace, MobileFace (Chen et al., 2018), MobileNet (Howard et al., 2017), CosFace (Wang et al., 2018) and SphereFace (Liu et al., 2017). The attack generation process is shown in Figure.4.

Figure 4. The attack generation process of the proposed method F-MIM.

Feature-based Masking Iterative Method (F-MIM)

In the proposed method, we use the pair of images of the LFW dataset. In the preprocessing stage, we resize the images in to the size of 112*112 because the face recognition systems are scale-variant. After resizing the images, the detection of the face image is performed using the multi-task cascade neural network (MTCNN) (Zhang et al., 2016). The MTCNN contains three main phases namely, proposal convolutional neural network(P-CNN), Refine convolutional neural network (R-CNN) and output convolutional neural network(O-CNN) to classify the face or non-face and generates the bounding box around the face and also returns the landmark locations of the face. The landmark locations are the prominent locations of the face. By making changes in the prominent locations of the face, the model can be misled. The binary masking technique is applied to the facial features to hide the information of the face so that model is unable to identify the correct person's identity. In the proposed attack, the

binary mask is used to modify the pixel of the prominent regions by using Equation.1. The binary mask helps to alter the bits 0 to 1 or vice-versa by using the boolean operator.

$$\boldsymbol{S_m = \sim S_L \; ; \; where \; S_m \in \{0, 1\}} \quad (1)$$

Where S_L is the number of landmark points.

Then, modified locations of landmarks are added in the image X_i using the dot product using the Equation.2.

$$\boldsymbol{X_i^p = S_m \odot X_i} \quad (2)$$

Where the number of images $i \in \{1,2,3 \ldots k\}$ and the binary mask S_m is applied by using the pixel-wise product on image X. Then, the output image X_i^p obtained using the filter is fused with the existing MIM adversarial attack (Dong et al., 2018) to increase the transferability. In Equation.3, x_{i+1} represents the computation of velocity vector in gradient direction to avoid the poor local maxima obtained in each iteration.

$$\boldsymbol{x_{i+1} = \mu \cdot x_i + \frac{\nabla_{X^p} D_f(X_i^p, X_i^t)}{\|\nabla_{X^p} D_f(X_i^p, X_i^t)\|}} \quad (3)$$

Where $\nabla_{X^p} D_f$ is the gradient loss of the image X_i^p and target image X_i^t, μ is the decay factor. Then, Equation.4 is used to calculate the X_{i+1}^p by using the gradient sign during each iteration.

$$\boldsymbol{X_{i+1}^p = \mathbf{clip}_{X,\varepsilon}\left(X_i^p + \alpha \cdot \mathbf{sign}(x_{i+1})\right)} \quad \textbf{(4)}$$

Empirically, the value of α is set as $\alpha = 4\varepsilon/2n$, the threshold value $\varepsilon = 8$ under the L_∞ norms and the total iteration $n = 100$. L_∞ norms compute the maximum difference of the corresponding pixels between the images (Sharif et al., 2018). The performance of proposed is tested for dodging and impersonate attacks. In the case of a dodging attack, the cosine distance between the generated images and the targeted images should be less than the threshold.

$$\boldsymbol{cosinedistance\ (X^p, X^t) < \varepsilon} \quad (5)$$

And, for impersonate attack, the cosine distance should be greater than the threshold

$$\boldsymbol{cosinedistance\ (X^p, X^t) > \varepsilon} \quad (6)$$

Dataset

In this research, we used the LFW dataset (Huang et al., 2008) for training and testing the proposed F-MIM attack. It consists of 13,234 face images. These are the images of 5,749

distinct identities in various illuminations, poses, and expressions. The dataset consists of 6,000 face pair images, including 3000 pair images of similar identities and 3000 pair images of dissimilar identities. The dataset is widely used by researchers for face verification and identification. Figure.5 shows the sample images of the LFW dataset.

Figure 5. The Sample images of LFW dataset (Huang et al., 2008)

Experimental Setup

The implementation of the F-MIM attack has been conducted on Google Colaboratory. It provides an open-source environment with a Tesla K80 GPU backend. It is helpful in the training and testing of various machine learning classifications.

Results

The quantitative results have been compared with renowned attacks that are FGSM (Goodfellow et al., 2014), BIM (Kurakin et al., 2016) and MIM (Dong et al., 2018). The attacker does not have any prior information of the target model under the proposed black-box attack. To compare and analyze the effectiveness of trained models namely ArcFace, MobileFace, MobileNet, CosFace and SphereFace using the attack Success Rate(SR) metric (Zhong et al., 2020a). An attack is considered successful if the obtained distance is less than the threshold and FR model is unable to predict the correct output.

The distance is computed between the generated image and the target image. The higher success rate, the better the transferability. With respect to transferability, by occluding the features proposed attack is more efficient. The attack success rate is computed using Equation.7.

$$\mathbf{SR}\ (\%) = \mathbf{100} \times \frac{\mathbf{1}}{\boldsymbol{K}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{K}} \left[\left\| \left(\boldsymbol{X}_{\boldsymbol{i}}^{(\boldsymbol{p})} \right) - (\boldsymbol{X}_{\boldsymbol{i}}) \right\|_{\mathbf{2}} < \boldsymbol{\varepsilon} \right] \tag{7}$$

Where K denotes the total number of images.

Table 1. The SR of proposed under the dodging attack with L_∞ norm metric

	Attacks	ArcFace	MobileFace	MobileNet	CosFace	SphereFace
ArcFace	FGSM	0.986	0.578	0.442	0.496	0.522
	BIM	1	0.862	0.749	0.674	0.732
	MIM	1	0.9	0.835	0.788	0.831
	F-MIM	**1**	**0.96**	**0.94**	**0.928**	**0.89**

The quantitative results of the proposed black-box attack have been shown in Table 1 and Table 2. The proposed F-MIM attack is performed under the dodging and impersonate attacks. Table1 represents the SR of the proposed attack on existing face recognition models under the dodging attack with L_∞ norm metric and Table 2. represents the SR of the proposed attack under the impersonate attack with L_∞ norm metric. The graphical representation of the attack comparison has been shown in Figure.6 and Figure.7.

Table 2. The SR of proposed under the impersonate attack with L_∞ norm metric

	Attacks	ArcFace	MobileFace	MobileNet	CosFace	SphereFace
ArcFace	FGSM	0.99	0.612	0.326	0.407	0.408
	BIM	1.0	0.79	0.474	0.523	0.577
	MIM	1.0	0.877	0.570	0.622	0.665
	F-MIM	**1.0**	**0.91**	**0.85**	**0.86**	**0.84**

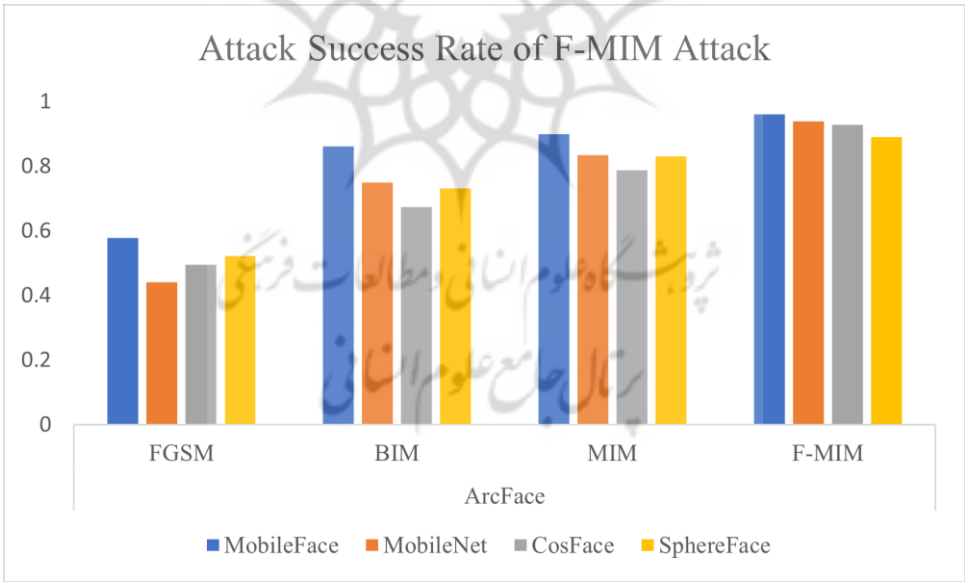

Figure 6. The comparison of proposed attack on target models under the dodging black-box attack with L_∞ norm metrics.

Conclusion

Deep learning has been greatly used for face recognition. In this paper, the proposed attack has been proposed to generate the perturbed image by using the binary mask to occlude the feature points of the face. The proposed has been performed under the dodging and

impersonating attack with L_∞ norm metric on the LFW dataset. The proposed black-box attack outperforms the existing attacks on ArcFace, MobileFace, MobileNet, CosFace and SphereFace FR models. The experimental result shows that the proposed attack improves transferability. Moreover, the robustness of the existing face recognition models can be increased on various neural network architectures, hyper-parameters, and datasets. In the future, the amount of pixel modification can be reduced by using different filtering techniques.

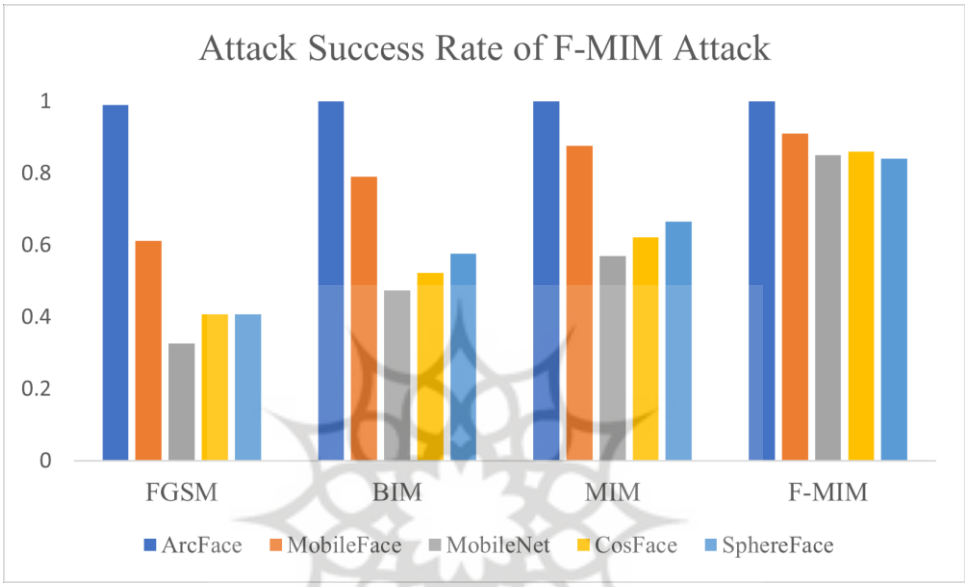

Figure 7. The comparison of proposed attack on target models under the impersonate black-box attack with L_∞ norm metrics.

Conflict of interest

The authors of this paper state that they do not have any competing financial interests or personal relationships that could have influenced their work. We would like to verify that there are no conflicts of interest related to this publication and that no significant financial support has been received that could have influenced the outcome of the research. This statement indicates that the authors have taken steps to ensure that their work is unbiased and free from any undue influence.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

A. Yahyaoui ,A. Jamil,J. Rasheed, M. Y. (2019). 1st International Informatics and Software Engineering Conference (IISEC-2019) : "Innovative Technologies for Digital Transformation" : proceedings book : 6-7 November 2019, Ankara/Turkey. 2, 1–4.

Agrawal, K., & Bhatnagar, C. (2023). M-SAN: a patch-based transferable adversarial attack using the multi-stack adversarial network. *Journal of Electronic Imaging*, 32(2), 023033.

Agrawal, K., & Bhatnagar, C. (2023, May). A Black-box based Attack Generation Approach to Create the Transferable Patch Attack. In 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 1376-1380). IEEE.

Agrawal, K., Bhatnagar, C., 2021. Bmim: Generating adversarial attack on face recognition via binary mask, in: 2021 International Conference on Intelligent Technologies (CONIT), pp. 1–5. doi:10.1109/CONIT51480.2021.9498370.

Biggio, B., Corona, I., Maiorca, D., Nelson, B., Srndi ˇ c, N., Laskov, P., Giacinto, G., Roli, F., 2013. Evasion attacks against machine learning at test time, in: Joint European conference on machine learning and knowledge discovery in databases, Springer. pp. 387–402.

Bressan, G. M., de Azevedo, B. C. F., & de Souza, R. M. (2020). A fuzzy approach for diabetes mellitus type 2 classification. Brazilian Archives of Biology and Technology, 63. https://doi.org/10.1590/1678-4324-2020180742

C. Yan, L. Meng, L. Li, et al., "Age-invariant face recognition by multi-feature fusion and decomposition with self-attention," ACM Transactions on Multimedia Computing, *Communications, and Applications (TOMM)* 18(1s), 1–18 (2022).

Chen, S., Liu, Y., GAO, X., & Han, Z. (2018, August). Mobilefacenets: Efficient cnns for accurate real-time face verification on mobile devices. In Chinese Conference on Biometric Recognition (pp. 428-438). Springer, Cham.

Deb, D., Zhang, J., Jain, A.K., 2020. Advfaces: Adversarial face synthesis, in: 2020 IEEE International Joint Conference on Biometrics (IJCB), *IEEE*. pp. 1–10.

Deng, J., Guo, J., Xue, N., Zafeiriou, S., 2019. Arcface: Additive angular margin loss for deep face recognition, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4690–4699.

Devi, U. R., & Uma, K. (2019). A Study on Fuzzy Expert System for Diagnosis of Diabetes Mellitus. International *Journal of Applied Engineering Research* (IJAER), 14(4), 129–139. https://www.ripublication.com/ijaerspl2019/ijaerv14n4spl_16.pdf

Dong, Y., Liao, F., Pang, T., Su, H., Zhu, J., Hu, X., Li, J., 2018. Boosting adversarial attacks with momentum, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 9185–9193.

Geman, O., Chiuchisan, I., & Toderean, R. (2017). Application of Adaptive Neuro-Fuzzy Inference System for diabetes classification and prediction. 2017 E-Health and Bioengineering Conference, EHB 2017, Dm, 639–642. https://doi.org/10.1109/EHB.2017.7995505

Goswami, G., Ratha, N., Agarwal, A., Singh, R., & Vatsa, M. (2018, April). Unravelling robustness of deep learning-based face recognition against adversarial attacks. *In Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1).

H. Qiu, D. Gong, Z. Li, et al., "End2end occluded face recognition by masking corrupted features," IEEE Transactions on Pattern Analysis and Machine Intelligence (2021).

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T. & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.

Howsalya Devi, R. D., Bai, A., & Nagarajan, N. (2020). A novel hybrid approach for diagnosing diabetes mellitus using farthest first and support vector machine algorithms. Obesity Medicine, 17, 100152. https://doi.org/10.1016/j.obmed.2019.100152

Huang, G.B., Mattar, M., Berg, T., Learned-Miller, E., 2008. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments, in: Workshop on faces in'Real-Life'Images: detection, alignment, and recognition.

I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial exam-476 ples," arXiv preprint arXiv: 1412.6572 (2014).

Khalil, R. M., & Al-Jumaily, A. (2017). Machine learning based prediction of depression among type 2 diabetic patients. Proceedings of the 2017 12th International Conference on Intelligent Systems and Knowledge Engineering, ISKE 2017, 2018-Janua, 1–5. https://doi.org/10.1109/ISKE.2017.8258766

Khan, T., Singh, K., Manjul, M., Ahmad, M. N., Zain, A. M., & Ahmadian, A. (2022). A Temperature-Aware Trusted Routing Scheme for Sensor Networks: Security Approach. Computers & Electrical Engineering, 98, 107735.

Kumar, A., Singh, K., & Khan, T. (2021). L-RTAM: Logarithm based reliable trust assessment model for WBSNs. *Journal of Discrete Mathematical Sciences and Cryptography*, 24(6), 1701-1716.

Kumar, A., Singh, K., Khan, T., Ahmadian, A., Saad, M. H. M., & Manjul, M. (2021). ETAS: an efficient trust assessment scheme for BANs. IEEE Access, 9, 83214-83233.

Kurakin, A., Goodfellow, I., Bengio, S., et al., 2016. Adversarial examples in the physical world.

Liu, W., Wen, Y., Yu, Z., Li, M., Raj, B., Song, L., 2017. Sphereface: Deep hypersphere embedding for face recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 212–220.

Lukmanto, R. B., Suharjito, Nugroho, A., & Akbar, H. (2019). Early detection of diabetes mellitus using feature selection and fuzzy support vector machine. *Procedia Computer Science*, 157, 46–54. https://doi.org/10.1016/j.procs.2019.08.140

Mahmood Sharif, Sruti Bhagavatula, Lujo Bauer, and Michael K Reiter. Accessorize to a crime: Real and stealthy attacks on state-of-the-art face recognition. In Proceedings of the 2016 acm sigsac conference on computer and communications security, pages 1528–1540, 2016.

Massoli, F.V., Falchi, F., Amato, G., 2020. Cross-resolution face recognition adversarial attacks. *Pattern Recognition Letters* 140, 222–229.

Niswati, Z., Mustika, F. A., & Paramita, A. (2018). Fuzzy logic implementation for diagnosis of Diabetes Mellitus disease at Puskesmas in East Jakarta. Journal of Physics: Conference Series, 1114(1). https://doi.org/10.1088/1742-6596/1114/1/012107

Raj, R. S., Sanjay, D. S., Kusuma, M., & Sampath, S. (2019). Comparison of Support Vector Machine and Naïve Bayes Classifiers for Predicting Diabetes. 1st International Conference on Advanced Technologies in Intelligent Control, Environment, Computing and Communication Engineering, ICATIECE 2019, 41–45. https://doi.org/10.1109/ICATIECE45860.2019.9063792

Rajeswari, A. M., Sidhika, M. S., Kalaivani, M., & Deisy, C. (2018). Prediction of Prediabetes using Fuzzy Logic based Association Classification. Proceedings of the International Conference on Inventive Communication and Computational Technologies, ICICCT 2018, Icicct, 782–787. https://doi.org/10.1109/ICICCT.2018.8473159

Rozsa, A., Günther, M., & Boult, T. E. (2017, October). LOTS about attacking deep features. In 2017 IEEE International Joint Conference on Biometrics (IJCB) (pp. 168-176). IEEE.

Sarwar, M. A., Kamal, N., Hamid, W., & Shah, M. A. (2018). Prediction of diabetes using machine learning algorithms in healthcare. ICAC 2018 - 2018 24th IEEE International Conference on Automation and Computing: Improving Productivity through Automation and Computing, September, 1–6. https://doi.org/10.23919/IConAC.2018.8748992

Schroff, F., Kalenichenko, D., Philbin, J., 2015. Facenet: A unified embedding for face recognition and clustering, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 815–823.

Sharif, M., Bauer, L., Reiter, M.K., 2018. On the suitability of lp-norms for creating and preventing adversarial examples, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 1605–1613.

Sharif, M., Bhagavatula, S., Bauer, L., Reiter, M.K., 2016. Accessorize to a crime: Real and stealthy attacks on state-of-the-art face recognition, in: Proceedings of the 2016 acm sigsac conference on computer and communications security, pp. 1528–1540.

Swain, A., Mohanty, S., & Das, A. (2013). COMPARATIVE RISK ANALYSIS ON PREDICTION OF DIABETES MELLITUS USING MACHINE LEARNING APPROACH. 1, 3312–3317. International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) - 2016

Szegedy, W. Zaremba, I. Sutskever, et al., "Intriguing properties of neural networks," arXiv474 preprint arXiv: 1312.6199 (2013).

Thakkar, H., Shah, V., Yagnik, H., & Shah, M. (2021). Comparative anatomization of data mining and fuzzy logic techniques used in diabetes prognosis. *Clinical EHealth*, 4, 12–23. https://doi.org/10.1016/j.ceh.2020.11.001

Turk, M.A., Pentland, A.P., 1991. Face recognition using eigenfaces, in: Proceedings. 1991 IEEE computer society conference on computer vision and pattern recognition, *IEEE Computer Society*. pp. 586–587.

Undre, P., Kaur, H., & Patil, P. (2015). Improvement in prediction rate and accuracy of diabetic diagnosis system using fuzzy logic hybrid combination. 2015 International Conference on Pervasive Computing: Advance Communication Technology and Application for Society, ICPC 2015, 00(c). https://doi.org/10.1109/PERVASIVE.2015.7087029

Verma, D, M. N. (2017). using Data mining classification Techniques. 2017 International Conference on Intelligent Sustainable Systems (ICISS), Iciss, 533–538.

Vijiyakumar, K., Lavanya, B., Nirmala, I., & Sofia Caroline, S. (2019). Random forest algorithm for the prediction of diabetes. 2019 IEEE International Conference on System, Computation, Automation and Networking, ICSCAN 2019, 1–5. https://doi.org/10.1109/ICSCAN.2019.8878802

Wang, H., Wang, Y., Zhou, Z., Ji, X., Gong, D., Zhou, J., Li, Z., Liu, W., 2018. Cosface: Large margin cosine loss for deep face recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 5265–5274.

X. Wei, Y. Guo, and J. Yu, "Adversarial sticker: A stealthy attack method in the physical world," IEEE Transactions on Pattern Analysis and Machine Intelligence, 1–1 (2022).

Y. Dong, H. Su, B. Wu, et al., "Efficient decision-based black-box adversarial attacks on face recognition," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 7714–7722 (2019).

Y. Dong, Q.-A. Fu, X. Yang, et al., "Benchmarking adversarial robustness on image classification," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 321–331 (2020).

Y. Li, X. Yang, B. Wu, et al., "Hiding faces in plain sight: Disrupting ai face synthesis with 484 adversarial perturbations," arXiv preprint arXiv: 1906.09288 (2019).

Yang, L., Song, Q., Wu, Y., 2021. Attacks on state-of-the-art face recognition using attentional adversarial attack generative network. Multimedia Tools and Applications 80, 855–875.

Zhang, K., Zhang, Z., Li, Z., Qiao, Y., 2016. Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters 23, 1499–1503.

Zhong, Y., Deng, W., 2020a. Towards transferable adversarial attack against deep face recognition. IEEE Transactions on Information Forensics and Security 16, 1452–1466.

Zhou, Z., Tang, D., Wang, X., Han, W., Liu, X., Zhang, K., 2018. Invisible mask: Practical attacks on face recognition with infrared. arXiv preprint arXiv:1803.04683.

Bibliographic information of this paper for citing:

Agrawal, Khushabu & Bhatnagar, Charul (2023). F-MIM: Feature-based Masking Iterative Method to Generate the Adversarial Images against the Face Recognition Systems. *Journal of Information Technology Management*, 15 (Special Issue), 80-93. https://doi.org/10.22059/jitm.2023.95247