

Received 4 July 2024, accepted 25 July 2024, date of publication 29 July 2024, date of current version 12 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3435527

RESEARCH ARTICLE

Enhancing Security in Multimodal Biometric Fusion: Analyzing Adversarial Attacks

SHAIMA M. ALGHAMDI¹, SALMA KAMMOUN JARRAYA¹, AND FARIS KATEB²

¹Computer Science Department, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia

²Information Technology Department, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia

Corresponding author: Shaima M. Alghamdi (saliaalghamdi0006@stu.kau.edu.sa)


This work was supported in part by the Institutional Fund Project IFPIP:671-612-1443; and in part by the Ministry of Education and King Abdulaziz University, Deanship of Scientific Research (DSR), Jeddah, Saudi Arabia.

ABSTRACT Biometric recognition has become essential for secure and reliable access control in high-security systems such as surveillance, law enforcement, and smart cities. While deep learning models offer exceptional performance in biometric recognition, they are susceptible to security challenges such as adversarial attacks. Current research addresses the vulnerability of single modalities systems. However, there is a gap in understanding the impact of adversarial attacks on fusion levels in multimodal biometric systems. This research aims to fill this gap by thoroughly assessing the vulnerability of multimodal biometric systems to adversarial attacks to enhance their security in real-world high-security applications. This research investigates the most secure fusion level for combining behavioral and biological biometric traits in multimodal biometric authentication systems under adversarial attacks. We assess the security of different fusion levels by employing the Fast Gradient Sign Method (FGSM). Also, our study contributes to the field by evaluating the extent of perturbations needed to generate effective adversarial attacks and identifying the fusion level that offers the highest security under different perturbations. Our experiments thoroughly analyze multimodal fusion levels using attack success rate, accuracy, precision, and recall under clean and adversarial data conditions. According to our results, the input fusion level offers the most secure level among the three fusion levels. In various adversarial attack settings, it demonstrates an average attack success rate of 16.62% on DenseNet201 and 32.30% on ArcFace architectures. This research presents an important analysis to support further investigations into the security of multimodal biometric systems and building defense methods for such systems.

INDEX TERMS Adversarial attacks, FGSM, fusion levels, multimodal biometric recognition, surveillance.

I. INTRODUCTION

Biometrics recognition explores and analyzes human traits to provide secure, reliable, and accurate access control to reduce the reliance on less secure and traditional methods such as passwords. Such traits are unique and measurable for each individual. It ranges from biological that rely on static physical characteristics (e.g., face, iris, and fingerprints) to behavioral traits that identify individuals based on activity patterns (e.g., gait and keystrokes). As biometric systems

The associate editor coordinating the review of this manuscript and approving it for publication was Zhe Jin .

have a wide range of applications across various sectors such as law enforcement [1], education [2], healthcare [3] smart cities surveillance [4], and banking [5].

With the widespread use of deep learning, it is becoming increasingly common and used in different applications. It has significantly improved the performance of biometric systems. Convolutional neural networks (CNNs) have improved the feature extraction for image-based biometrics. Multimodal biometric systems that combine multiple biological and behavioral characteristics for identification offer improved performance over single-modality systems, especially in environments where a single biometric might be

affected by environmental conditions or insufficient. Recent research has demonstrated that these systems enhance biometric authentication performance and strengthen resilience against security limitations that makes them essential for high-security settings [6], [7].

While biometric recognition provide enhanced performance and overall better security compared to traditional measures, they are vulnerable to adversarial attacks. Recent research shows that machine learning and deep learning models are exposed to new vulnerabilities, known as adversarial attacks. Adversarial attacks add a slight and often imperceptible modification of the original data that alters the data to deceive systems in critical ways. These attacks pose serious security threats by undermining the reliability and integrity of biometric authentication systems. Adversarial attacks craft inputs designed to cause the model to make a mistake [8]. These inputs exploit how models process data by using the model's parameters against it. For example, subtle perturbations in facial recognition data can make a recognition system misidentify. As noted in foundational research by Szegedy et al. and Goodfellow et al., these vulnerabilities can be exploited to bypass security systems discreetly. It makes investigating the development of adversarial defense a critical area of future research [9], [10]. It is important to balance security with privacy concerns. Due to the sensitivity of biometric data, its protection from unauthorized access is a top priority [11], [12].

This research focuses on face and gait modalities because they non-intrusively capture data from a distance without direct contact. Such modalities suit continuous authentication in public or work settings. They cover biological and behavioral traits (i.e., rich facial features and unique walking patterns).

Despite the advancements in multimodal biometric authentication, a significant research gap remains in understanding the vulnerabilities associated with the fusion levels. Current literature extensively covers the susceptibilities of individual biometric traits to adversarial attacks but pays limited attention to how these vulnerabilities affect the fusion levels. This gap is essential to investigate because the fusion process on different levels (i.e., input, feature, or decision levels) can introduce unique security challenges such as adversarial attacks, spoofing, and evasion attacks specifically designed to exploit the fusion logic of multimodal authentication not present in single-modal systems.

Based on our previous research that implements three fusion levels on the face and gait modalities using ArcFace and DensNet201 deep learning architectures. This research aims to assess the security of multimodal biometric authentication systems against adversarial attacks. Additionally, we will conduct thorough evaluations of the three fusion-level performances using a variety of metrics to establish a baseline performance. Moreover, we will generate and analyze adversarial images to identify the most secure fusion level. Finally, the research assesses the impact of the attacks'

perturbations on each fusion level's behavior to distinguish adversarial attacks. This paper extends our research to shed light on the resilience of the three fusion levels (input, feature, and decision).

Our contributions are summarized as follows:

- 1) **Identify the most secure fusion level** for combining face and gait modalities in multimodal biometric authentication systems. By employing FGSM adversarial attacks to uncover vulnerabilities to different fusion levels.
- 2) **Assessing and identifying the range of perturbations** that make adversarial attacks highly successful and imperceptible to users and detection methods.

II. LITERATURE REVIEW

Adversarial attacks involve creating input data modified from the original data to fool the deep learning model into misclassification or mispredictions. The modifications on the input can be imperceptible to the human eye but lead to significant errors in the deep learning models' outputs. An adversarial image refers to the modified version of an original image by altering the image through, e.g., the addition of calculated perturbation to deceive machine learning models. An impersonation attack is a type of security threat where the attacker manipulates the biometric images to mimic another person's identity (person with authorized access) to trick biometric recognition systems into granting access to the impersonated person. Adversarial attacks can be classified using different aspects, such as attack capacity and specificity. The attack capacity refers to the attacker's knowledge of the target model under attack. In attack capacity, the attacks are either white-box where the attacker has in-depth knowledge of the model's architecture and parameters. White-box attacks enable the attacker to tailor their inputs to exploit specific weaknesses. On the other hand, black-box attacks occur when the attacker lacks insight into the model's parameters and details. This leads to attacks being based on the observable outputs of the model. Attack specificity refers to the attack's objective. It is categorized into targeted and non-targeted attacks. Targeted attacks focus on manipulating the model to produce a particular incorrect result (impersonation attack). Non-targeted attacks generate errors that lead the model to misidentification [8] and [13].

This section reviews the adversarial attacks, specifically on the face and gait separately, and the security challenges of fused modalities on different fusion levels. Additionally, we shed light on current research's limitations and potential investigation areas and present our contribution relative to the research gap.

A. ADVERSARIAL ATTACKS ON FACE RECOGNITION

The investigation of deep learning vulnerability started with Szegedy et al. [9] pointing out that deep learning models can be deceived into misclassifying images with

high confidence by adding perturbation to images without disrupting the original appearance of the image. Moreover, this misclassification happens across different classifiers with different architectures and is trained on various datasets. Goodfellow et al. [10] further investigated the deep neural network susceptibility to adversarial samples in relation to the linearity of the models. They proposed FGSM, a one-step algorithm that leverages the gradients of the models to create adversarial images. Since this discovery, other researchers have proposed various ways to craft adversarial attacks. Ryu et al. [14] proposed generating adversarial examples by attaching noise markers and adjusting their colors and locations to facial modalities with the aim of fooling deep learning models. Additionally, they investigated the attack's transferability on two face recognition models. They measured the attack's effectiveness using attack success rates with variations to the generation method settings, such as the marker's location and color adjustments. Massoli et al. [15] generate two classes of adversarial attacks, employing various methods to determine which group of adversarial attacks poses a significant threat to the face recognition model. In addition, the authors investigated the generalizability of the detection method to different groups of attacks. The first type of attack is a classification attack targeting the class labels generated using three algorithms: Momentum Iterative Fast Gradient Sign Method [16], Basic iterative method [10], Carlini-Wagner [17]. Moreover, this attack assesses the detection method proposed in their previous work in [18]. The other type is the deep representation attack; unlike the classification type, this type targets the representation to deceive the face recognition generated using the distance between the feature extracted by ResNet50 with the assistance of the KNN algorithm in the optimization process. Kakizak and Yoshida [19] address generating adversarial examples with large perturbations to deceive the current defense techniques focused on small perturbations. The authors use image translation to translate the source image to an unrestricted adversarial example (i.e., not restricted to small perturbation and not noticeable to the human eye), which was first proposed by [20]. Zhu et al. [21] employ GANs as a generation method with large distortion. Using CycleGan [22], the authors transfer makeup effects on face images. Moreover, they use a designated adversarial attack GAN subnetwork to generate the examples after the makeup transfer. The research focus on the balance between maintaining the source image identifiability and adding sufficient perturbations. Another research that exploits GANs by Yang et al. [23] in impersonation attacks. The proposed method Attentional Adversarial Attack Generative Network (A3GN) generates images similar to the source while holding the feature representation of the target images. Lin et al. [24] propose adversarial attack generation for the adversarial images with perturbations that are imperceptible to human eyes and can fool deep learning-based face recognition models. The proposed generation method leverages facial landmark detection and superpixel

segmentation to target several deep learning models at different attack capacities, such as white-box and black-box attacks. Bisogni et al. [25] explore the vulnerabilities of deep learning models against cross-spectral adversarial attacks, which sheds light on their limitations against attacks that are not limited to the spectral domain in which they were trained. Massoli et al. [26] examine adversarial attacks on face recognition in cross-resolution to mimic real-world conditions. The research shows that the success rate of adversarial attacks varies with image resolution. Lower-resolution images are more susceptible to attacks. Cross-resolution model demonstrates greater resilience to attacks compared to the base models.

B. ADVERSARIAL ATTACKS ON GAIT RECOGNITION

Maqsood et al. [27] proposed a new detection approach for deep learning-based gait recognition against adversarial attacks using reinforcement learning (Q-learning). Moreover, it targets patch-based black-box adversarial attacks. The authors leverage reinforcement learning to determine the optimal placement of an image patch that fools the model. He et al. [28] exploit GANs to target sequence-based gait recognition using state-of-the-art models. They generate adversarial attacks by inserting a few adversarial gait silhouettes into the original sequence to balance imperceptibility and achieve high attack success rates. Meijuan et al. [29] proposed a novel GAN-based attack generation approach that generates realistic videos and preserves discriminative details to deceive gait recognition systems. The proposed method renders new adversarial samples combining the source foreground and silhouette with the target's background by applying mask R-CNNs. The evaluation uses two protocols for the target scene and background from the same dataset or cross datasets. The authors measure the attack's recognition rate (i.e., attack success rate) in which the model is deceived into misidentification to identify the target individual. Honghao et al. [30] use the shadow model to mimic the target gait with the goal of deceiving the recognition model to misclassifying the gait of an attacker as that of a legitimate user (increasing false acceptance rate) or to failing to recognize a legitimate user's gait (increasing false rejection rate).

Table 1 provides an overview of the modalities, attack generation, datasets, targeted models, and evaluation key findings in the discussed studies with the aim of understanding the varied approaches applied in adversarial attack generation for deep learning face and gait recognition.

C. SECURITY CHALLENGES OF MULTIMODAL BIOMETRIC FUSION

This section broadly explores the security challenges associated with biometric fusion across various modalities, extending beyond face and gait. Jomaa et al. [31] explore the fusion of fingerprint and heart signal to improve the security of biometric systems at the decision level against presentation attacks. The authors utilize a multilayer authentication

approach by employing a fine-tuned CNN to distinguish between real and spoofed biometrics. Their novel method combines these two modalities to effectively decrease false match rates and false non-match rates. It demonstrates improved performance in preventing various presentation attacks.

Cui et al. [32] investigate the resilience of large multimodal models at feature and decision levels to image-based adversarial attacks across various tasks like image classification, image captioning, and Visual Question Answering. It shows vulnerability to visual adversarial inputs.

Yin et al. [33] explores adversarial attacks specifically designed for vision-language tasks at feature and decision levels. It exploits vulnerabilities in pre-trained models. The study demonstrates that these multimodal systems can be compromised through carefully crafted adversarial inputs that disrupt the model's ability to interpret visual or textual information correctly.

Despite the considerable advancements in developing techniques for generating and defending against adversarial attacks for single modalities, we still need to investigate how these models and fusion levels perform under attacks that combine different modalities. This limitation presents a crucial area for highlighting the need for investigation to ensure security across different fusion levels.

III. METHODOLOGY

This section will cover our multimodal face and gait biometric authentication on the three commonly used fusion levels: input, feature, and decision fusion, to help shed light on their security issues. Additionally, we explain the attack generation method used and its parameters.

A. MULTIMODAL FACE AND GAIT IDENTIFICATION

Multimodal biometric Identification by integrating features face and gait has been shown to improve the authentication performance and overcome the limitations of single-modalities [34], [35], [36], [37]. This research contributes to filling the existing gaps in the literature by introducing fusion at various levels to further elevate biometric authentication systems' reliability.

In this research, we adopted state-of-the-art architectures for multimodal biometric authentication. Arcface [38] with ResNet34 [39] and DensNet201 [40] are the backbone architectures to capture necessary biometric features for the identification. In face identification, we used CNN-based Max-Margin Object Detection to detect facial features due to its reliable performance [41]. It employs a margin-based loss function that ensures the detected features are distinctively separated in the feature space. This detection method improves the discriminative power of the model and aims to handle diverse environmental challenges and facial expressions in real-world applications.

We employed Gait Energy Images (GEIs) in gait identification introduced by [42] to represent gait modality. They are a compact representation of gait data. They are

generated by averaging silhouettes of a person captured over a complete walking cycle. It starts by segmenting the silhouette of the walking figure from the background in video frames. Secondly, aligning and normalizing the silhouettes across the gait cycle. Finally, averaging the frames to produce a single image that captures the dynamic motion pattern of an individual's gait. Such representation is valuable in gait identification due to their ability to encapsulate motion dynamics in a highly descriptive and computationally efficient form utilized in various applications [43], [44], [45].

Combining both face and gait modalities comprises an improved method for augmenting the efficacy and dependability of authentication systems. This following sections will detail the three discrete fusion levels: input-level, feature-level, and decision-level [46]. Each level maximizes the respective modalities' inherent advantages to optimize identification performance.

1) INPUT FUSION

At the input-level, our fusion process merges biometric data from disparate sources during the first stages of data processing prior to any feature extraction. This stage involves the creation of a dataset representation of the subjects facial images and GEIs through a vertical stacking [37]. This approach is depicted in Figure 1. It demonstrates the combined data constructs a unified representation to exploit the complementary nature of the spatial features inherent in each modality. The newly created unified representation is later fed into two distinct deep learning architectures (i.e., ArcFace and DenseNet201) for extracting complex features for the identification.

2) FEATURE FUSION

Feature fusion combines multiple feature vectors from different modalities to create a unified data representation using each feature set. We perform feature fusion by combining two architectures and adopting attention methods, as shown in Figure 1. Attention models are used in various applications and have proven effective for prioritizing essential features such as [47]; the authors propose the Multiattention-Net model. It uses modified squeezed residual blocks to refine feature representation significantly. Zhang et al. [48] introduce a self-attention mechanism into the convolutional GANs (SAGAN) model. This allows the GAN to model long-range dependencies better and improve the generated image quality. It generates details at every image location by considering information from the entire image. It leads to better feature fusion in different parts of the image. Dai, et al. [49] proposes a novel scheme called Attentional Feature Fusion (AFF) to improve the performance of various neural networks by addressing challenges in feature fusion across different scales and layers. They introduce a Multi-Scale Channel Attention Module (MS-CAM) to manage semantic and scale inconsistencies better. In this research, the

TABLE 1. Adversarial attack research for face and gait biometric recognition systems.

Ref	Modality	Attack Generation	Datasets	Target Model	Evaluation
Ryu et al. [14]	Face	Noise Markers	Collected from seven volunteers and K-Face dataset	MobileNetV2, Inception-ResNetV2	Attack Success Rate: 4.76%-14.29%
Massoli et al. [15]	Face	L-BFGS, FGSM, BIM, MI-FGSM, Carlini-Wagner, and deep features attacks	VGGFace2 Test Set	Se-Net-50	Attack Success Rate: Targeted: 95.6%-96.3%. Un-targeted: 96.8%
Kakizak et al. [19]	Face	Image translation	VGGFace, VGGFace2, CelebA	VGG16 and ResNet50	Attack Success Rate: Targeted (white-box): 90%, Targeted (black-box): 80%
Zhu et al. [21]	Face	GAN-based subnetworks (Makeup Transfer and Adversarial Attack), FGSM, PGD	Collected Makeup Face Dataset	AlexNet, SqueezeNet, VGG16, ResNet50, InceptionV3, DenseNet121, and LightCNN29	Attack Success Rate Targeted GAN-based: AlexNet: 98.50%; SqueezeNet: 95.25%; VGG16: 100%; ResNet50: 90%; InceptionV3: 97%; DenseNet121: 95.50%; LightCNN29: 93.75%; Average: 95.78%; FGSM: Average on all models: 1.07%; PGD: Average on all models: 29.02%
Yang et al. [23]	Face	A3GN	CASIA-WebFace, LFW, CFP-FP, AgeDB-30	Softmax, SphereFace, and ArcFace	Fake Accuracy: 97.52% - 99.66%
Lin et al. [24]	Face	Novel generation method, FGSM, BIM, PGD	CASIA-WebFace	FaceNet, Inception-ResNet (IR152 and IRSE50), ArcFace	Average Attack Success Rate: FaceNet: (61.44% - 90.72%) IR152: (0.38% - 4.38%) IRSE50: (14.86% - 23.94%) ArcFace: (24.64% - 36.70%)
Bisogni et al. [25]	Face	Cross-spectral adversarial attacks, FGSM	HITSZ, VIS-TH	VGG16, VGG19, InceptionV3, DenseNet121, and Inception-ResNetV2	Post-attack accuracy (original spectrum): 2% - 4%, Transposed spectrum drops to nearly 0%
Massoli et al. [26]	Face	Jacobian saliency map attack, Elastic net attack, Carlini and wagner, and PGD	VGGFace2, SC-face	Backbone architectures ResNet-50 with Squeeze-and-Excitation unites for the base model and cross resolution model	Attack Success Rates: Low Resolution: nearly 100%
Maqsood et al. [27]	Gait (GEI)	Adversarial patch generation	CASIA-B	New model	Max success rate: 77.59%
Ziwen He et al. [28]	Gait (Sequence)	GAN	CASIA-B	GaitSet	Avg. success rate: 15.11% - 69.30%
Meijuan et al. [29]	Gait (Videos)	GAN	CASIA-A, CASIA-B	GaitSet, CNN-Gait	Attack success rate: GaitSet: 57.00% to 82.00% CNN-Gait: 63.00% to 94.00%

architectures extract face and gait features from both models: ArcFace for facial features and DenseNet201 for gait features. The extracted features are passed to the Squeeze-and-Excitation (SE) Block introduced in [50]. By incorporating SE blocks, we aim to contribute to a more robust biometric identification system. Our model's attention mechanism applies channel-wise attention to the feature vectors. The blocks squeeze global spatial information into a channel descriptor and then excite the original features by reweighting them using this descriptor. This process emphasizes the

essential features and suppresses the less useful ones. Thus enhancing the model's focus on relevant features. Subsequently, the gait features are fed to global average pooling. Both face and gait features are then enhanced using SE blocks. The outputs from the SE blocks are concatenated to form a single feature vector, creating a new feature set that harnesses the detailed information from both models. In addition, a dropout layer is used for model generalization. Finally, a dense softmax layer consolidates the identification process.

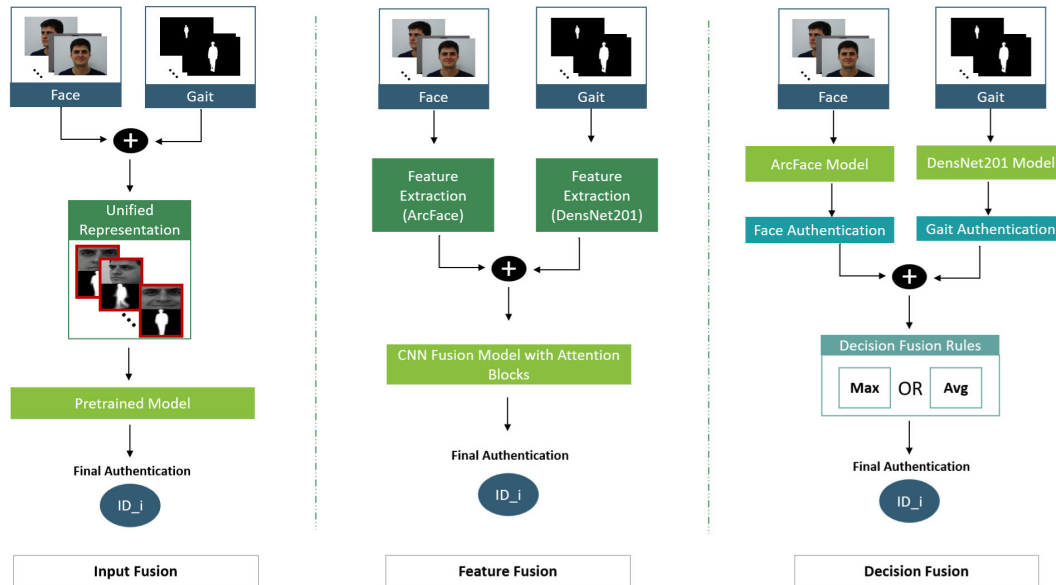


FIGURE 1. All fusion Levels: Combining two facial and GEI modalities on three fusion levels (input, feature and decision).

3) DECISION FUSION

Decision fusion combines the outcomes from multiple models. Each model trained on distinct modalities contributes to a final authentication decision. This practical application of decision fusion is essential as it maximizes the independent performance of each authentication model and leads to better overall performance in real-world biometric systems. Decision fusion is highly used in the multimodal biometric system. For instance, the study by Sivarathinabala et al. demonstrated the robustness of using multimodal features at the decision level to improve system reliability by combining face and gait data in a smart security setup [35]. Similarly, the approach by Maity et al. demonstrates the importance of decision-level fusion in achieving high accuracy and resilience against spoofing attacks in complex biometric systems [34]. In this research, we have used two separate models for our analysis: the face data is processed using the ArcFace model, and DenseNet201 handles the gait data. Once the face and gait data are individually processed and, the respective models generate their decisions. Next, it combines them to achieve a final identification result. To this end, we implement two fusion rules (i.e., maximum and average).

- 1) Maximum Rule: Select the decision with the highest confidence score from the models. It is beneficial in scenarios where one modality is significantly more reliable than others. Thus relying on the most accurate model available.
- 2) Average Rule: This rule computes the average of the confidence scores from both models. It offers a balance, mitigating outliers, and consistent authentication.

By comparing the outcomes of these two fusion rules, we aim to secure fusion for multimodal biometric authentication. Figure 1 visually represents the decision fusion level as described.

B. ADVERSARIAL ATTACKS GENERATION

Adversarial attacks introduce precisely crafted perturbations to input images leading the model to misclassification. The attacks threaten the reliability and security of deep learning models. Such attacks have various and evolving generation methods. In this research, we used the **Fast Gradient Sign Method (FGSM)** proposed by Goodfellow et al. [10]. FGSM algorithm is straightforward, simple, and computationally effective. By leveraging FGSM, we aim to quantify the susceptibility of these models and fusion levels to adversarial attacks. It exploits the gradients of the loss concerning the input image. Equation 1, explains how the perturbations are calculated. Figure 2 illustrates FGSM attack generation. It consists of a series of images arranged as follows:

- 1) Original Images (a): The leftmost images depict the initial inputs the model accurately classifies under normal conditions. These serve as the baseline for understanding the effect of the perturbations.
- 2) Perturbation Visualizations (b): The middle images showcase the perturbations. It is calculated as per Equation 1. These visualizations are important for understanding the magnitude and direction of the changes applied to the original images.
- 3) Adversarial Images (c): The rightmost images display the adversarial examples. The adversarial image is created by adding the calculated perturbations to the original images. These images demonstrate the significance of minimal changes that can cause misclassifications.

The sign function is applied to the gradient to determine the direction of the perturbation for each pixel, ensuring the perturbation is small and controlled by ϵ but effective in misclassifications.

$$X_{adv} = x + \epsilon \cdot \text{sign}(\nabla_x J(\theta, x, y)) \quad (1)$$

where x is the original image, ϵ is the perturbation magnitude, $\nabla_x J(\theta, x, y)$ the gradient of the loss, and y is the true label. Figure 3 demonstrates the impact of increasing epsilon values on the visibility of perturbations:

- **Target vs. Imposter Images:** Initially, a target image (the genuine person intended to be recognized) and an imposter image (a different individual) are presented. The imposter serves as the base for adversarial example generation.
- **Adversarial Examples Across Epsilon Values:** Subsequent images demonstrate the progression of adversarial examples generated from the imposter image with increasing values of ϵ (0.001, 0.01, 0.02, 0.05, and 0.1). Each image is a result of adding the perturbations calculated for the respective ϵ value, showcasing how higher values make perturbations more visible but also more effective at deceiving the model.

IV. EXPERIMENTS

This section covers the experimental setup in different aspects, such as datasets, a multimodal identification baseline on three fusion levels, and an adversarial attack configuration. It also covers the evaluation metric used to measure the vulnerability to the attacks. Finally, it covers the results and discussion of our findings.

A. EXPERIMENTS SET-UP

This section covers datasets description and preparation. In addition to the adversarial attack generation configuration.

1) DATASETS DESCRIPTION AND PREPARATIONS

Due to the limitation of publicly available datasets that combine both modalities, our approach leverages distinct datasets for each modality. To combine independent face and gait datasets, we integrate two separate face datasets: Cambridge Olivetti Research Lab (ORL) [51] and FEI face dataset [52] in addition to the gait Chinese Academy of Sciences (CASIA-B) dataset [53]. ORL includes 10 images for each of its 41 subjects, while the FEI dataset provides 14 images for each of its 200 subjects. CASIA-B has 124 subjects under variations in viewing angles, clothing, carrying bags, and walking speed. This approach allows us to cover facial and gait features by producing data from these independent sources. Each modality contributes unique information to create a dataset in which each subject has multiple biometric modalities. The mapping was performed on two levels. The first was subject-level mapping, in which we ensured a one-to-one mapping between the subjects from face and gait datasets. The second is instance-level mapping, where the datasets are not well-balanced enough to be mapped, which would introduce biases and affect the overall performance. To ensure equal representation, we used data augmentation to help achieve a more equitable representation of various traits across both modalities.

TABLE 2. Summary of multimodal datasets.

Multimodal Dataset	Original Dataset	No. of Subs	No. of Instances
FEI_CASIAB	FEI with CASIA-B	124	For each of the 110 GEI from the CASIA-B, we'll randomly select 98 GEI to map with the corresponding 98 image face images in the FEI dataset.
ORL_CASIAB	ORL with CASIA-B	41	Each of the 110 GEI from the CASIA-B dataset with the corresponding 110 face images in the ORL dataset

In preparation for creating impersonation adversarial images, we randomly paired target subjects with corresponding impersonators (target-imposter pair). The sample size of the multimodal datasets encompasses 50% of the available subjects. The potential impersonators are within the same dataset as the target in this setup. The objective is to manipulate a single image per pair.

2) ADVERSARIAL ATTACKS CONFIGURATION

This research employs a white-box attack approach in which the attacker fully knows the target model, including parameters and architecture. Additionally, our focus is targeted attacks (impersonation attacks) aiming to mislead the model into incorrectly assigning a chosen label to an adversarial image. As for the generation method, in our analysis of FGSM, we experimented with varying perturbation values (epsilons) to understand its impact on fusion levels.

B. EVALUATION METRICS

To assess the effectiveness of our multimodal biometric authentication model under normal and adversarial conditions, we employ several key performance metrics:

- **Accuracy:** This metric evaluates the overall correctness of the model by measuring the proportion of true results (both true positives and true negatives) among the total number of cases examined.
- **Precision:** Measures the ratio of correct positive observations to the total predicted positives. High precision indicates a low rate of false positives.
- **Recall:** Also known as sensitivity, this metric assesses the model's ability to correctly identify all actual positives for each class. This highlights its effectiveness in detecting genuine instances.
- **F1-Score:** The harmonic mean of precision and recall. It provides a single measure to balance both metrics.
- **Attack Success Rate:** Represents the proportion of successful adversarial attacks. It measures the model's vulnerability by calculating how often it incorrectly rejects an authorized user or accepts an imposter.

These metrics collectively offer a comprehensive overview of the model's performance and robustness. Moreover, it provides insights into authentication efficacy and resilience to adversarial attacks.

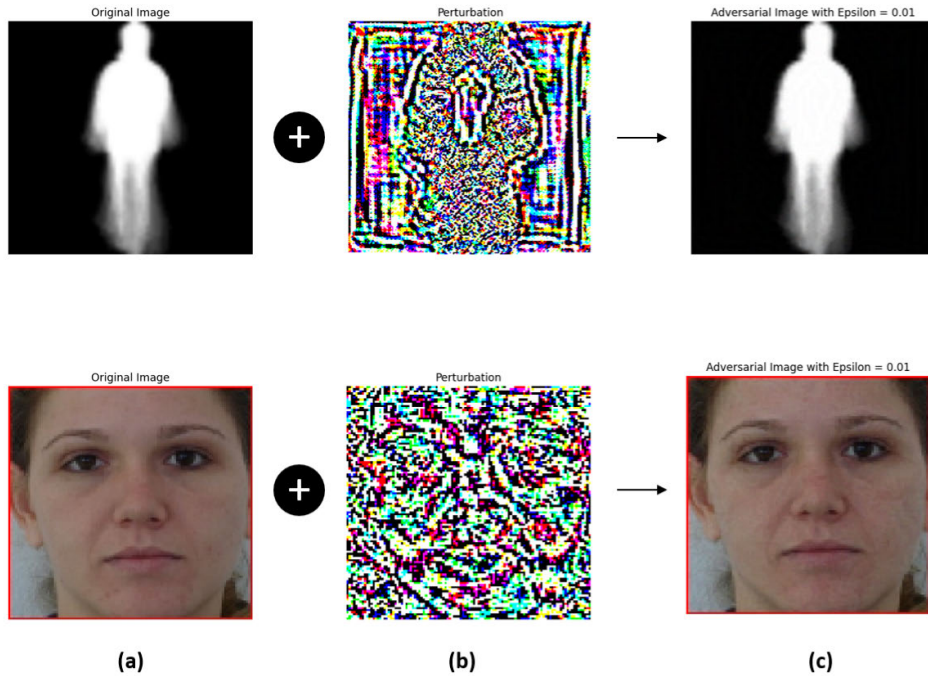


FIGURE 2. illustration of fgsm Adversarial Attack Generation.

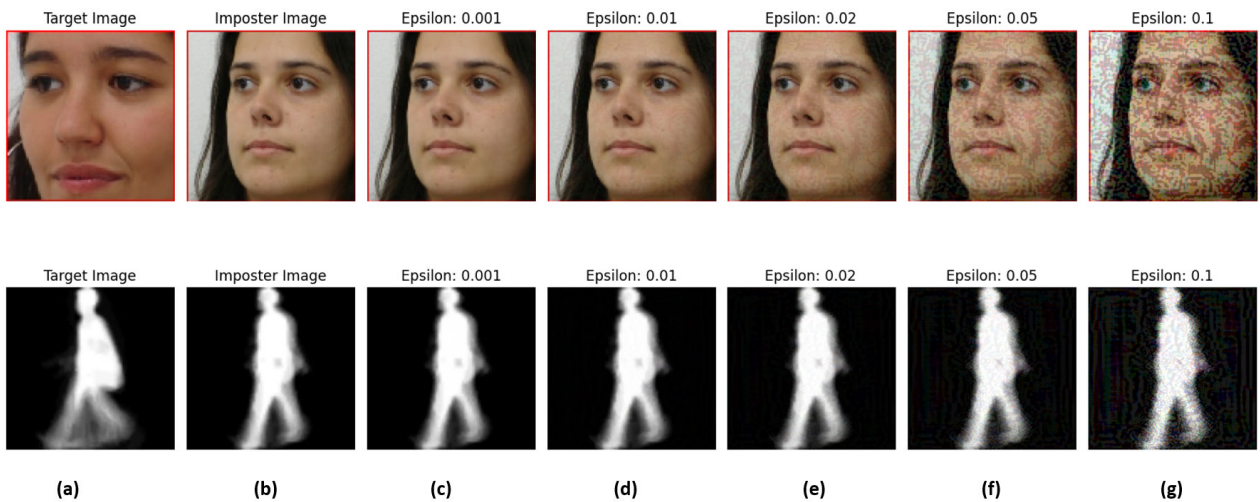


FIGURE 3. Comparative visualization of adversarial perturbations at various epsilon values.

C. RESULTS

In the initial phase of our study, we establish the fusion level performance baseline before the adversarial attacks to set a comparative standard against which the impact of subsequent adversarial attacks will be measured. Furthermore, we evaluate and compare each fusion level’s relative security and performance trade-offs. Additionally, we analyze the balance between security and performance. Moreover, we assess and analyze the impact of the perturbations on the fusion levels.

1) BASELINE PERFORMANCE FOR MULTIMODAL AUTHENTICATION

Figure 4 compares precision, recall, F1 score, and accuracy across these fusion levels. it illustrates each fusion level’s

performance under normal conditions to provide a baseline against which the impact of adversarial attacks can be measured. It compares the baseline and sample performance metrics across different fusion levels. The performance is consistent across different fusion levels. All metrics demonstrate high values. The data indicate that the system performs well under clean data conditions for each fusion.

2) MOST SECURE FUSION-LEVEL FOR MULTIMODAL AUTHENTICATION

In assessing the security of our multimodal authentication system against adversarial attacks, we use the success rate metric of these attacks across various fusion levels to indicate

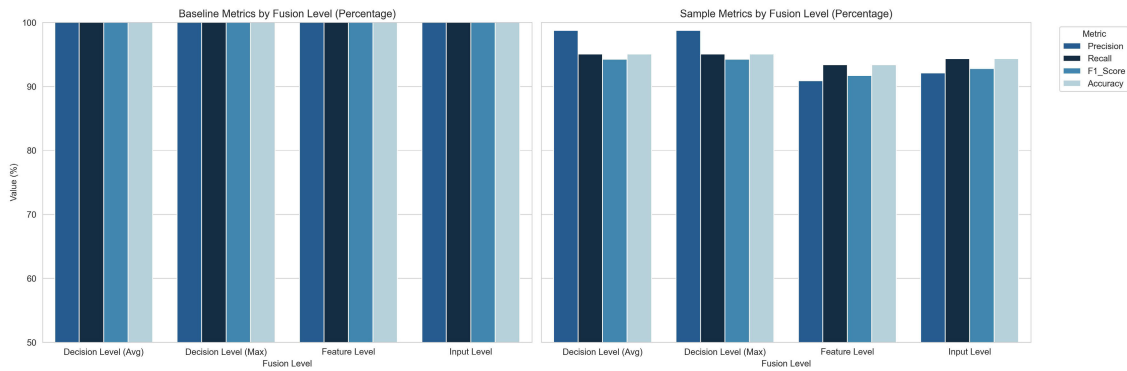


FIGURE 4. Baseline performance of the fusion models under clean data using the complete test dataset and on the sample before attack generation.

each fusion level’s vulnerability. As depicted in Figure 5, the FEI_CASIAB and ORL_CASIAB datasets were utilized to evaluate the average success rates of adversarial attacks targeting the input, decision (average and maximum), and feature fusion levels. The input fusion level demonstrated the lowest average success rate of adversarial attacks. This level provides a higher resilience against such perturbations. This robustness is attributed to the preprocessing and integration of raw data at this stage. Thus, it hinders the attacker’s ability to manipulate the system’s behavior effectively. Within the decision fusion, using the maximum rule resulted in a lower success rate of attacks compared to the average rule. This outcome highlights the maximum rule’s ability to rely on the most confident decisions. It reduces the influence of adversarial changes in the data to deceive the model. Contrarily, the feature level exhibited a significantly higher success rate for adversarial attacks. It indicates a vulnerability in how features are integrated and weighted. This level’s susceptibility suggests that the attack can manipulate the representation space where features from different modalities are combined. Therefore, it makes it easier to affect the authentication. The input and decision levels’ superior security performance (maximum rule) encourages their use in high-security scenarios.

Table 3 complements this analysis by showing the average success rates across the studied fusion levels and datasets.

Figure 6 compares the impact of adversarial perturbations across these fusion levels by varying the epsilon values, which dictate the magnitude of the perturbations. Input Level: Utilizes ArcFace and DensNet201_MLP. The graph compares the success rates on two datasets, FEI_CASIAB and ORL_CASIAB. Typically, the input level should show robustness to adversarial attacks since the perturbations would need to be significant to influence the combined raw data effectively. Feature Level: Employs an attention-enhanced model (ArcFace_DensNet201_Attention) and assesses its robustness on the same datasets. Given that feature-level fusion involves integrated feature sets, this level might show a higher susceptibility to attacks if adversarial perturbations manipulate the features effectively. Decision Level (Max and Avg): This uses ArcFace_DensNet201 and

assesses attack success rates on the same datasets. Decision-level fusion, depending on whether the Max or Avg rule is used, might exhibit different vulnerabilities to adversarial attacks. The Max rule tends to rely on the highest-confidence decision, possibly reducing susceptibility to perturbations that do not significantly affect the most confident outputs.

TABLE 3. Fusion levels attack success rate on different models and datasets.

Model	Dataset	Input	Feat	Dec (Max)	Dec (Avg)
ArcFace	FEI_CASIAB	31.61%			
DensNet201_MLP		13.23%			
ArcFace_DensNet Attention			84.19%		
ArcFace_DensNet201				84.51%	83.87%
ArcFace	ORL_CASIAB	33.00%			
DensNet201_MLP		20.00%			
ArcFace_DensNet Attention			91.00%		
ArcFace_DensNet201				71.00%	68.00%

3) MULTIMODAL BIOMETRIC AUTHENTICATION ACCURACY VS SECURITY TRADE-OFF

Figure 7 illustrates the trade-offs between the success rates of adversarial attacks and clean sample accuracy across different fusion levels within the multimodal authentication framework. The scatter plot outlines the relationship between the average success rate of adversarial attacks and the accuracy of clean data for each fusion level—input, feature, and decision—employing different models. Input Level Fusion (ArcFace, DensNet201_MLP): These points demonstrate a strong performance on clean data while maintaining a lower success rate for attacks. This suggests that the input level fusion has high security due to the complexity of merging unprocessed data. Decision Level Fusion (ArcFace_DensNet201 for both Avg and Max): Positioned with moderate attack success rates and reasonable clean data accuracy. It shows a balanced fusion trade-off after Input fusion. Feature Level Fusion: These points plot higher

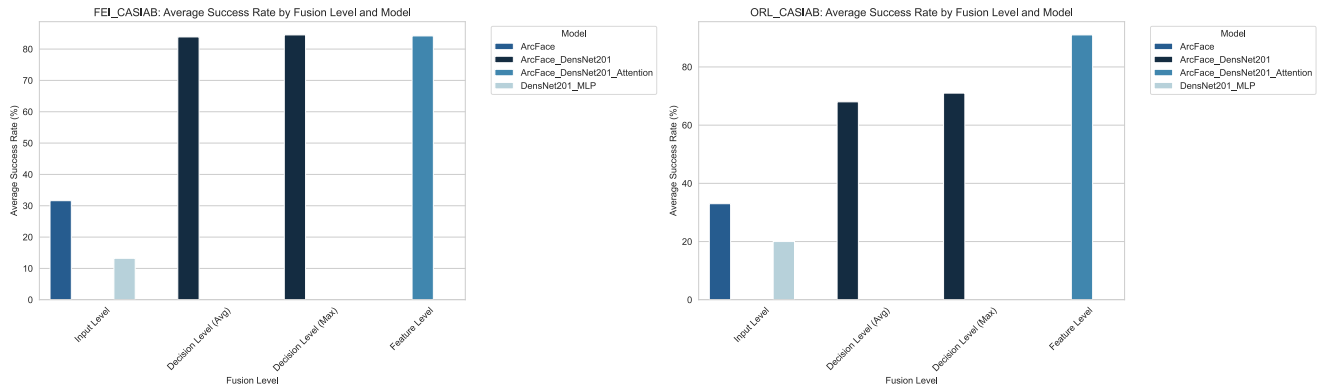


FIGURE 5. Average success rate for each fusion level.

Success Rate by Fusion Level, Model, and Epsilon

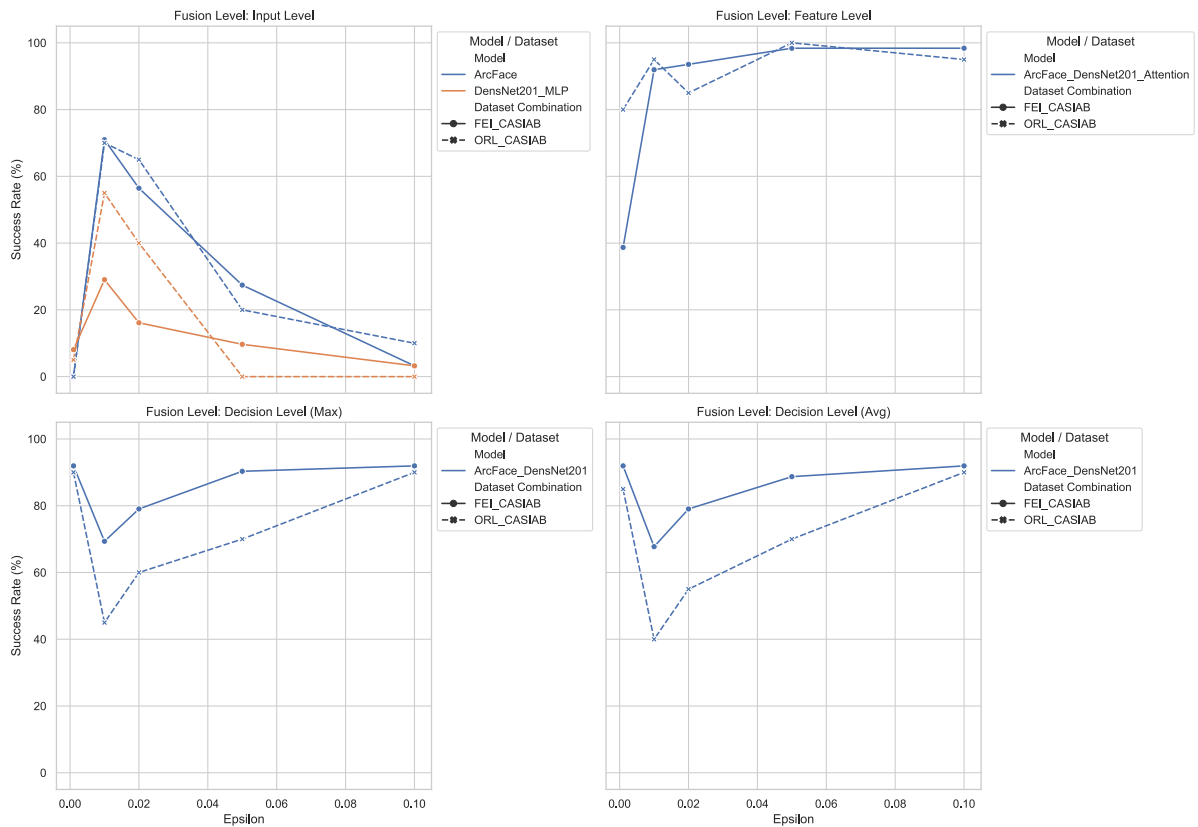


FIGURE 6. Detailed attack success rate for each fusion level, model and epsilon.

on clean sample accuracy and attack success rate. Despite high performance on clean samples, the elevated success rate of attacks indicates potential vulnerabilities.

4) IMPACT OF FGSM PERTURBATIONS ON FUSION-LEVELS

Figure 8 presents the differential impacts of increasing adversarial perturbation levels (epsilon) on the success rates across various fusion levels. Each fusion level exhibits distinct trends that reflect its inherent robustness or susceptibility to adversarial attacks. While some fusion levels, such as

the input fusion, demonstrate substantial resilience, others, particularly feature, and average-based decision fusions, reveal specific vulnerabilities that adversarial attacks could exploit. Input Level: This level consistently declines in attack success rate as epsilon increases, which is indicative of high robustness. The input level fusion’s ability to learn more complex representations that generalize effectively across different scenarios allows it to resist even significant adversarial perturbations effectively. This means that preprocessing and early integration of modalities at the input level provides a strong defense mechanism against

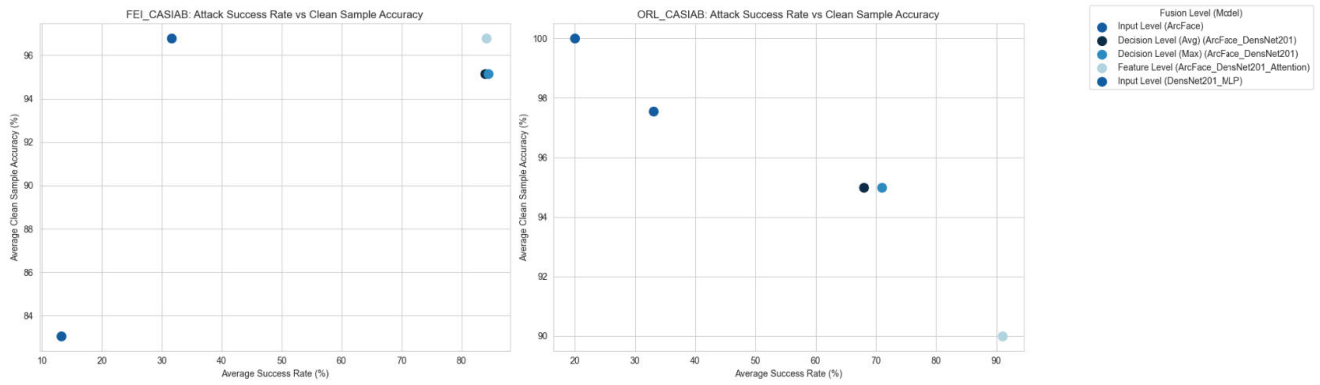


FIGURE 7. Balancing performance vs. security.

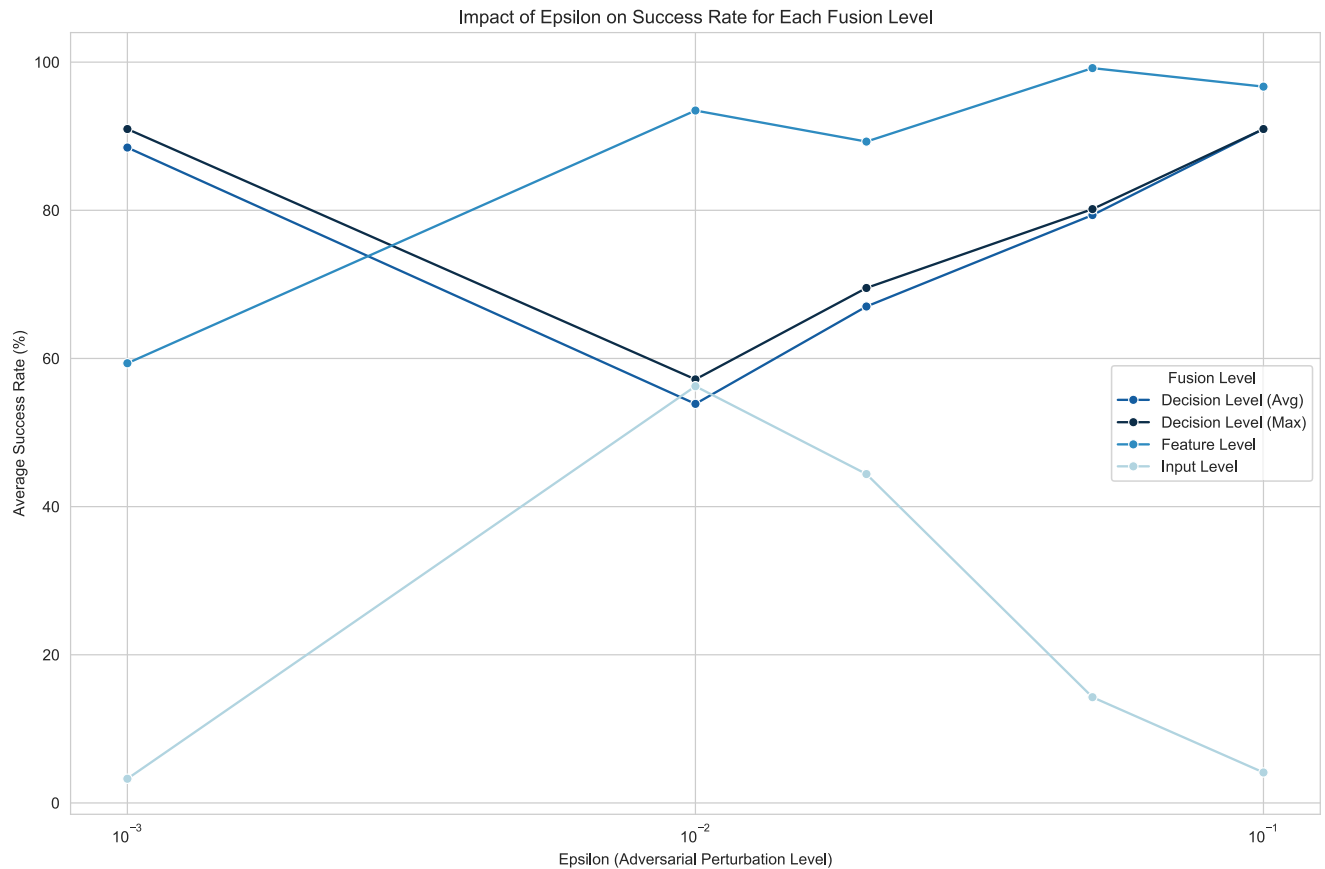


FIGURE 8. The impact of epsilon values on fusion levels.

adversarial influences. Feature Level: Unlike other levels, the feature level demonstrates an initial increase in attack success rate with moderate epsilon values, peaking before it plateaus. This behavior indicates a critical threshold beyond which further increases in perturbation do not significantly affect the system’s vulnerability. The initial increase in susceptibility might be attributed to merging features from different modalities, which, while enhancing performance under normal conditions, may also streamline the attack surface for adversarial exploits. Decision Level (max and average): This fusion strategy exhibits an intriguing trend where the success rate of attacks initially increases as epsilon

grows but then begins to stabilize. This suggests that while average-based decision fusion can initially accommodate mild perturbations, it becomes more susceptible as they become perceptible and impactful. Both decision fusion rules show an initial increase in success rate with epsilon, followed by a decline. This suggests an optimal point of vulnerability at a certain perturbation level beyond which the fusions start to respond to the attack more effectively.

D. COMPARISON WITH PREVIOUS RESEARCH

Previous research demonstrates high success rates of adversarial attacks on single-modality systems. Our research shows

that merging multiple biometric traits can achieve higher robustness against adversarial attacks compared to single-modality systems. The average attack success for input fusion using the ArcFace model and DensNet201 ranges between 13.23% and 31.61%. These rates are significantly lower compared to the highest single-modality attack success rates, such as Zhu et al.'s GAN-based attacks on VGG16 (93.75% - 100%) and Massoli et al.'s targeted attacks on Se-Net-50 (95.6% - 96.3%). This establishes that input-level fusion provides significant robustness against adversarial attacks. The Feature fusion model using ArcFace_DensNet Attention achieves an attack success rate of 84.19%. Its performance under attack still improves over some of the most effective attacks reported in the literature, such as the white box targeted attacks (90%) by Kakizak et al. At this stage of fusion, further enhancements are needed. As for decision fusion, it achieves 84.51% (Max) and 83.87% (Avg) success rates. These rates are comparable to high success rates in single-modality systems like Massoli et al.'s low-resolution attacks (nearly 100%). This indicates that while decision-level fusion is beneficial, further improvements are necessary to enhance robustness against adversarial attacks.

Our research stresses the vital role of multimodal fusion in strengthening the robustness of biometric systems against adversarial attacks. Compared to biometric systems in the literature, our fusion levels generally achieve lower attack success rates, indicating improved resilience.

V. CONCLUSION

This research addresses the vulnerability of multimodal biometric recognition across three fusion levels (input, feature, and decision) that combine behavioral and biological characteristics. It focused on identifying the most secure fusion levels by employing FGSM adversarial attacks to uncover vulnerabilities. In addition to assessing the impact of perturbations that make adversarial attacks successful. Our findings offer insights into the three fusion levels, where input fusion demonstrated superior performance under attack, to aid future research in building defence methods to enhance overall security in multimodal systems.

ACKNOWLEDGMENT

This work was supported in part by the Institutional Fund Project IFPIP:671-612-1443; and in part by the Ministry of Education and King Abdulaziz University, Deanship of Scientific Research (DSR), Jeddah, Saudi Arabia.

REFERENCES

- [1] F. Jansen, J. Sánchez-Monedero, and L. Dencik, "Biometric identity systems in law enforcement and the politics of (voice) recognition: The case of Süp," *Big Data Soc.*, vol. 8, no. 2, Jul. 2021, Art. no. 205395172110636.
- [2] M. Hernandez-de-Menendez, R. Morales-Menendez, C. A. Escobar, and J. Arinez, "Biometric applications in education," *Int. J. Interact. Des. Manuf. (IJIDeM)*, vol. 15, pp. 365–380, Sep. 2021.
- [3] J. Mason, R. Dave, P. Chatterjee, I. Graham-Allen, A. Esterline, and K. Roy, "An investigation of biometric authentication in the healthcare environment," *Array*, vol. 8, Dec. 2020, Art. no. 100042.
- [4] M. Kumar, K. S. Raju, D. Kumar, N. Goyal, S. Verma, and A. Singh, "An efficient framework using visual recognition for IoT based smart city surveillance," *Multimedia Tools Appl.*, vol. 80, no. 20, pp. 31277–31295, Aug. 2021.
- [5] P. M. A. B. Estrela, R. D. O. Albuquerque, D. M. Amaral, W. F. Giozza, and R. T. D. S. Júnior, "A framework for continuous authentication based on touch dynamics biometrics for mobile banking applications," *Sensors*, vol. 21, no. 12, p. 4212, Jun. 2021.
- [6] R. Ryu, S. Yeom, S.-H. Kim, and D. Herbert, "Continuous multimodal biometric authentication schemes: A systematic review," *IEEE Access*, vol. 9, pp. 34541–34557, 2021.
- [7] A. Tarannum, Z. U. Rahman, L. K. Rao, T. Srinivasulu, and A. Lay-Ekuakille, "An efficient multi-modal biometric sensing and authentication framework for distributed applications," *IEEE Sensors J.*, vol. 20, no. 24, pp. 15014–15025, Dec. 2020.
- [8] Y. Xu, K. Raja, R. Ramachandra, and C. Busch, "Adversarial attacks on face recognition systems," in *Handbook of Digital Face Manipulation and Detection*. Cham, Switzerland: Springer, 2022, pp. 139–161.
- [9] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus, "Intriguing properties of neural networks," in *Proc. 2nd Int. Conf. Learn. Represent.*, 2014, pp. 1–10.
- [10] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," *Stat.*, vol. 1050, p. 20, Jan. 2015.
- [11] C.-W. Lien and S. Vhaduri, "Challenges and opportunities of biometric user authentication in the age of IoT: A survey," *ACM Comput. Surv.*, vol. 56, no. 1, pp. 1–37, Jan. 2024.
- [12] A. Sarkar and B. K. Singh, "A review on performance, security and various biometric template protection schemes for biometric authentication systems," *Multimedia Tools Appl.*, vol. 79, nos. 37–38, pp. 27721–27776, Oct. 2020.
- [13] F. Vakhshiteh, A. Nickabadi, and R. Ramachandra, "Adversarial attacks against face recognition: A comprehensive study," *IEEE Access*, vol. 9, pp. 92735–92756, 2021.
- [14] G. Ryu, H. Park, and D. Choi, "Adversarial attacks by attaching noise markers on the face against deep face recognition," *J. Inf. Secur. Appl.*, vol. 60, Aug. 2021, Art. no. 102874.
- [15] F. V. Massoli, F. Carrara, G. Amato, and F. Falchi, "Detection of face recognition adversarial attacks," *Comput. Vis. Image Understand.*, vol. 202, Jan. 2021, Art. no. 103103.
- [16] Y. Dong, F. Liao, T. Pang, H. Su, J. Zhu, X. Hu, and J. Li, "Boosting adversarial attacks with momentum," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9185–9193.
- [17] N. Carlini and D. Wagner, "Towards evaluating the robustness of neural networks," in *Proc. IEEE Symp. Secur. Privacy (SP)*, May 2017, pp. 39–57.
- [18] F. Carrara, R. Becarelli, R. Caldelli, F. Falchi, and G. Amato, "Adversarial examples detection in features distance spaces," in *Proc. Eur. Conf. Comput. Vis. (ECCV) Workshops*, 2018, pp. 313–327.
- [19] K. Kakizaki and K. Yoshida, "Adversarial image translation: Unrestricted adversarial examples in face recognition systems," 2019, *arXiv:1905.03421*.
- [20] Y. Song, R. Shu, N. Kushman, and S. Ermon, "Constructing unrestricted adversarial examples with generative models," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 1–12.
- [21] Z.-A. Zhu, Y.-Z. Lu, and C.-K. Chiang, "Generating adversarial examples by makeup attacks on face recognition," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 2516–2520.
- [22] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2242–2251.
- [23] L. Yang, Q. Song, and Y. Wu, "Attacks on state-of-the-art face recognition using attentional adversarial attack generative network," *Multimedia Tools Appl.*, vol. 80, no. 1, pp. 855–875, Jan. 2021.
- [24] C.-Y. Lin, F.-J. Chen, H.-F. Ng, and W.-Y. Lin, "Invisible adversarial attacks on deep learning-based face recognition models," *IEEE Access*, vol. 11, pp. 51567–51577, 2023.
- [25] C. Bisogni, L. Cascone, J.-L. Dugelay, and C. Pero, "Adversarial attacks through architectures and spectra in face recognition," *Pattern Recognit. Lett.*, vol. 147, pp. 55–62, Jul. 2021.
- [26] F. V. Massoli, F. Falchi, and G. Amato, "Cross-resolution face recognition adversarial attacks," *Pattern Recognit. Lett.*, vol. 140, pp. 222–229, Dec. 2020.
- [27] M. Maqsood, S. Yasmin, S. Gillani, F. Aadil, I. Mehmood, S. Rho, and S.-S. Yeo, "An autonomous decision-making framework for gait recognition systems against adversarial attack using reinforcement learning," *ISA Trans.*, vol. 132, pp. 80–93, Jan. 2023.

- [28] Z. He, W. Wang, J. Dong, and T. Tan, "Temporal sparse adversarial attack on sequence-based gait recognition," 2020, *arXiv:2002.09674*.
- [29] M. Jia, H. Yang, D. Huang, and Y. Wang, "Attacking gait recognition systems via silhouette guided GANs," in *Proc. 27th ACM Int. Conf. Multimedia*, Oct. 2019, pp. 638–646.
- [30] H. Guo, Z. Wang, B. Wang, X. Li, and D. M. Shila, "Fooling a deep-learning based gait behavioral biometric system," in *Proc. IEEE Secur. Privacy Workshops (SPW)*, May 2020, pp. 221–227.
- [31] R. M. Jomaa, M. S. Islam, H. Mathkour, and S. Al-Ahmadi, "A multilayer system to boost the robustness of fingerprint authentication against presentation attacks by fusion with heart-signal," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 34, no. 8, pp. 5132–5143, Sep. 2022.
- [32] X. Cui, A. Aparcedo, Y. K. Jang, and S.-N. Lim, "On the robustness of large multimodal models against image adversarial attacks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2024, pp. 24625–24634.
- [33] Z. Yin, M. Ye, T. Zhang, T. Du, J. Zhu, H. Liu, J. Chen, T. Wang, and F. Ma, "VLATTACK: Multimodal adversarial attacks on vision-language tasks via pre-trained models," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 36, 2024, pp. 52936–52956.
- [34] S. Maity, M. Abdel-Mottaleb, and S. S. Asfour, "Multimodal low resolution face and frontal gait recognition from surveillance video," *Electronics*, vol. 10, no. 9, p. 1013, Apr. 2021.
- [35] M. Sivarathinabala, S. Abirami, M. Deivamani, and M. Sudharsan, "A smart security system using multimodal features from videos," *Pattern Recognit. Image Anal.*, vol. 29, no. 1, pp. 89–98, Jan. 2019.
- [36] S. A. F. Manssor, S. Sun, and M. A. M. Elhassan, "Real-time human recognition at night via integrated face and gait recognition technologies," *Sensors*, vol. 21, no. 13, p. 4323, Jun. 2021.
- [37] A. Sharma, N. Jindal, A. Thakur, P. S. Rana, B. Garg, and R. Mehta, "Multimodal biometric for person identification using deep learning approach," *Wireless Pers. Commun.*, vol. 125, no. 1, pp. 399–419, Jul. 2022.
- [38] 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. (CVPR)*, Jun. 2019, pp. 4685–4694.
- [39] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [40] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.
- [41] S. Maji and J. Malik, "Object detection using a max-margin Hough transform," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1038–1045.
- [42] J. Han and B. Bhanu, "Individual recognition using gait energy image," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 2, pp. 316–322, Feb. 2006.
- [43] J. N. Mogan, C. P. Lee, K. M. Lim, and K. S. Muthu, "VGG16-MLP: Gait recognition with fine-tuned VGG-16 and multilayer perceptron," *Appl. Sci.*, vol. 12, no. 15, p. 7639, Jul. 2022.
- [44] A. Mehmood, M. A. Khan, M. Sharif, S. A. Khan, M. Shaheen, T. Saba, N. Riaz, and I. Ashraf, "Prosperous human gait recognition: An end-to-end system based on pre-trained CNN features selection," *Multimedia Tools Appl.*, vol. 83, no. 5, pp. 14979–14999, Apr. 2020.
- [45] O. Elharrouss, N. Almaadeed, S. Al-Maadeed, and A. Bouridane, "Gait recognition for person re-identification," *J. Supercomput.*, vol. 77, no. 4, pp. 3653–3672, Apr. 2021.
- [46] D. Ramachandram and G. W. Taylor, "Deep multimodal learning: A survey on recent advances and trends," *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 96–108, Nov. 2017.
- [47] S. Nathan, M. P. Beham, A. Nagaraj, and S. Roomi, "Multiattention-net: A novel approach to face anti-spoofing with modified squeezed residual blocks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2024, pp. 1013–1020.
- [48] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 7354–7363.
- [49] Y. Dai, F. Giesecke, S. Oehmcke, Y. Wu, and K. Barnard, "Attentional feature fusion," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021, pp. 3559–3568.
- [50] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.
- [51] F. S. Samaria and A. C. Harter, "Parameterisation of a stochastic model for human face identification," in *Proc. IEEE Workshop Appl. Comput. Vis.*, Dec. 1994, pp. 138–142.
- [52] C. E. Thomaz and G. A. Giralaldi, "A new ranking method for principal components analysis and its application to face image analysis," *Image Vis. Comput.*, vol. 28, no. 6, pp. 902–913, Jun. 2010.
- [53] S. Zheng, J. Zhang, K. Huang, R. He, and T. Tan, "Robust view transformation model for gait recognition," in *Proc. 18th IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 2073–2076.



SHAIMA M. ALGHAMDI received the B.S. degree in computer science from Taif University, Taif, Saudi Arabia, in 2019. She is currently pursuing the M.S. degree in computer science with King Abdulaziz University, Jeddah, Saudi Arabia. She is also an AI & Digital Solutions Specialist. Her research interests include AI, gen-AI, and computer vision.



SALMA KAMMOUN JARRAYA received the Ph.D. degree in computer science from Sfax University, Tunisia. Currently, she is an Associate Professor with the CS Department, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia. She is a member of the Mir@cl Laboratory, University of Sfax. Her research interests include computer vision, video and image processing, machine learning, and knowledge discovery in images and video. She is a primary investigator of many research projects in artificial intelligence. She has served on technical conference committees and as a reviewer for many international conferences and journals, educational programs, and research projects.



FARIS KATEB received the Ph.D. degree in computer science from the University of Colorado at Colorado Springs, USA. Currently, he is an Assistant Professor and the Head of the IT Department, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia. His research interests include computer vision and image processing applications, such as object detection, face recognition, and adversarial examples. He is also working on the natural language process for Arabic and English. He participates as a speaker or a presenter in conferences and artificial intelligence areas and a member of the advisory board in other departments.

...