A Secure and Private Ensemble Matcher Using Multi-Vault Obfuscated Templates

Babak Poorebrahim Gilkalaye Shubhabrata Mukherjee Reza Derakhshani
School of Science and Engineering, University of Missouri-Kansas City
bpktk@mail.umkc.edu smpw5@umsystem.edu derakhshanir@umkc.edu

***Abstract*—Generative AI has revolutionized modern machine learning by providing unprecedented realism, diversity, and efficiency in data generation. This technology holds immense potential for biometrics, including for securing sensitive and personally identifiable information. Given the irrevocability of biometric samples and mounting privacy concerns, biometric template security and secure matching are among the most sought-after features of modern biometric systems. This paper proposes a novel obfuscation method using Generative AI to enhance biometric template security. Our approach utilizes synthetic facial images generated by a Generative Adversarial Network (GAN) as "random chaff points" within a secure vault system. Our method creates n sub-templates from the original template, each obfuscated with m GAN chaff points. During verification, s closest vectors to the biometric query are retrieved from each vault and combined to generate hash values, which are then compared with the stored hash value. Thus, our method safeguards user identities during the training and deployment phases by employing the GAN-generated synthetic images. Our protocol was tested using the AT&T, GT, and LFW face datasets, achieving ROC areas under the curve of 0.99, 0.99, and 0.90, respectively. Our results demonstrate that the proposed method can maintain high accuracy and reasonable computational complexity comparable to those unprotected template methods while significantly enhancing security and privacy, underscoring the potential of Generative AI in developing proactive defensive strategies for biometric systems.**

I. Introduction

Biometric user verification has become ubiquitous, but its vulnerabilities to various attacks and privacy breaches emphasize the need for enhanced security measures. Recent advances in Generative AI (GenAI) have transformed the field of machine learning, offering new opportunities for improving biometric security and privacy. GenAI's ability to generate highly realistic and diverse synthetic data has been leveraged in various applications, including biometric security [17], [25]. When it comes to biometric security and privacy, GenAI may be used to synthesize biometric data for template obfuscation thus safeguarding user privacy. Unlike traditional methods, GenAI-produced chaff points can be highly varied and realistic, making it exceedingly difficult for attackers to distinguish between genuine and synthetic data, significantly enhancing biometric template security [8], [46], [49].

Biometric systems have been vulnerable to various attack vectors, including sensor hijacks, injection of forged biometric templates, and network attacks on servers [34]. Such vulnerabilities may expose biometric systems to correlation attacks [53], presentation attacks, replay attacks, Man-in-the-Middle attacks, Denial-of-Service attacks, or Brute-Force attacks [10], [20], [28], [40]. Security incidents and privacy breaches through social media platforms, healthcare data, and government services [36], [38], [44] exemplify these vulnerabilities. The need for securing biomedical and biometric reference data is highlighted by the substantial increase in data breaches across US institutions and hospitals, including data theft, unauthorized access, and hacking, as reported by the HIPAA journal [22].

To protect stored biometric references from such attacks, researchers have proposed various security schemes, such as random projection, feature disentanglement, fuzzy vaults with auxiliary (helper) data, and deep learning-based concealable multi-biometrics to mitigate the earlier-mentioned privacy and security issues [1], [3], [43], [47]. However, many of these approaches are either computationally too expensive or not completely secure against certain attack vectors. In this work, we propose an encrypted vault approach to securely store biometric templates. Our approach uses mutually exclusive embeddings generated by various deep learning models, combined and then securely encrypted using SHA-512-based encryption. This technique incorporates 2000-4000 AI-generated face embeddings as chaff points, stored alongside the genuine template in a secure vault. This makes it computationally infeasible to distinguish the original template from synthetic data. This attribute makes our approach robust against brute-force attacks.

The reference feature vector, derived from the enrollment image and used for comparison with other incoming images, is called an enrollment template. For example, in the context of a Convolutional Neural Network (CNN) trained for face matching, this could be the flattened input to an FC layer. If the target image corresponds to the same person or object, for instance by the corresponding templates having a cosine similarity below a set threshold, the claimant is accepted. In this case, the claimant is genuine, and the acceptance is a true positive or genuine accept. If the claimant is an impostor (the incoming image is not from the same identity) and they manage to pass the template comparison, the result is a false or impostor accept. A genuine comparison is anticipated to yield a higher similarity, indicating that the feature vectors derived from the images closely match.

Fig. 1 describes a fuzzy extractor scheme and its terminology. We will use the same structure and terminology for our proposed obfuscation method. The Gen function generates a secret key and auxiliary data P from a biometric template t. The Rep function takes a biometric query q and the auxiliary data as its input. If dist(q, t) is smaller than a preset threshold, then it generates the correct key. In other words, the auxiliary data P helps to remove noise from query q in order to generate t. The auxiliary data P is required to be *secure* and reveal limited information about template t to preserve the privacy of template t. To satisfy the information-theoretic security, the entropy loss must be at most $\mathcal{L}$, as further detailed below [14], [13], [6].

The rest of this paper is organized as follows: Section II reviews related work. Section III describes our methodology. Section IV describes the security capability of our scheme. Section V outlines the experimental setup. Section VI presents our results and their implications. Finally, we conclude with a summary of our method's unique attributes, challenges faced, and future work.

II. Related Work

Many methods have been introduced over the years to preserve the privacy of biometric data. For instance, [32] used a de-identification technique, but anonymity was not guaranteed. Other methods explored encryption-based schemes [30], but these are vulnerable to various attacks, including those that exploit high False Acceptance Rates (FAR), where an attacker attempts to gain unauthorized access by presenting multiple fraudulent biometric samples until one is falsely accepted as genuine. Our approach, demonstrably secure against such attacks (Section IV), offers significant improvements. Techniques utilizing Convolutional Neural Networks (CNNs) have been used to protect face templates while maintaining good matching performance [21]. Blockchain-based storage of biometric templates has also been discussed [11], but it often incurs high execution times. Adversarial learning for concealing information within a learned space has also been explored [29], focusing on disentangling identity from attribute information to mitigate soft-biometric information in face templates [3], [42]. Protecting source data by combining biometric modalities (e.g., fingerprint and iris scans) using various fusion techniques has been studied [35], [39]. While enhancing security through redundancy, it introduces a single point of failure, making the entire system vulnerable if one trait is compromised. It is also possible to show the unlinkability of the protected templates using different existing measures such as maximal leakage [37]. In recent years, Generative AI (GenAI) has gained traction in biometric privacy. GenAI's ability to generate highly realistic synthetic biometric data shows promise for enhancing privacy and security. For instance, [16] and [26] demonstrated the effectiveness of Generative Adversarial Networks (GANs) in creating synthetic facial images that can obfuscate original data, and thus protect user identities. Similarly, [48] explored using GenAI to generate diverse biometric data, improving the robustness of biometric systems against various attacks. Recent studies have investigated GenAI's role in creating synthetic multimodal biometric data, for instance combining facial images with other modalities such as voice and fingerprints [7], [45]. This approach not only enhances security but also provides additional layers of privacy protection by making it more difficult to link synthetic data to real individuals. Furthermore, [51] highlights the potential of GenAI in generating synthetic data for training biometric systems, reducing reliance on real biometric data, and thus mitigating some privacy concerns.

One can leverage existing measures, such as maximal leakage [37], to demonstrate the unlikability of protected templates. Our approach leverages such GenAI advances to provide a robust and secure method for protecting biometric templates, addressing a number of the limitations of previous methods while being computationally efficient.

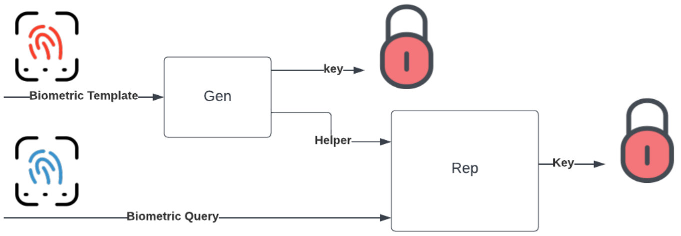


Fig. 1. Fuzzy extractor scheme

III. Methodology

A. Main Idea

In this section, we will first informally explain the main idea behind the proposed method, and then provide the formalized algorithm. The idea is to divide the biometric template into m sub-templates and hide them with n chaff points to make it computationally impractical for an adversary to find the template. One solution would be to hide a template t with n chaff points in a public vault, $V = [r_1, ..., t, ..., r_n]$, and store $f(t)$, where f is a one-way function such as SHA-256. Now the question is how hard it would be for an adversary to find the index of t in V. To make it computationally impractical for the adversary to find t, n must be a very big number (λ (security parameter)), as the best strategy to find the template t is to compute $f(x)$ for all x in the vault to find the template t. However, this method is also computationally impractical for the legitimate user. By way of example, assume we will have $f(t)$ and one vault containing $1 + 2^{80}$ vectors as auxiliary data. We assume the inserted chaff points are indistinguishable from the template t in the collection. By indistinguishable chaff points, we mean that the chaff points represent realistic face templates that an attacker cannot differentiate from the actual template. We achieve this using StyleGAN2 to generate synthetic face embeddings that closely mimic the statistical properties of genuine templates, making it extremely challenging to identify the original among them. For matching a query q, we compute the distance of q from all points inside the vault and retrieve m (depending on the threshold of the system) closest points to q: $p_1, p_2, ..., p_m$, and compute $f(p_1), f(p_2), ..., f(p_m)$ and check if there are any matches between these values and $f(t)$. This system is secure under certain conditions, but impractical because during the authentication the user needs to compute the desired distance of the query q vs. $1+2^{80}$ vectors.

Thus we propose generating n sub-templates $t_1, t_2, ..., t_n$ from template t, using protocol $\mathcal{G}$, such that $\mathcal{G}(t) : t_1, t_2, ..., t_n$ and hide each of them with m chaff points. In the end, we store $f([t_1, t_2, ..., t_n])$ (, where f is a one-way function), in n vaults such that each vault has $m+1$ vectors. To make it computationally impractical for the adversary to find the template by brute force, we need m^n to be very large. Unlike the earlier impractical example, we can choose m and n to garner computational security against brute force attacks while having a practical solution. In the forthcoming experimental demonstration, we use $m = 2000$, and $n = 5$, so that $m^n = 2^{54.82}$ to get acceptable security against a brute force attack. When matching query q, we generate $q_1, q_2, ...q_n$ using protocol $\mathcal{G}(q) : q_1, q_2, ..., q_n$ (protocol $\mathcal{G}$ will be explained later). The distances between q_i and all $m+1$ vectors in the corresponding vault are then computed and the top s vectors from each vault are retrieved. The threshold of the system is s and the distance can be computed by any distance metric, such as cosine similarity. At this point, we have retrieved sn vectors. Next, we construct all s^n possible combinations of vectors $v_1, v_2, ...v_{s^n}$ and compute $f(v_1), f(v_2), ...f(v_{s^n})$ and check if there are any matches between the stored value of $f([t_1, t_2, ..., t_n])$ and the $f(v_i)$. If q is a genuine vector, it should be able to retrieve $t_1, t_2, ..., t_n$ with the threshold s. If q is an imposter, it will fail to retrieve $t_1, t_2, ...t_n$ from the vaults. It should be noted that s cannot be a large number as it would render the biometric matching computationally impractical.

B. True Positive Rate (TPR) Improvement

It can be seen that a query needs to satisfy all n vaults in order to succeed. Thus we relax this condition and build the system such that a query can succeed if it satisfies k out of n vaults. To do so, after the generation of sub-templates $t_1, t_2, ...t_n$, we need to store all $\binom{n}{k}$ possible combinations of $f(t_i)$. To satisfy the asymptotic security, we need to make sure m^k is computationally impractical for a computationally bounded adversary, where m is the number of chaff points per vault.

C. The Algorithm

Our algorithm *Alg* consists of a GEN protocol, a REP protocol, a key generation protocol $\mathcal{K}$, a sub-template Generator $\mathcal{G}$, a one-way function f, and a security parameter γ. We assume that the template t is sampled from the distribution $\mathcal{D}$.

Key Generation $\mathcal{K}(\gamma)$: This protocol chooses n (number of vaults) and m (number of chaff points) so that $m^n > 2^\gamma$. Then it generates mn random samples from distribution $\mathcal{D}$ for mn chaff points $K = c_{11}, c_{12}..., c_{1m}, c_{21}, c_{22}, c_{21}, ...c_{2m}, ..., c_{nm}$. Two types of random numbers R_{1n} and R_{2n} are generated. Where R_{1n} are n scalars and R_{2n} are small vectors with the same dimension as the template. These will be used to hide the template in each vault further, as follows. Let d be the cosine similarity of two vectors, then:

$d(A, B) = d(A, rB)$ for any arbitrary scalar r, and:

$d(A, B) \approx d(A, r_1B + R_1)$ for any arbitrary scalar r_1 and small vector R_1

Sub-template Generator $\mathcal{G}(t, n)$: This protocol takes template t and the number of vaults n as its inputs and generates n independent sub-templates $T = (t_1, t_2, ..., t_n)$. In the experimental section, we propose one solution for such a function.

GEN: GEN(T, K) This protocol places each sub-template t_i and its corresponding chaff points $c_{i,j}$, for all $j = 1, .., m$ inside a vault in no specific order. It then stores $P_1 = f(t_1, t_2, ...t_n)$ and outputs the helper P:

REP: REP(C, q, tr): With q representing a new biometric query using protocol $\mathcal{G}$, $Q = (q_1, q_2, , ...q_n)$ is generated and the closest tr vectors are retrieved from each vault. The threshold of the system is denoted by tr. Note that tr is different from the original well-known face matching threshold. tr may be adjusted to increase or decrease the False Acceptance Rate (FAR). In other words, if $tr = 1$ we only retrieve the closest vector in each vault. This will drastically reduce the False Positive Rate (FPR), but simultaneously it reduces the TPR.

The following are retrieved from the n vaults:
$\{w_{11}, w_{12}, ..., w_{1tr}\}$, $\{w_{21}, w_{22}, ..., w_{2tr}\}$...
$\{w_{n1}, w_{n2}, ..., w_{ntr}\}$

All $\binom{n}{tr}$ possible combinations of these vectors are created and passed to the one-way function f. If there is any match between the generated hash values and the stored hash value P_1, then the query q is authenticated and the protocol outputs 1 (success), otherwise, it outputs 0 (fail).

IV. SECURITY

An ideal biometric security scheme needs to satisfy ***Revocability***, ***Diversity***, ***Accuracy***, and ***Security*** [27] requirements. Our proposed scheme meets all these requirements, as detailed below.

Revocability: A biometric system needs to be revocable because if it is compromised, or if the user wishes to no longer be identified, they should be able to revoke their reference. Our scheme's templates are revocable by choosing a new set of chaff points and a new sub-template generator. This process is further facilitated by the ease of GAN-based chaff generation.

Diversity: A secure template must not allow cross-matching across databases, thereby ensuring the user's privacy. In our system, for a false positive to occur, the imposter must meet the authentication criteria of each vault and successfully retrieve all segments of the template, hence our system mitigates the cross-matching as compared to systems using one model. It also could be verified throughout the experimental section, the FPR is very small, because an imposter needs to satisfy multiple vaults to be able to positively match.

Accuracy: Our method shows minimum accuracy degradation with a 0.99 area under the curve (AUC) for both the AT&T [4] and Georgia Tech (GT) face databases [31].

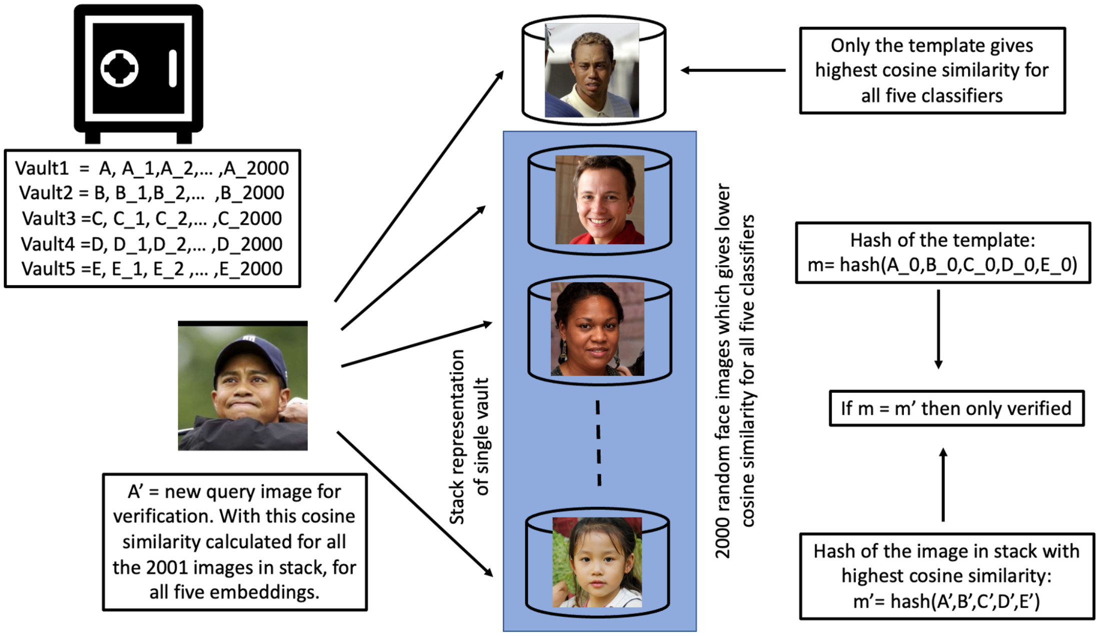

Vault1 = A, A_1,A_2,… ,A_2000
Vault2 = B, B_1,B_2,… ,B_2000
Vault3 =C, C_1, C_2,… ,C_2000
Vault4 =D, D_1,D_2,… ,D_2000
Vault5 =E, E_1, E_2 ,… ,E_2000
A' = new query image for verification. With this cosine similarity calculated for all the 2001 images in stack, for all five embeddings.
Stack representation of single vault
2000 random face images which gives lower cosine similarity for all five classifiers
Only the template gives highest cosine similarity for all five classifiers
Hash of the template: m= hash(A_0,B_0,C_0,D_0,E_0)
If m = m' then only verified
Hash of the image in stack with highest cosine similarity: m'= hash(A',B',C',D',E')

Fig. 2. Embedding generated by the secure hashing verification

Security: To show that our system is secure, we need to show that the auxiliary data $P = (P_1, P_2)$ is not revealing too much information and there is enough entropy in this system. Chaff points are generated from the same distribution that the corresponding sub-template is generated. This is achieved using StyleGAN2, which was originally trained on the Flickr-Faces-HQ (FFHQ) [24]dataset containing 70,000 high-quality, diverse human face images. This training ensures that the generated chaff points closely mimic the statistical properties and distribution of real face embeddings. Hence we can say the sub-templates inside each vault are indistinguishable from other chaff points there. Getting this indistinguishability between the sub-template and chaff points inside a vault is straightforward; the challenge is to achieve indistinguishability between sub-templates and chaff points. In other words, if an adversary knows that t_1 is a sub-template, other t_i should remain indistinguishable from chaff points, we achieve this by using independent models, given two models M_1, M_2, and template t and a random face r, let $===$ be coming from the same image.

$$|Pr[M_1(t) {=}{=}{=} M_2(t)] - Pr[M_1(t) {=}{=}{=} M_2(r)]| = \epsilon$$

To achieve this, we must make sure that the sub-templates are independent. If we assume everything looks completely random inside vaults, then we can claim the best strategy to find the template for any adversary is to check all possible combinations of elements in vaults, which we have made impractical by choosing proper n,m. As a result, our approach provides superior security against brute force attacks, as well as faster face acceptance and rejection capabilities.

It is clear that given a face template t, a random face r, and two biometric models $M1$ and $M2$, $y_1 = M1(t)$ and $y_2 = M2(t)$ cannot be completely independent. However, in our case, we are using big neural networks. With this assumption and if $M1$ and $M2$ have completely different structures, it is not going to be easy to distinguish between $M1(t)$ and $M1(r)$ given $M2(t)$. In other words, if we tell the adversary $M2(t)$ belongs to person A, and then if we ask him to tell us which one of $M1(t)$ and $M1(r)$ belongs to the same person, it will not be easy for the adversary to solve this problem.

One approach is to reverse engineer the template to recreate it and see which one of reverse $M1(t)$ and reverse $M1(r)$ is closer to the reverse $M2(t)$. We have addressed this attack by adding randomness; now the adversary needs to distinguish between $R_2M1(t)+R_2'$ and $R_2M1(r)+R_2''$, given $R_1M2(t)+R_1'$.

V. Experiments

A. Experimental Setup

Per common practice in face recognition, we use cosine similarity for matching deep face templates [33]. Cosine similarity between any two vectors is given by:

$$\cos(\mathbf{A},\mathbf{B}) = \frac{\mathbf{A}\mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} \mathbf{A}_i\mathbf{B}_i}{\sqrt{\sum_{i=1}^{n} (\mathbf{A}_i)^2}\sqrt{\sum_{i=1}^{n} (\mathbf{B}_i)^2}} \quad (1)$$

Where A and B are the two feature vectors being compared.

In our research, a face image is transformed into an embedding using five different deep learning methods (Fig. 3). Each of these embeddings is produced by a different convolutional neural network (CNN) model. The 512-dimensional feature vectors $(\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3, \boldsymbol{e}_4, \boldsymbol{e}_5)$ are numerically distinct, as verified by cosine similarity comparisons showing low similarity scores between vectors of different vaults. A total of five secure vaults is constructed. Each vault contains 2001 embeddings, of which 2000 embeddings are generated using 2000 random images of human faces and only 1 embedding is generated from an intended face image. We employed StyleGAN2 [24] to generate realistic human face images for our chaff points. StyleGAN2 was chosen for its superior image quality, diversity, and photo-realism compared to alternatives like PGGAN [23] and StarGAN [9]. Its improved fine-scale detail, reduced artifacts, and disentangled latent space make it ideal for creating indistinguishable and diverse chaff points for our secure vault system. These images were then used to create random chaff points for the secure vault.[1] This approach improves the privacy safeguards for the proposed method during the development and evaluation of the Multi-Vault Obfuscated Templates algorithm. The same structure, consisting of 2000 embeddings generated using random images and 1 embedding from an intended face image (i.e. the template), is replicated in all five vaults. To ensure security, the hash value of the five different embeddings $H = hash(e1, e2, e3, e4, e5)$ is stored.

B. Verification Process

The process for a single vault is shown in Fig. 2. For a biometric verification match:

- Five distinct embeddings (A', B', C', D', E') are generated using five separate classifiers based on the image of the person seeking verification (referred to as the query).
- Each resultant embedding is compared against all embeddings across the corresponding vault, resulting in the calculation of a total of 2001*5 cosine similarity scores.
- The embedding with the highest cosine similarity score is selected from each vault, leading to the choice of five winners, one from each vault.
- A secure 64-bit hash, denoted as m', is constructed by combining the five embeddings selected in the previous step.
- The query is verified only if this hash matches the previously generated hash ($H = hash(e1, e2, e3, e4, e5)$) of the template. However, in our implementation, we selected the top two or three embeddings from each vault and generated all possible 5^2 or 5^3 hashes, and compared them with the stored hash. This allows for relaxing the FRR of the system.

A higher number of random images can be used to increase security. A lower and higher number of classifiers can be used depending upon the scalability requirements.

C. Pre-trained Models and Datasets

A set of five different pre-trained models have been used here to construct the end-to-end secure privacy framework. The first model is InceptionResnetV1; this particular model [41] has been pre-trained on the VGGFace2dataset. [5], [15]. As a second classifier, another variation of InceptionResnetV1 has been used, but this version was pre-trained with CASIA-Webface [50]. For both models, one and two, MTCNN [52] has been used as a preprocessing for face cropping. More precisely, MTCNN cropped a 250x250 dimension face to improve the model accuracy. As a third and fourth classifier, two different versions of insightFace [12] have been used. The third pretrained classifier, named "buffalo", is an implementation of the ResNet50 model pretrained on the WebFace600K dataset. This dataset contains 600,000 unique identities and is derived from the larger WebFace42M dataset [54] which contains 42 million identities. The fourth pre-trained model antelopev2 is an implementation of ResNet100 pretrained using Glint360K [2], which contains 17091657 images of 360232 individuals. The fifth model is an arcface-based implementation of ResNet100 [18]; for this version, some additional preprocessing has been used including image resizing, detection of faces using a face detector, and face alignment.

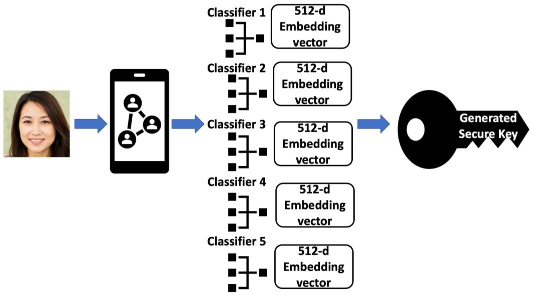


Fig. 3. Generating key using five different classifiers

Our method was evaluated using three widely recognized face datasets: the AT&T Database of Faces (AT&T), Labeled Faces in the Wild (LFW), and Georgia Tech Face Database (GT). AT&T contains 400 images of 40 subjects with controlled variations, LFW includes over 13,000 unconstrained face images collected from the web, and GT comprises 750 color images of 50 subjects with diverse facial expressions and illumination conditions. These datasets were selected to provide a comprehensive evaluation across different controlled and uncontrolled conditions, varying numbers of subjects, and diverse image characteristics, allowing us to assess our method's performance in a range of scenarios.

VI. Result and Analysis

Our results are mainly described from two perspectives, accuracy and time complexity performance. Fig. 4 shows the comparison between the distribution of cosine similarity scores for genuine and imposter face images. The leftmost distribution is for the genuine cosine similarities, which mostly lie between 0.8 to 1. The middle distribution is for the imposter cosine similarity computed from the target template and other (non-mated) class images from the *same* dataset (e.g., same AT&T dataset but different identities), with the score distribution mostly covering 0.2 to 0.4. The rightmost distribution also represents imposter cosine similarities but is computed by comparing the target templates and the random

[1] To achieve this, we conducted an iterative process, generating 2000 embeddings from 2000 realistic face images obtained from the website thispersondoesnotexist.com

face images generated by GAN (as mentioned in the experimental setup), with the scores being mostly between -0.25 and 0.25. The constrained feature space of faces within a single dataset promotes higher cosine similarity (0.2-0.4) due to inherent structural similarities. Conversely, the inclusion of random external faces introduces greater vector variance, leading to a broader cosine similarity distribution (-0.25 to 0.25). These comparisons show how well the cosine similarity can distinguish the embeddings generated from face images. The corresponding ROC AUC of cosine similarity scores reinforces this fact.

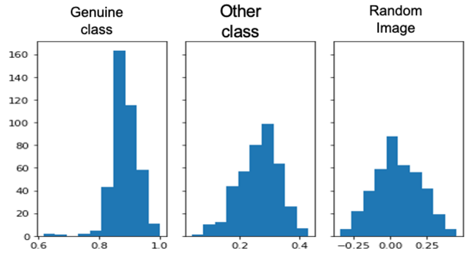


Fig. 4. Distribution of Cosine Similarity for Genuine, Other, and Random Facial Images

A. Individual Classifier Performance

TABLE I
INDIVIDUAL CLASSIFIER AUC FOR DIFFERENT DATASETS

Dataset used	***Classifier used***	***AUC***
AT&T	InceptionResNetV1-VGG	.9999
AT&T	InceptionResNetV1-Casia	.9996
AT&T	InsightFace Buffalo	.9952
AT&T	InsightFace Antelope v2	.9978
AT&T	ResNet100 with preprocessing	1.000
GT	InceptionResNetV1-VGG	.9998
GT	InceptionReNnetV1-Casia	.9995
GT	InsightFace Buffalo	.9780
GT	InsightFace Antelope v2	.9745
GT	Resnet100 with preprocessing	1.000
LFW	InceptionResNetV1-VGG	.9904
LFW	InceptionResNetV1-Casia	.9909
LFW	InsightFace Buffalo	.9070
LFW	InsightFace Antelope v2	.9175
LFW	ResNet100 with preprocessing	.9701

All five classifiers chosen for the secure hash pipeline performed exceptionally well on the AT&T dataset, achieving an AUC exceeding 0.99 (shown in Table I). Similar success was observed on the GT dataset, where at least three classifiers surpassed an AUC of 0.99, and the remaining two achieved an AUC exceeding 0.97. The LFW dataset [19] yielded comparable results, with at least three classifiers again exceeding an AUC of 0.99. The designation 'genuine' indicates a positive class, and 'imposter' is marked as a negative class.

TABLE II
END-TO-END BENCHMARKING PERFORMANCE

Dataset	***Chaff points***	***TPR***	***TNR***	***Classifier used***
GT	4000	81.20%	100.00%	5
GT	2000	84.63%	100.00%	5
GT	4000	91.46%	100.00%	4
AT&T	4000	91.58%	100.00%	5
AT&T	2000	93.27%	100.00%	5
AT&T	4000	96.19%	100.00%	4

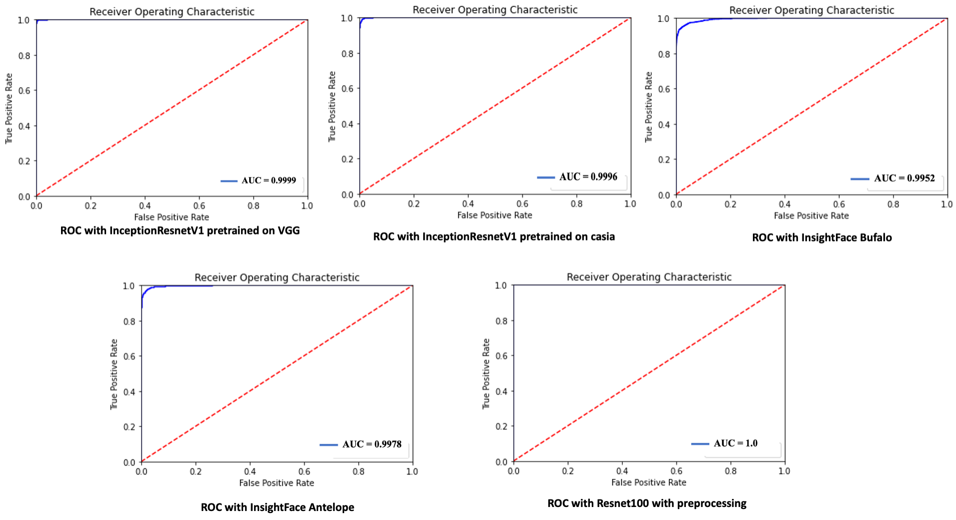
B. End-to-end Benchmarking

As shown in Table II, the True Negative Rate (TNR) remained mostly at 100% for the proposed framework when tested with both AT&T and GT datasets. However, when the number of chaff points was increased from 2000 to 4000, the true positive rate slightly decreased. It was also observed that if the system had to satisfy only 4 out of all 5 classifiers, the true positive rate could improve even with more chaff points. Using a higher number of chaff points can increase the computation complexity, which in turn may make the system robust against attack. Introducing additional classifiers strengthens the system by exponentially increasing the imposter's comparison workload ($(n+1)^C > n^C$), thus making the system further resilient against brute force attacks. Here, C is the number of classifiers and n is the number of chaff points. A combination of 4000 chaff points and 4 classifiers yielded the best TPR and TNR.

C. ROC Analysis

Figures 6, 7, and 8 depict the study ROC curves for various datasets, all generated using the presented secure hash pipeline. Instead of evaluating individual classifiers, these ROC curves assess the performance of the complete biometric enrollment and authentication pipeline, which we refer to as the registration process. Within this pipeline, a sample is considered a true positive if a genuine person is successfully registered, and a false positive otherwise. The ROC curves (Figures 6, 7, 8) depict the trade-off between security and accuracy in the registration process. These curves show how the rate of incorrectly accepting imposters (FPR) increases as we loosen the criteria for accepting genuine users (TPR). Imagine a vault containing 2001 images. The strictest system would only accept a single image with the highest cosine similarity score, while the most lenient would accept all 2001. Practically, accepting the top five highest-scoring images might be a good balance. This means the person can be authenticated if their true identity is among the top five closest matches out of the 2001 comparisons. As the ROC curves show, there's a point beyond which accepting more images doesn't significantly improve the chance of correctly identifying a genuine person, while increasing the risk of mistakenly accepting someone else.

As shown in Table III, the AT&T and Georgia Tech Face datasets achieve an AUC greater than 0.99, while the LFW dataset, with its challenging variability in pose, expression, and lighting, yields a lower AUC of 0.9042.



ROC with InceptionResnetV1 pretrained on VGG

ROC with InceptionResnetV1 pretrained on casia

ROC with InsightFace Bufalo

ROC with InsightFace Antelope

ROC with Resnet100 with preprocessing

Fig. 5. ROC Performance of Individual Classifiers Based on the AT&T Face Dataset

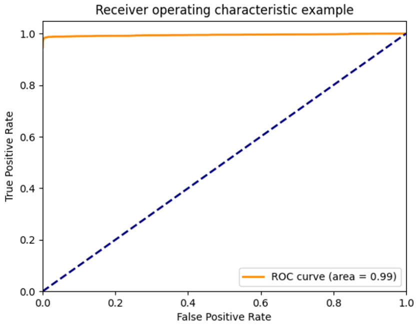


Fig. 6. End-to-end Performance Based on the AT&T Face Dataset

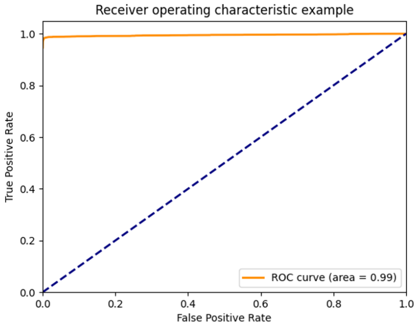


Fig. 7. End-to-end Performance Based on the GT Face Dataset

TABLE III
END-TO-END AUC PERFORMANCE ON DIFFERENT DATASETS

Dataset for benchmarking	***AUC***
AT&T face dataset	0.9939
Georgia Tech Face dataset	0.9942
LFW dataset	0.9042

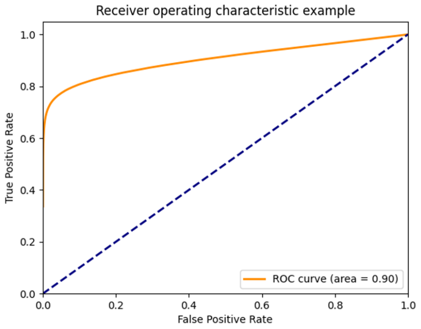


Fig. 8. End-to-end Performance Based on the LFW Face Dataset

D. *Time Complexity Analysis*

TABLE IV
TIME COMPLEXITY PERFORMANCE

Event	***Time taken in seconds***
Face detection and embedding	1.15
Matching (cosine similarity)	0.31
Secure hash generation	0.01
Complete end-to-end procedure	1.47

Our system, developed on a Windows laptop with an Intel i9 CPU and NVIDIA GeForce RTX 3080 GPU, demonstrates efficient performance. Face detection, preprocessing, and embedding creation take 1.15 seconds, matching 0.31 seconds, and secure hash generation 0.01 seconds(Table IV). The process completes in an average of 1.47 seconds. Comparable non-protected systems report 0.1 to 0.5 seconds for feature extraction and matching. Our system maintains competitive performance while introducing additional security performance.

VII. CONCLUSION

In this paper, we proposed a cryptographic biometric template security protocol leveraging GenAI to enhance the security and privacy of face recognition. By harnessing GenAI's ability to generate highly realistic and diverse synthetic data, we demonstrated a 2^{γ} security against brute-force attacks. Our experimental results on the AT&T, GT, and LFW datasets showed that our protocol preserves the average accuracy of the models used to generate the embeddings. We introduced a novel ROC curve computation method that evaluates our entire biometric pipeline, including multiple classifiers and GAN-generated chaff points. This comprehensive approach captures the synergistic effects of our multi-layered security system, unlike traditional ROC analyses that typically assess individual components separately. By utilizing GAN-generated synthetic data, we eliminated the risk of exposing real individuals' identities during training and verification, hence providing additional privacy protections. This work showcases GenAI's potential in constructing robust and secure end-to-end biometric systems. Future work may include exploring other GenAI methods to create more diverse samples and integrate them with other modalities, such as fingerprints, iris, and voice to develop multimodal biometric systems that are more resilient to attacks.

As GenAI continues its advance, one can envision a future where biometric security protocols continuously evolve to stay ahead of potential threats, enabling a new wave of secure and private biometric systems.

Acknowledgment

This work was supported in part by a grant from CBL, an NSF IUCRC. Dr. Derakhshani is also a consultant for Jumio.

References

[1] E. Abdellatef, N. A. Ismail, S. E. S. Abd Elrahman, K. N. Ismail, M. Rihan, A. El-Samie, and E. Fathi. Cancelable multi-biometric recognition system based on deep learning. *The Visual Computer*, 36(6):1097–1109, 2020.

[2] X. An, X. Zhu, Y. Gao, Y. Xiao, Y. Zhao, Z. Feng, L. Wu, B. Qin, M. Zhang, D. Zhang, et al. Partial fc: Training 10 million identities on a single machine. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1445–1449, 2021.

[3] B. Bortolato, M. Ivanovska, P. Rot, J. Križaj, P. Terhörst, N. Damer, P. Peer, and V. Štruc. Learning privacy-enhancing face representations through feature disentanglement. In *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*, pages 495–502, 2020.

[4] A. L. Cambridge. The database of faces. cam-orl.co.uk/facedatabase.html.

[5] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman. Vggface2: A dataset for recognising faces across pose and age. In *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*, pages 67–74. IEEE, 2018.

[6] E.-C. Chang and Q. Li. Hiding secret points amidst chaff. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 59–72. Springer, 2006.

[7] A. Chen et al. Enhancing biometric security with gans. *Information Fusion*, 77:23–34, 2022.

[8] S. Chen, R. Patel, and L. Nguyen. Advancements in generative ai for biometric applications. *Pattern Recognition Letters*, 159:112–120, 2022.

[9] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8789–8797, 2018.

[10] A. K. Das, M. Wazid, N. Kumar, A. V. Vasilakos, and J. J. P. C. Rodrigues. Biometrics-based privacy-preserving user authentication scheme for cloud-based industrial internet of things deployment. *IEEE Internet of Things Journal*, 5(6):4900–4913, 2018.

[11] O. Delgado-Mohatar, J. Fierrez, R. Tolosana, and R. Vera-Rodriguez. Biometric template storage with blockchain: A first look into cost and performance tradeoffs. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 2829–2837, 2019.

[12] J. Deng, J. Guo, N. Xue, and S. Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4690–4699, 2019.

[13] Y. Dodis, R. Ostrovsky, L. Reyzin, and A. Smith. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. *SIAM journal on computing*, 38(1):97–139, 2008.

[14] Y. Dodis, L. Reyzin, and A. Smith. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. In *International conference on the theory and applications of cryptographic techniques*, pages 523–540. Springer, 2004.

[15] T. Esler. facenet-pytorch. https://github.com/timesler/facenet-pytorch, 2020.

[16] A. Fang et al. A survey on gans for biometric security. *IEEE Transactions on Information Forensics and Security*, 17:2451–2467, 2022.

[17] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[18] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[19] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, October 2007.

[20] A. Husseis, J. Liu-Jimenez, I. Goicoechea-Telleria, and R. Sanchez-Reillo. A survey in presentation attack and presentation attack detection. In *2019 International Carnahan Conference on Security Technology (ICCST)*, pages 1–13. IEEE, 2019.

[21] A. K. Jindal, S. Chalamala, and S. K. Jami. Face template protection using deep convolutional neural network. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 575–5758, 2018.

[22] T. H. Journal. Healthcare data breach statistics, 2024.

[23] T. Karras, T. Aila, S. Laine, and J. Lehtinen. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2017.

[24] T. Karras, S. Laine, and T. Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410, 2019.

[25] T. Karras, S. Laine, and T. Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8110–8119, 2020.

[26] A. Liu et al. Improving biometric template security using gans. *Pattern Recognition Letters*, 157:20–30, 2023.

[27] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar. *Handbook of fingerprint recognition*. Springer Science & Business Media, 2009.

[28] M. Martinez-Diaz, J. Fierrez-Aguilar, F. Alonso-Fernandez, J. Ortega-Garcia, and J. Siguenza. Hill-climbing and brute-force attacks on biometric systems: A case study in match-on-card fingerprint verification. In *Proceedings 40th Annual 2006 International Carnahan Conference on Security Technology*, pages 151–159, 2006.

[29] A. Morales, J. Fierrez, R. Vera-Rodriguez, and R. Tolosana. Sensitivenets: Learning agnostic representations with application to face images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(6):2158–2164, 2021.

[30] I. Natgunanathan, A. Mehmood, Y. Xiang, G. Beliakov, and J. Yearwood. Protection of privacy in biometric data. *IEEE Access*, 4:880–892, 2016.

[31] A. V. Nefian. Georgia tech face database, 1999.

[32] E. Newton, L. Sweeney, and B. Malin. Preserving privacy by de-identifying face images. *IEEE Transactions on Knowledge and Data Engineering*, 17(2):232–243, 2005.

[33] H. V. Nguyen and L. Bai. Cosine similarity metric learning for face verification. In *Asian conference on computer vision*, pages 709–720. Springer, 2010.

[34] Z. Rui and Z. Yan. A survey on biometric authentication: Toward secure and privacy-preserving identification. *IEEE Access*, 7:5994–6009, 2019.

[35] K. Sasireka and R. Rajesh. Dual biometric authentication scheme for privacy protection. In *2014 International Conference on Communication*

and Network Technologies, pages 105–108, 2014.

[36] O. F. Segun and F. B. Olawale. Healthcare data breaches: Biometric technology to the rescue. *Int. Res. J. Eng. Technol.*, 4(11):946–950, 2017.

[37] H. O. Shahreza, Y. Y. Shkel, and S. Marcel. Measuring linkability of protected biometric templates using maximal leakage. *IEEE Transactions on Information Forensics and Security*, 2023.

[38] P. Singh. Aadhaar and data privacy: biometric identification and anxieties of recognition in india. *Information, Communication & Society*, 24(7):978–993, 2021.

[39] S. P. Singh and S. Tiwari. A dual multimodal biometric authentication system based on woa-ann and ssa-dbn techniques. *Sci*, 5(1):10, 2023.

[40] D. F. Smith, A. Wiliem, and B. C. Lovell. Face recognition on consumer devices: Reflections on replay attacks. *IEEE Transactions on Information Forensics and Security*, 10(4):736–745, 2015.

[41] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Thirty-first AAAI conference on artificial intelligence*, 2017.

[42] P. Terhörst, D. Fährmann, N. Damer, F. Kirchbuchner, and A. Kuijper. On soft-biometric information stored in biometric face embeddings. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 3(4):519–534, 2021.

[43] U. Uludag and A. Jain. Securing fingerprint template: Fuzzy vault with helper data. In *2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)*, pages 163–163, 2006.

[44] I. ur Rehman. Facebook-cambridge analytica data harvesting: What you need to know. *Library Philosophy and Practice*, pages 1–11, 2019.

[45] A. Wang et al. Gan-based synthetic data for biometric security. *IEEE Access*, 11:123456–123467, 2023.

[46] M. Wang, E. Johnson, and D. Brown. Exploring the potentials of generative ai for future biometrics. In *Proceedings of the International Joint Conference on Biometrics (IJCB)*, pages 123–130. IEEE, 2023.

[47] Y. Wang and K. N. Plataniotis. An analysis of random projection for changeable and privacy-preserving biometric verification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 40(5):1280–1293, 2010.

[48] A. Yang et al. Template obfuscation in biometrics using gans. *Journal of Biometrics*, 34:89–98, 2022.

[49] J. Yang, J. Smith, and A. Lee. Generative ai in biometrics: A comprehensive survey. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 4(1):45–67, 2022.

[50] D. Yi, Z. Lei, S. Liao, and S. Z. Li. Learning face representation from scratch. *arXiv preprint arXiv:1411.7923*, 2014.

[51] A. Zhang et al. Recent advancements in generative ai. *Journal of Machine Learning Research*, 23(1):1–45, 2022.

[52] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE signal processing letters*, 23(10):1499–1503, 2016.

[53] X. Zhou, S. D. Wolthusen, C. Busch, and A. Kuijper. Feature correlation attack on biometric privacy protection schemes. In *2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pages 1061–1065, 2009.

[54] Z. Zhu, G. Huang, J. Deng, Y. Ye, J. Huang, X. Chen, J. Zhu, T. Yang, J. Lu, D. Du, et al. Webface260m: A benchmark unveiling the power of million-scale deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10492–10502, 2021.