Computers, Materials & Continua

Tech Science Press

DOI: 10.32604/cmc.2023.038556

ARTICLE

Deep Learning-Based Robust Morphed Face Authentication Framework for Online Systems

Harsh Mankodiya[1], Priyal Palkhiwala[1], Rajesh Gupta[1,*], Nilesh Kumar Jadav[1], Sudeep Tanwar[1], Osama Alfarraj[2], Amr Tolba[2], Maria Simona Raboaca[3,4,*] and Verdes Marina[5]

[1]Department of Computer Science and Engineering, Institute of Technology, Nirma University, Ahmedabad, 382481, India

[2]Computer Science Department, Community College, King Saud University, Riyadh, 11437, Saudi Arabia

[3]Doctoral School, University Politehnica of Bucharest, Bucharest, 060042, Romania

[4]Department of Hydrogen and Fuel Cell, National Research and Development Institute for Cryogenics and Isotopic Technologies, Ramnicu Valcea, 240050, Romania

[5]Department of Building Services, Faculty of Civil Engineering and Building Services, Technical University of Gheorghe Asachi, Iasi, 700050, Romania

*Corresponding Authors: Rajesh Gupta. Email: rajesh.gupta@nirmauni.ac.in; Maria Simona Raboaca. Email: simona.raboaca@icsi.ro

Received: 18 December 2022 Accepted: 20 June 2023 Published: 31 October 2023

ABSTRACT

The amalgamation of artificial intelligence (AI) with various areas has been in the picture for the past few years. AI has enhanced the functioning of several services, such as accomplishing better budgets, automating multiple tasks, and data-driven decision-making. Conducting hassle-free polling has been one of them. However, at the onset of the coronavirus in 2020, almost all worldly affairs occurred online, and many sectors switched to digital mode. This allows attackers to find security loopholes in digital systems and exploit them for their lucrative business. This paper proposes a three-layered deep learning (DL)-based authentication framework to develop a secure online polling system. It provides a novel way to overcome security breaches during the face identity (ID) recognition and verification process for online polling systems. This verification is done by training a pixel-2-pixel *Pix2pix* generative adversarial network (GAN) for face image reconstruction to remove facial objects present (if any). Furthermore, image-to-image matching is done by implementing the Siamese network and comparing the result of various metrics executed on feature embeddings to obtain the outcome, thus checking the electorate credentials.

KEYWORDS

Artificial intelligence; discriminator; generator; *Pix2pix* GANs; Kullback-Leibler (KL)-divergence; online voting system; Siamese network

1 Introduction

Elections in a democratic country play a crucial role in electing a leader for the nation. Elections are conducted at particular time intervals. It is a way citizens can voice their opinions and exercise their rights. However, with the issues of maintaining credibility at the polling booths, they also raise concerns about certain malpractices. Everyone knows several election problems, such as using black money, horse trading, false promises, and violence. One such issue is duplicate or dummy votes in the online polling system [1], where the electorates can vote through another's unique vote ID in electronic voting machines (EVMs) and favors a particular candidate. Rikwith et al. [2] presented a research work to scale up the performance of EVM using fingerprint and face recognition using the R307 sensor and Raspberry Pi boards. Though it has been proven to be much better than ballot boxes [3], it has been reported to pose some issues of bias, accessibility, and accountability. Other research has been proposed; for instance, Mondal et al. [4] optimized EVMs through deep learning (DL)-based face recognition. A Region-based Convolutional Neural Network (R-CNN)-based approach has been described for face detection and feature extraction. Moreover, references [4,5] proposed a one-step authentication framework using a biometric system. This literature shows that the face recognition scenario under occlusion due to objects such as masks, glasses, etc., has not been considered. Such an inflection from the norm can disrupt systems, making them ineffective in detecting the faces.

Further, there have been many reported cases of phishing attacks on online voting systems, compromising the framework of the system and, consequently, its credibility. Nisha et al. [6,7] have discussed various possibilities of firewall infringements and reported security attacks in the past years, which have questioned the authenticity of the system. In the online voting system, the chances of malpractice are quite high. One such issue is logging in by an attacker under a different registered user account by morphing their face using various props. In this way, the attacker fulfills the face-matching feature and can vote by the wrong means. Thus, the system is susceptible to duplicate votes, which questions its credibility. Reference [8] discussed the advancements in morph generation and morphometric detection in face morphing attacks. It details various techniques to detect morphed images [8] effectively. Moreover, since 2020, the world has been struck by a global pandemic resulting in many deaths worldwide. It has been advised to avoid going to crowded places and mandatory to always wear masks. In such a scenario, online voting systems are considered safe for the elections. However, due to the mask, many voters can maneuver the verification process of the online polling system. Our system will work with real-time images, i.e., images recorded from the input video stream. Hence, there is a need to employ a technique for morph detection of real-time images.

The generative adversarial networks (GANs) framework can be used to solve the problem of morph detection through the reconstruction of the image. In the context of morph face detection, GANs can be used to generate a set of morphed images that are similar to the real images in the dataset. The GANs analyze the image and if find any change in the face feature space; then it removed the change to acquire the real image. The morphed images can then be used to train a classifier to detect morphed faces. GANs framework was first proposed by Goodfellow et al. [9] to estimate generative models via an adversarial process. It was a breakthrough discovery and gave excellent results when implemented on different datasets. Moreover, reference [9] proposed the idea of simultaneously training two models: a generative model and a discriminative model, which demonstrates a minimax two-player game. Various research works have been proposed for face recognition techniques under occlusion. In [10], the authors presented GANs with Visual Geometry Group (VGG)-19 as the base architecture developed for unmasking a masked face and regenerating a realistic image of the person. Several research works have been published by researchers across the globe presenting image-to-image translation techniques. It is observed that varied DL architectures like StarGAN [11], cycleGAN [12], DualGAN [13,14], etc., are trained to achieve the task of image-to-image translation. One such research work [15] was carried out by Isola et al., proposing a system for image-to-image translation using conditional GANs (cGANs). This research work uses the *Pix2pix* GAN-based framework, specifically employed for processing an image as the input in the online voting system. The problem of image morphing in an online voting system addressed in this paper requires implementing the reconstruction process of an input image recorded from the user's webcam. *Pix2pix* GAN architecture is hence used at the initial layer for this purpose.

Furthermore, the reconstructed image is matched with the user's image stored in the universal database, i.e., image matching has to be performed in the system. Hence, another network architecture, the Siamese network [16], is trained. It is a neural network that compares two image input matrices and finds their similar confidence score. The greater the value of this score, the more similar the input images are. In [17], the authors proposed a method to understand the general similarity function in the Siamese network using a hybrid convolutional neural network (HybridCNN) [18] as their principal network. Table 1 shows the relative comparison of existing state-of-the-art approaches and their research gaps. This paper proposes an authentication framework to verify user credentials in light of the aforementioned issues. The framework proposed in this paper aims to check the user credentials in a three-step manner. Firstly, it matches the unique ID with the same present in the universal database.

Table 1: A relative comparison of various state-of-the-art approaches

Author	Year	Methodology	Objective	Research gaps
Shah et al. [19]	2020	Blockchain technology	Blockchain-enabled online voting system that decentralizes control to make sure that fair voting occurs.	Not considered face recognition as an authentication mechanism.
Kumar et al. [20]	2020	Deep learning	Uses deep learning for face detection and recognition to authenticate users in the online voting system.	Fails to consider detection and recognition of occluded faces that can act as adversaries to evade the detection system.
Jafar et al. [21]	2021	Blockchain technology	Uses blockchain technology to ensure fair voting in the system.	Fails to consider face detection and recognition to ensure a single vote counts from a single individual.
Pooja et al. [22]	2021	Deep learning and blockchain technology	Combination of blockchain technology and deep learning to ensure fair and monitored online voting.	Authentication using face recognition is not implemented, and occluded faces can act as an adversary to evade the detection system.

(Continued)

Table 1 (continued)

Author	Year	Methodology	Objective	Research gaps
Parmar et al. [23]	2021	Blockchain technology	Secure and authenticated blockchain-based E-voting with face recognition.	System reliability and availability needs to be focused.
Alvi et al. [24]	2022	Blockchain technology	A blockchain-based secure mechanism for digital voting system.	Data storage cost is high and One Time Password is not implemented.
Proposed approach	2023	Deep learning	A three-layered DL-based authentication framework for on-line polling.	–

Consequently, it scans the user's live facial features, processes them using GANs trained model, and stores them encrypted in another database. Subsequently, these features are matched with the original image and uploaded with ID in the universal database to check if the user has a right to vote in the elections. Moreover, a separate encrypted database is also maintained in which the credentials of people who have already voted are stored to strengthen the performance of the polling system. Hence, in case of a security breach in which an attacker attempts to delete the electorate ids from the database, the system can still use the scanned facial features and check if similar embedding already exists in the encrypted database. Authentication fails if user embeddings already exist in this database.

1.1 Research Novelty

In recent years, numerous research works have been proposed to make online voting systems more secure and scalable. However, some of these frameworks [19,22] have not considered incorporating face recognition with authentication mechanisms. For example, Pooja et al. [22] proposed a secure deep-learning mechanism for fair online voting without any real-time implementation. Next, Alvi et al. [24] considered blockchain technology for digital voting systems, but there is no discussion on the One Time Password (OTP) mechanism for security purposes. To address the security gap in the literature, a new three-layered online voting system has been proposed that deploys a GANs framework for face recognition and verification. The proposed system [21] aims to enhance the security of online voting systems, which are still vulnerable to several security and data tampering attacks. The proposed system uses a dataset of original and partially covered faces to train the GANs model. The model aims to recognize and reconstruct faces under occlusion and authenticate them through a 2-step process using a trained Siamese network. The first step involves encoding the facial data, and the second step compares it with the encoded data in the database to ensure authenticity. Previous research works, such as [20] and [22], have used face detection without considering authentication for face recognition. The proposed system goes further to address this gap by incorporating GANs framework for face recognition and authentication to enhance the security of online voting systems. While some previous frameworks have partially solved issues related to privacy, authenticity, and robustness [21] in online voting systems, they are still vulnerable to security breaches and data tampering attacks. The proposed system aims to address this vulnerability by incorporating GANs framework for face recognition and verification. The system's effectiveness and efficiency can be tested through simulations and experiments to ensure its viability in real-world scenarios. Following are the research contributions of the proposed work.

1.2 Research Contributions

The main contributions of this paper are as follows:

- We propose a systematic framework to ameliorate the security of the online voting system. It is achieved by implementing three layered-system, i.e., GAN layer/reconstruction layer, the authentication, and the validation layer. The reconstruction of the image is consummated through *Pix2pix* GANs framework. Face detection is achieved on a video stream and is fed to this layer for processing and reconstruction.
- The framework consists of two-factor authentication, which integrates GANs framework with the credential matching process with the system's universal database.
- Further, a Siamese network is utilized for authentication purposes (image id verification), wherein a similarity confidence score is generated to interpret the output and to offer proactive decision-making.
- At the validation layer, the Kullback Leible (KL)-divergence metric calculates the scores between the feature embeddings obtained from the Siamese network and those in the encryption database. Thus, it ensures security and privacy to the online systems using the proposed three-layered authentication.

1.3 Organization of the Paper

The flow of the paper is as follows. Section 2 discusses the system model and problem formulation. Section 3 explains the proposed approach for the stated problem statement. Section 4 presents the outcomes of implementing the trained model, and Section 5 concludes the paper and provides insights into future works.

2 System Model and Problem Formulation

2.1 System Model

This section gives an overview of the proposed system to achieve security goals for the online voting system. Fig. 1 outlines the system model for handling security breaches with face id recognition and verification. It assumed that to log in to the system, access to the webcam should be enabled. Furthermore, it is assumed that privacy terms and conditions should be accepted before registering and logging in, especially ones emphasizing that video stream data is being monitored and stored in real time. When a login attempt is made, the input credentials are verified with the credentials stored in the universal database to check if the user is registered. After a successful verification, the webcam, which is kept on from the beginning, records the input video stream and transmits it to the AI interface. The AI interface consists of 3 layers: GANs, authentication, and validation layers. At the GANs layer, face detection and reconstruction are carried out using the *Pix2pix* GANs framework. The trained *Pix2pix* model detects any facial objects, like a face mask, glasses, etc., present on the face and reconstructs it and consequently stores the feature vectors obtained from the resulting image in an encrypted database. A trained Siamese network is implemented in the next layer, which checks for similarity scores between the reconstructed image and the original image id from the universal database. The final layer is the validation layer, where the KL divergence score is calculated between feature embeddings obtained and stored from the previous layer. The final portal gets activated only after successfully executing the AI interface and its decision. The layers and their functioning are discussed in detail in the upcoming sections.

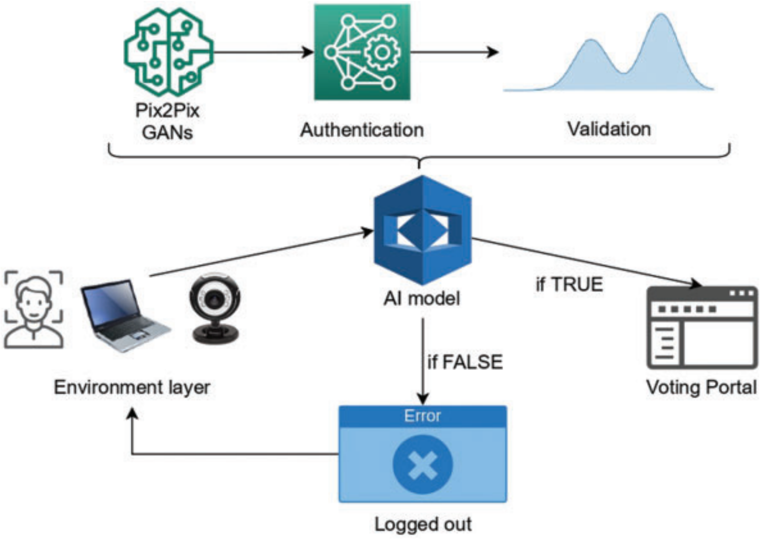


Figure 1: System model to achieve security goals for the online voting system

2.2 Problem Formulation

With the digitization wave, almost every sector has adopted online strategies to increase business or to make services available to users at their convenience. However, traditional electronic voting systems are not flexible and have a lot of hassles. A paramount event for any democratic state is conducting elections online by developing robust and secured voting interfaces. With online polling systems, more people who cannot reach the allotted voting booths can cast their votes easily. The online voting system can provide convenience by increasing accessibility. Moreover, the automated systems prevent long queues and waiting times, making the process highly efficient. However, some security issues might exist that can disrupt the voting system. A security breach in a voting system can pose severe risks or even destroy a democratic state. Reference [6] detailed various phishing attacks that can occur in the online voting system.

This paper has addressed a possible issue that can cause a security breach in the most crucial issue of an online voting system. The module discussed here is face id recognition and verification. Consider the following scenario: a voter V with a unique voter id V_{id} log into the online system. We assume a universal database Ψ containing the existing and registered unique voter id and id image V_{image}. A citizen can only cast their vote if, firstly, $V_{id}\varepsilon\Psi$ and face input image V_{inp} from the camera video stream and V_{image} match with each other. (Note: It is assumed that proper facilities of webcam and internet are available to the voter.) The physical appearance of a person, especially a face, can be easily morphed to look like someone else. This can lead to fallacious registration of the vote by mismatched voter id and the actual id face.

$$V^{a}_{id}, V^{a}_{inp} \xrightarrow{\text{(resemble)}} V^{a}_{id}, V^{b}_{image} \tag{1}$$

where a and b are two instances from the universal database. Hence, fake voter registration is a concern caused by physically doctoring one's appearance. This paper aims to prevent such altercation by employing image reconstruction using a DL approach. The reconstructed image $V^x_{inp'}$ for a person, x can be produced, which can be compared with V^x_{image} with authenticated voter id V^x_{id}. Trained image verification pipeline $IV\left(V^x_{id}, V^x_{inp'}\right)$ generates a confidence score λ for image similarity.

$$\lambda = IV\left(V^x_{id}, V^x_{inp'}\right) \tag{2}$$

All the registered votes are stored in a separate database Δ. A major issue with such a two-layered approach is the possibility of re-registration of votes using the same voter id in case of an attack α_{pkc}. An attack α_{pkc} (primary key compromise attack) can be developed and deployed such that it queries to delete the tokenization voter id V^x_{id} for a specific person x, such that a new redundant vote can be recast again from the same voter id V^x_{id}.

$$\alpha_{pkc} = \text{DELETE FROM } \Delta \text{ WHERE } id = V^x_{id} \tag{3}$$

Hence, to solve the security issues identified above, a three-layered framework is proposed in the forthcoming section.

3 Proposed Approach

The problem formulation has discussed a few possible problems in implementing a classic system. Fig. 2 presents the AI layer of the proposed system model that aims to prevent shortcomings of the online electoral system. This section discusses the proposed three-layered framework to prevent physical security breaches for the approach taken into consideration.

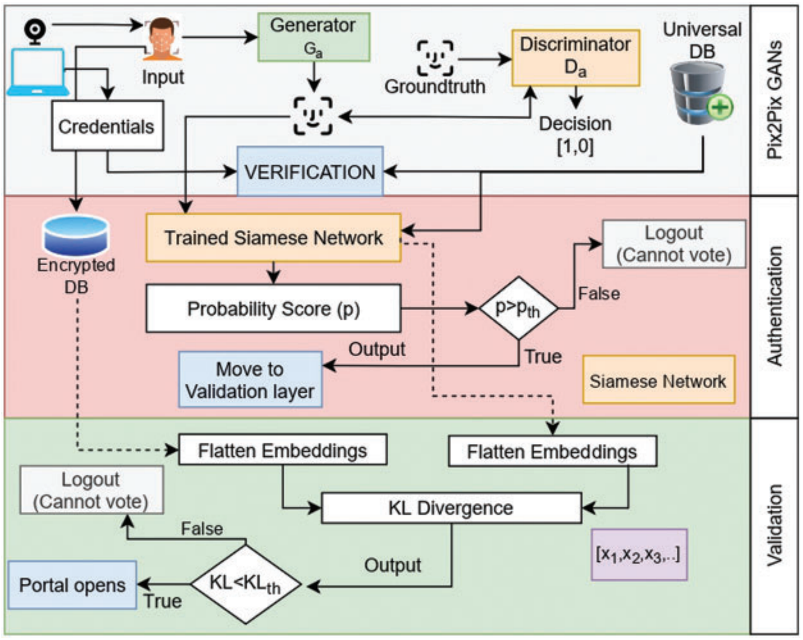


Figure 2: The proposed AI layer

3.1 GAN Layer

The GAN layer is the foremost layer in the decision layer of the proposed framework. This layer consists of a trained generative model that aims to generate a reconstructed image of the input video stream of the electorate's face. In this subsection, we describe the generative model training, prediction, and dataset.

Dataset Description—No publicly available dataset contains image pairs of faces with and without attached objects such as specs, masks, hats, sunglasses, etc. Hence, we created a synthetic dataset that matches the use case description. The data collection step involves obtaining face images of different people with and without facial objects attached to them. A total of 20 image pairs were collected from 5 different people. Images of the high resolution were later resized to 256 × 256 to reduce computational expenses. The images are captured inside a light-illuminated room of a machine learning laboratory on the university campus. A dataset is prepared such that an image pair is captured and obtained for a person. This pair contains two different images of the same person. During the dataset preparation process, we restricted the face images to only the front face, not the lateral one. Furthermore, we only considered cases with only one face per frame. Detecting multiple faces in a frame is another research area to work on in the future. Given the novelty of the research, we constrained our dataset to only clear-face images and exempted frames with blurred faces and cases with different illumination conditions. Fig. 3 shows a sample image pair present in the dataset. Label image (Fig. 3b) here refers to the image that the generator sub-layer is supposed to reconstruct/regenerate given an original input image (Fig. 3a) with several attached facial objects, i.e., a blue face mask and glasses.

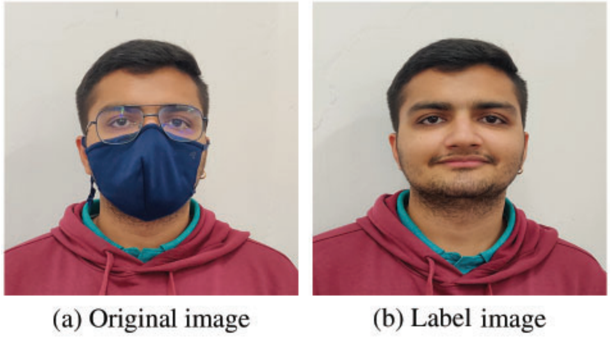

(a) Original image
(b) Label image

Figure 3: Sample images pair from the dataset

Dataset Preparation and Processing—The data preparation and processing step involves making the image data fit for training and testing. Let I be the set of P image pairs denoted by i_p.

$$I = \{i_1, i_2, \ldots, i_p, \ldots, i_P\} \tag{4}$$

$$i_p = \left(o_p, l_p\right) \tag{5}$$

$$\forall_{\mathrm{p}} 1 \leq p \leq P$$

where o_p and l_p are the original and label images, respectively, for the p^{th} instance in the dataset. Each $i_p = \left(o_p, l_p\right) \in I, o_p,$ and l_p with size ($h' \times w'$) are resized to a static shape ($h \times w$).

$$i = \left(o_p^{h'\times w'}, l_p^{h'\times w'}\right) \rightarrow \left(\mathrm{o}_{\mathrm{p}}^{\mathrm{h}\times\mathrm{w}}, l_p^{h\times w}\right) \tag{6}$$

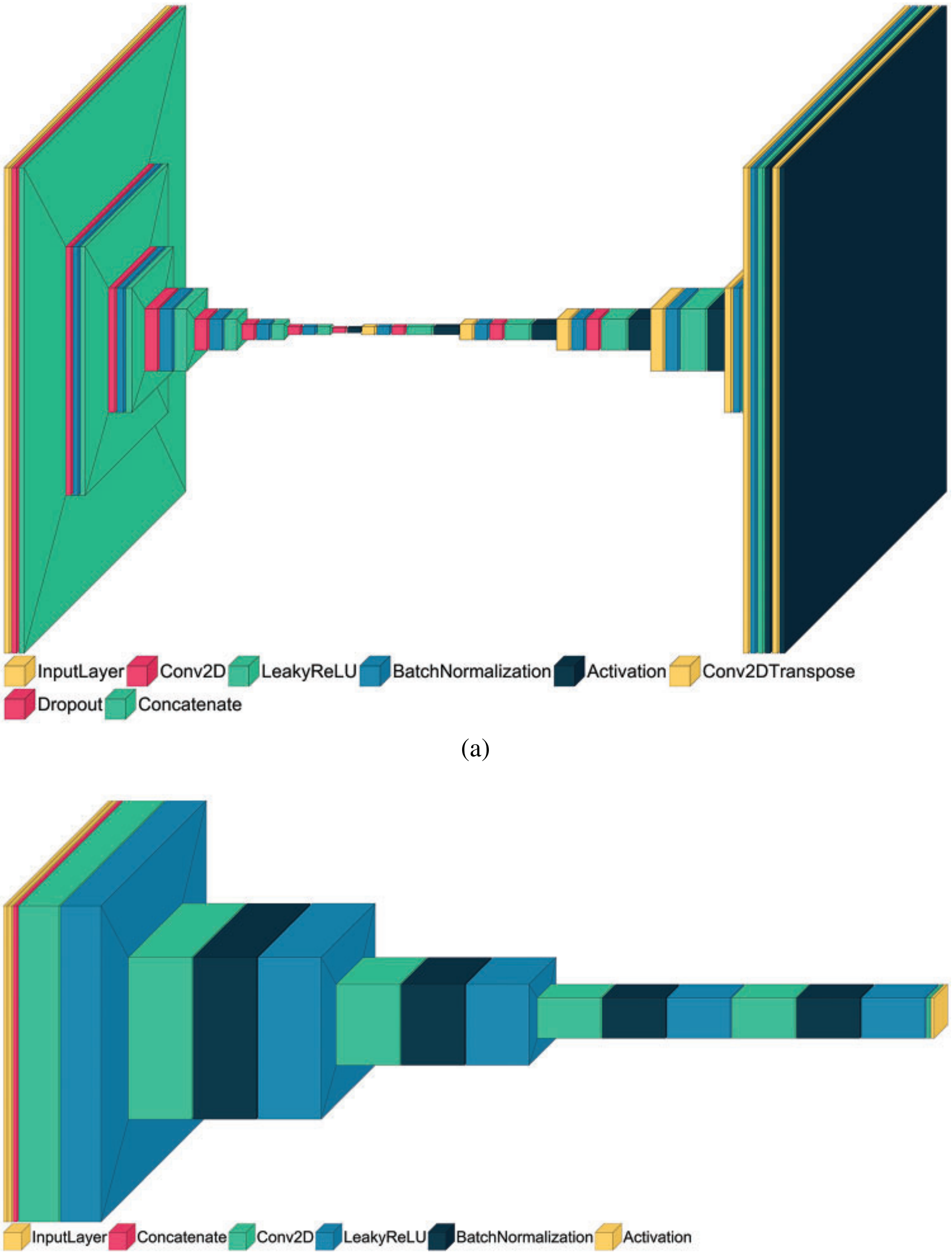
Note: Total number of color channels in the images is kept constant, i.e., three.

The final pre-processed dataset I_m is obtained, where m denotes the total number of instances of image pairs. Furthermore, the image pixel values having a range [0, 255] are also scaled to a new range of [−1, 1] using the min-max scaler.

$$i_p = \frac{i_p - 127.5}{127.5} \tag{7}$$

Generative Adversarial Networks—The generative adversarial networks are deep learning-based generative models. Such a network [9] of the model was first developed and presented by Goodfellow. A basic GAN architecture is divided into two main sub-models: generator and discriminator. The generator model learns to mimic the original training data, whereas the discriminator [25] learns to classify it as real or generated. GANs can be used in face recognition to generate realistic synthetic faces that can be used for training face recognition models. This can be useful when there is a limited amount of real face data available for training or when the data needs to be augmented to improve the robustness of the model. In an online voting system, GANs can be used to generate synthetic faces of voters, which can be used to facilitate remote voting. This can be useful when voters cannot physically attend a voting station during a pandemic or in remote areas. The GAN-generated faces can be used to verify the voter's identity, and the system can compare the synthetic face with the real face of the voter using face recognition technology. The pre-trained generator model weights can be preserved to generate newer instances with nearly equivalent distributions as the training data. There are several GAN architectural variants. These variants are applied to different fields involving the usage of AI. This paper uses the image-to-image (I2I) translation method, which renovates/reconstructs the input face images by learning a masking operation. *Pix2pix* GANs [26], which employs supervised learning, is used here for the I2I translation task. However, it is important to note that there are ethical and legal considerations around using GANs in online voting systems, as there is a risk of impersonation and fraud. Therefore, robust security measures and regulations must be implemented to ensure the voting process's integrity. Therefore, we consider an authentication mechanism with the Siamese network along with GANs, which strengthens the current voting systems.

GANs Training—*Pix2pix* GAN is trained on the collected image dataset. We were motivated by the work of Isola et al. [26,27], where they constructed an efficient model. However, the training process is inspiringly carried out from [28], showing an efficient training time. The generator model and the discriminator are trained separately on a given training batch. Fig. 4a shows the generator sub-layers for the GAN model. A generator here has two components: an encoder and a decoder, which makes it a U-Net [29] like architecture. Fig. 4b shows the discriminator sub-layer for the model. The discriminator aims to train the generator for better image generation. For input data represented by set I, consider generator sub-model G and discriminator sub-model D. Also, let the fully connected GAN model be represented as $\zeta\left(G, D\right)$. Consider an original image $\chi^{o}_{h\times w\times 3}$ be passed to the generator sub-model G.



(a)

(b)

Figure 4: (a) Generator sub-model and (b) Discriminator sub-model

$$Z_{h\times w\times 3} = G\left(\chi^{0}_{h\times w\times 3}\right) \tag{8}$$

$$\chi^{0}_{h\times w\times 3} \in i^{0p}_{p}$$

where Z is the image reconstructed by the generator sub-layer in a single pass, it can be observed that the image shape remains maintained. The generated image Z and the original image χ can be passed to the discriminator to classify it as real or fake. The GANs loss function $L_{G,D}$ can be defined as follows:

$$L_{G,D} = -(c_{\chi} log(\hat{y} + (1 - c_{\chi}) log(1 - \hat{y})) \tag{9}$$

$$\hat{y}_{d_1 \times d_2 \times d_3} = D\left(\left[\chi^0, Z\right]\right) \tag{10}$$

$$c_{\chi} \rightarrow \mathrm{d}_1 \times \mathrm{d}_2 \times \mathrm{d}_3$$

where c refers to the class labels of the input image, i.e., whether the image is real or fake. Note that c has dimensions equivalent to the discriminator's output layer. Moreover, a separate generator loss is added to the overall model loss to train the model more concretely and have near-to-ground truth generations. Generator loss L_G is defined as L_1 distance of Z generated from χ^0 w.r.t the real label images χ^l. After the GAN model is trained, the reconstructed image of the input stream is obtained as Z. Here,

$$Z = G\left(\chi^0\right) \tag{11}$$

$$L_G = \left|\left|\chi^l - Z\right|\right| \tag{12}$$

$$\chi^l_{h \times w \times 3} \in i^{lp}_p$$

Hence the final loss function L_{GAN} [26] is written as

$$L_{GAN} = argmin_G max_D L_G + L_{G,D} \tag{13}$$

Training configuration for the GAN layer is provided in Table 2.

Table 2: Training configuration for GAN model

Epochs	200
Learning Rate	0.0002
Optimizer	Adam

3.2 *Authentication Layer*

The authentication layer aims to cross-check the reconstructed input face image with the existing unique login and image ids database. The objective of this layer is to employ a trained Siamese network [16] for the task of image verification.

Siamese Network Training—Similar to the GAN layer, the authentication layer consists of a deep neural network that does the task of image id verification. A simple Siamese network is trained for the same purpose. It is a type of neural network architecture that is designed to compare two different inputs and determine how similar they are. It was originally proposed for signature verification but has since been applied to various tasks, such as face recognition, image retrieval, and natural language processing. A Siamese network consists of two identical sub-networks which share the same weights and architecture. The two sub-networks take in the two input samples to be compared and extract their features. The outputs of the two sub-networks are then compared using a similarity metric such as Euclidean distance or cosine similarity, which produces a similarity score that indicates how similar the two inputs are. The advantage of using a Siamese network is that it can be trained on a small dataset of pairs of inputs, rather than a large dataset of individual samples. This makes it particularly useful for tasks where labeled data is scarce, such as one-shot learning and few-shot learning. This model takes in two input images of the same shape and outputs a confidence score pertaining to their mutual similarity. A higher output probability for a given pair of images implies a greater visual similarity in them. Training configuration for the Siamese network is provided in Table 3. In addition, Fig. 5 shows the network architecture constructed for image verification. We refer to reference [30] to construct and train this model.

Table 3: Training configuration for Siamese model

Epochs	256
Learning Rate	0.001
Optimizer	Adam

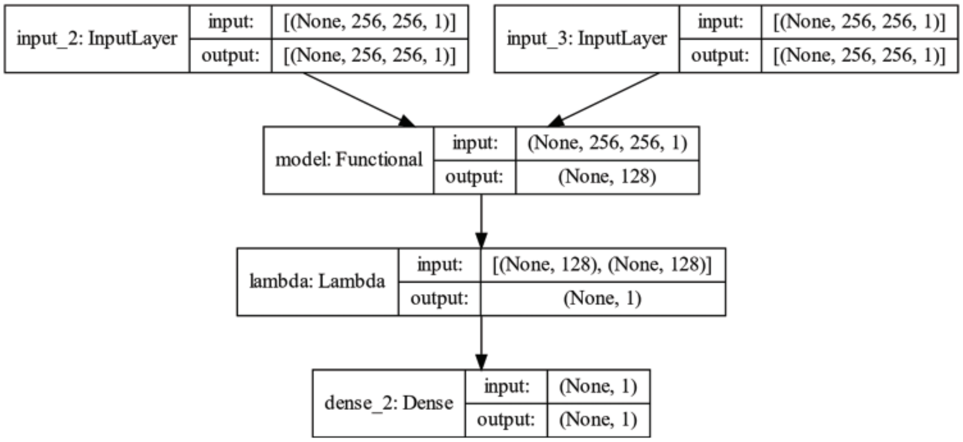


Figure 5: Siamese network architecture

Consider a training dataset S of size s containing the images ι_{id} and ι_{gen} to be compared and their ground truth class labels μ. It is important to note here that generated image ids are resized to a static shape of 256 × 256 × 1 and also transformed to grayscale.

$$S = \cup_s \left(\left[\iota_{id}^k, \iota_{gen}^k\right], \mu\right) \tag{14}$$

$$\iota = 256 \times 256 \times 1 \tag{15}$$

$$\forall_k 1 \leq k \leq s$$

The objective of the Siamese network can be mathematically represented as follows:

$$\underset{L_{Siamese}}{min}\ L_{Siamese} = -vlog\left(\hat{v}\right) - (1 - v)\,log\left(1 - \hat{v}\right) \tag{16}$$

$$\hat{v} = \sigma\left(distance\right) \tag{17}$$

$$diatance = \sum_{k=1}^{S} \sqrt{\left(E_{id}^k - E_{gen}^k\right)^2} \tag{18}$$

$$E^{k}_{id_{128\times 1}} = Functional\left(\iota^{k}_{id}\right) \tag{19}$$

$$E^{k}_{gen_{128\times 1}} = Functional\left(\iota^{k}_{gen}\right) \tag{20}$$

where E refers to the Siamese feature embeddings obtained at the output of the functional layer in Fig. 5. Also σ refers to the Sigmoid activation function.

Siamese Predictions—A universal database Ψ is considered to contain tuples of unique electorate ids ω and image ids O.

$$\Psi = \cup_{size}\left(\omega_i, O_i\right) \tag{21}$$

$$\forall_i i \leq size, \{size, i\} \in W$$

For a correctly identified and existing unique id ω_j, where $j \leqslant size$, image id O_j, and GAN-generated image Z are passed on to the trained Siamese network Γ. A probability score $p_{\omega j}$ is obtained, which is used further for decision-making.

$$p_{\omega_j} = \Gamma\left(\left[O_j, Z\right]\right) \tag{22}$$

If $p_{\omega j} \geqslant p_{th}$, the feature embeddings for the generated image E^{Z}_{gen} are stored in a database, and the credential verification process proceed to its third layer, i.e., the validation layer.

3.3 Validation Layer

This is the only layer where no neural network is trained and used for predictions. This layer works as the final step for electorate id verification. For the given authenticated unique ids from the second layer, feature embedding $E^{Z_z}_{gen}$ for the generated GAN image are stored in a database called the encryption database Φ of size R.

$$\Phi = \cup_R E^{Z_z}_{gen} \tag{23}$$

$$\forall_z z \leq R, R \varepsilon W$$

An authenticated image with feature embeddings E^{ω}_{gen} (ω refers to the user's unique id) is queried through the database, and divergence values are calculated using a particular metric. The main objective of the validation layer is to obtain divergence scores between existing data of feature embeddings and new feature embeddings to check for redundancy. KL-divergence [31] is used as a metric here. A higher KL-divergence score suggests a greater difference in values of the two input distributions proving that the electorate image id has still not been registered.

$$KL(E^{\omega}_{gen}||E^{Z_z}_{gen}) = \sum_{e=1}^{128} E^{\omega}_{gen}[e]\, log_2 \frac{E^{\omega}_{gen}[e]}{E^{Z_z}_{gen}[e]} \tag{24}$$

$$E^{Z_z}_{gen}[e] \in \Phi$$

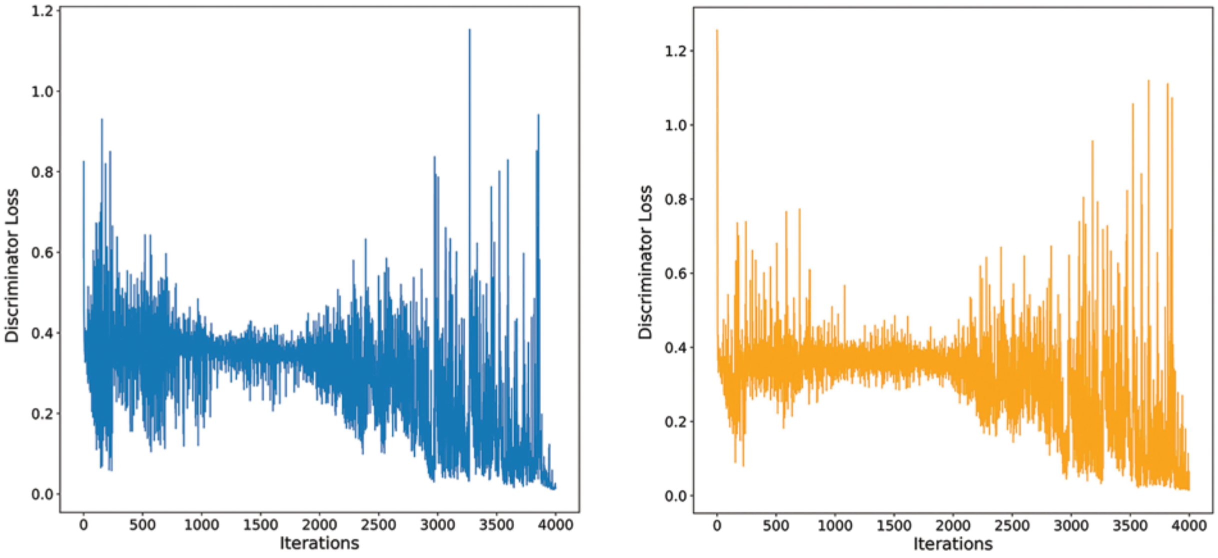
The detailed analysis of results for all three layers is carried out in the forthcoming section.

4 Results and Discussions

This section presents a detailed description of the performance and evaluation of the individual components of the proposed three-layered system framework.

4.1 GANs Results

The purpose of the generator in the proposed architecture is to extract and reconstruct the original face without external facial objects such as glasses and face masks. *Pix2pix*, which is a GANs-based approach described in the proposed architecture, is trained and used for the predictions. Fig. 6 shows the loss curves *vs.* iterations for the discriminator of the GAN model. The models were trained for 4000 iterations, and it can be observed that the overall losses do decrease at the end of the total epochs. The loss curve for the generator sub-layer of the *Pix2pix* GAN architecture is also plotted to check its training performance. Fig. 7 shows the training loss curve for the generator sub-layer. As the model learns to reconstruct the original images by masking out the facial objects, the relative divergence between the predicted training and the original images depicted by the generator loss function also decreases with the iterations.



(a) Loss 1

(b) Loss 2

Figure 6: Discriminator losses for the proposed GAN model

As the generator model gets trained gradually, the reconstructed generated output for a given input image also improves. A series of images for different facial configurations are passed through the generator at various intermediate training steps to obtain several reconstructed images. These images are juxtaposed to evaluate and visualize the iterative training of the generator sub-layer. Fig. 8 shows the regeneration of original images using the generator model to obtain faces without the face mask and glasses.

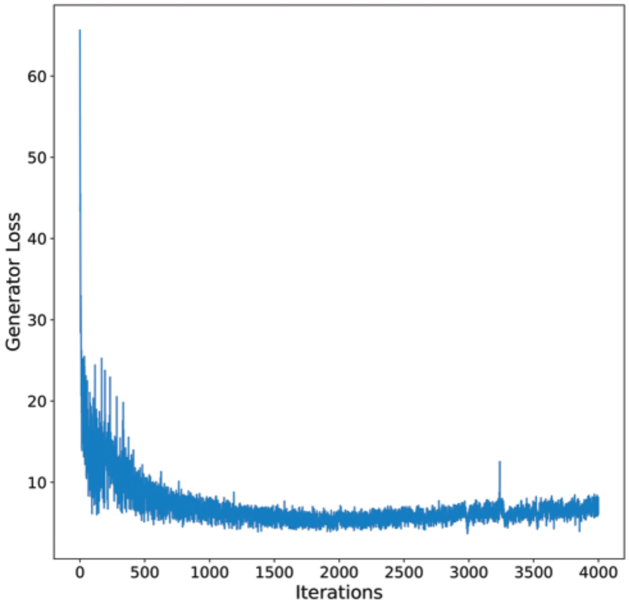


Figure 7: Training curve for the generator loss function

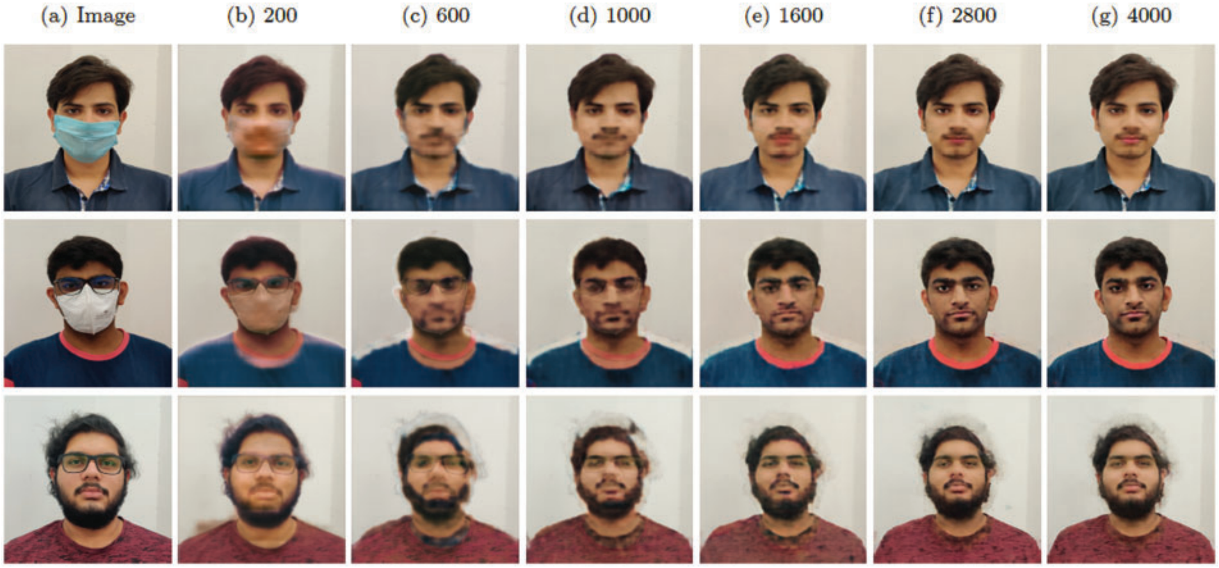


Figure 8: Reconstructing the original images using the trained generator sub-layer. (a) Original image from the dataset. (b) Generator output images at iterations (b) 200 (c) 600 (d) 1000 (e) 1600 (f) 2800 (g) 4000

4.2 Siamese Network-Based Results

The Siamese network is used to validate the id image existing in the universal credential database. The reconstructed image by the generator of the trained GAN architecture can be compared with the id image using a trained Siamese network, as shown in Fig. 8. A return of a higher probability score suggests a greater similarity between the compared input images. Fig. 9 shows the training loss curve for the Siamese network. The model loss value decreases as the model trains gradually. Fig. 10 shows similar images that are taken into consideration for comparison. A similarity confidence score of 0.72 was obtained when these images were passed to the proposed Siamese network. This essentially implies that the input images are highly likely to be similar and of the same person. Fig. 11 depicts the dissimilar images that are taken into consideration. A confidence score of 0.02 was obtained on the trained Siamese network implying that the images are not of the same person, which stands out to be the truth.

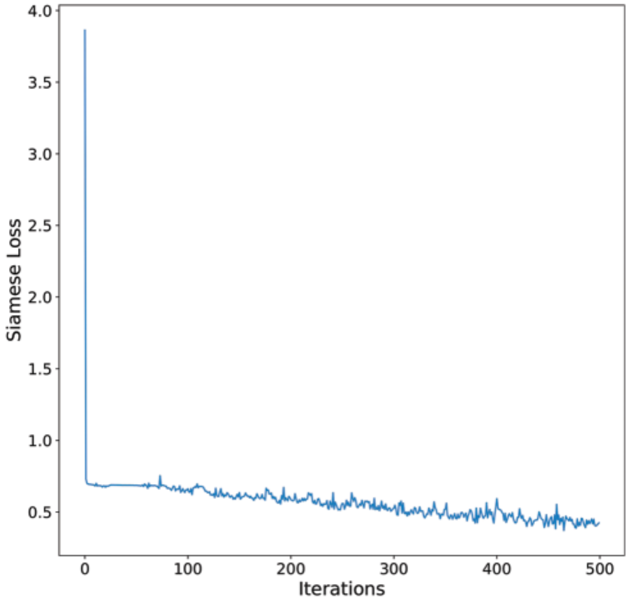
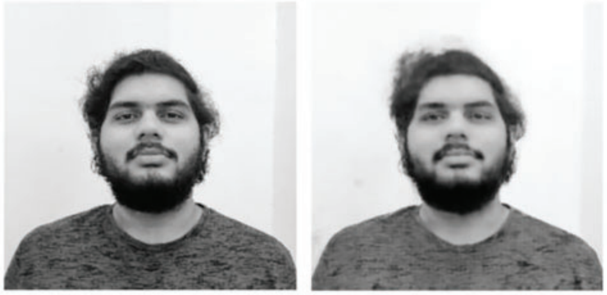


Figure 9: Siamese network's training loss w.r.t total iterations



(a) Original image id (b) Generated image

Figure 10: Comparing original id image (a) and reconstructed image (b) using the Siamese network

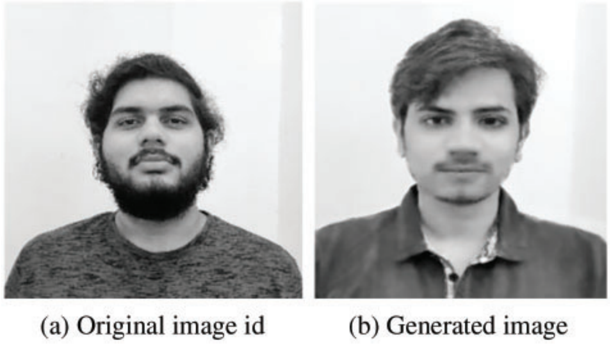

(a) Original image id (b) Generated image

Figure 11: Comparing original image id (a) and test image (b) using the Siamese network

4.3 Validation Layer Results

The third layer of the proposed framework, as discussed earlier, is the validation layer. The Siamese network embeddings for each input image are stored unanimously. The results of the authentication parameter defined by the KL-divergence are obtained. Figs. 12 and 13 show the image embeddings obtained from the Siamese network. For images, Figs. 10a and 10b, a KL-divergence of 1.1855 is obtained, implying that similar images have similar face embeddings hence a lesser KL-divergence value. For images in Figs. 11a and 11b, KL-divergence of 50.1229 is obtained, implying that dissimilar images have far from equivalent face embeddings.

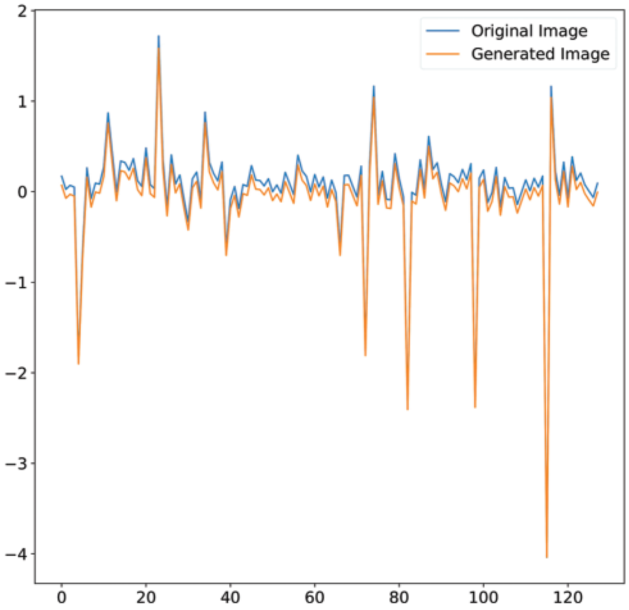


Figure 12: Siamese network embeddings comparison for Fig. 10 images. KL-divergence for the above-shown distributions is 1.1855

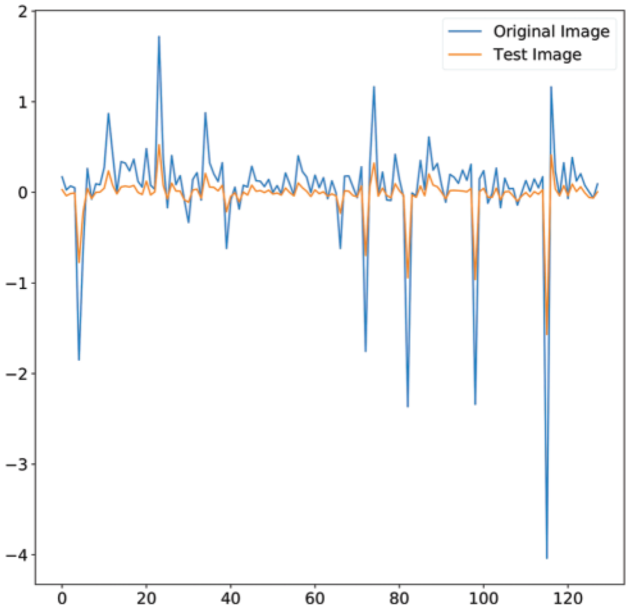


Figure 13: Siamese network embeddings comparison for Fig. 11 images. KL-divergence for the above-shown distributions is 50.1229

5 Conclusion and Discussions

This paper highlights one of the potential risks of organizing an online voting system. It confers a probable framework based on GANs model and Siamese network to avoid future security breaches. Generator and discriminator models have been trained, and their performances have been studied using loss functions. It gave at-par results with the reconstruction of facial features, as seen in the results section. The resulting metrics of the Siamese network, namely probability scores and KL-divergence, are considered for authentication and validation layers, respectively. These metrics have given promising results and shown greater KL divergence between different images, thereby successfully detecting fraud.

In the future, we plan to expand our dataset containing all the possible image orientations. The dataset would consist of the cases where a lateral view of the face image is detected, on which the system will state an unsuccessful login attempt. This will make the system highly efficient and robust. Furthermore, the dataset can also include faces under different illumination conditions and faces under occlusion with varying objects like hair, hat, etc. The research is currently focused on proposing a novel approach to an online voting system. In future works, we plan to optimize the system with training on a diversified dataset and enhance the framework.

Acknowledgement: The authors would also like to acknowledge all members of Sudeep Tanwar's Research Group (ST Lab) for their support in revising this manuscript. Further, the authors would like to acknowledge editorial board members of CMC and reviewers for providing technical comments to improve overall scientific depth of the manuscript.

Funding Statement: This work was funded by the Researchers Supporting Project Number (RSP2023R 102) King Saud University, Riyadh, Saudi Arabia.

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Rajesh Gupta, Nilesh Kumar Jadav, Priyal Palkhiwala, Osama Alfarraj, Sudeep Tanwar; data collection: Harsh Mankodiya, Priyal Palkhiwala, Rajesh Gupta, Nilesh Kumar Jadav, Amr Tolba; analysis and interpretation of results: Harsh Mankodiya, Sudeep Tanwar, Maria Simona Raboaca, Amr Tolba, Verdes Marina; draft manuscript preparation: Osama Alfarraj, Maria Simona Raboaca, Verdes Marina. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Not applicable.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] M. Ahmad, A. U. Rehman, N. Ayub, M. Alshehri, M. A. Khan *et al.,* "Security, usability, and biometric authentication scheme for electronic voting using multiple keys," *International Journal of Distributed Sensor Networks*, vol. 16, no. 7, pp. 1–16, 2020.

[2] S. L. Rikwith, D. Saiteja and R. Jayaraman, "Enhancement of electronic voting machine performance using fingerprint and face recognition," in *2nd Int. Conf. on Smart Electronics and Communication (ICOSEC)*, Trichy, India, pp. 757–763, 2021.

[3] D. A. Kumar and T. U. S. Begum, "Electronic voting machine a review," in *Int. Conf. on Pattern Recognition, Informatics and Medical Engineering (PRIME)*, Salem, India, pp. 41–48, 2012.

[4] I. Mondal and S. Chatterjee, "Secure and hassle-free EVM through deep learning-based face recognition," in *2019 Int. Conf. on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, Faridabad, India, pp. 109–113, 2019.

[5] S. Hussain, Y. Zikria, G. A. Mallah, C. M. Chen, M. D. Alshehri *et al.,* "An improved authentication scheme for digital rights management system," *Wireless Communications and Mobile Computing*, vol. 2022, no. 1871, pp. 1–11, 2022.

[6] S. Nisha and A. N. Madheswari, "Prevention of phishing attacks in voting system using visual cryptography," in *2016 Int. Conf. on Emerging Trends in Engineering, Technology and Science (ICETETS)*, Pudukkottai, India, pp. 1–4, 2016.

[7] S. Banerjee, V. Odelu, A. K. Das, S. Chattopadhyay, N. Kumar *et al.,* "Design of an anonymity- preserving group formation-based authentication protocol in global mobility networks," *IEEE Access*, vol. 6, pp. 20673–20693, 2018.

[8] S. Venkatesh, R. Ramachandra, K. Raja and C. Busch, "Face morphing attack generation & detection: A comprehensive survey," *IEEE Transactions on Technology and Society*, vol. 2, no. 3, pp. 128–145, 2020.

[9] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley *et al.,* "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[10] N. Ud Din, K. Javed, S. Bae and J. Yi, "A novel GAN-based network for unmasking of masked face," *IEEE Access*, vol. 8, pp. 44276–44287, 2020.

[11] Y. Choi, M. Choi, M. Kim, J. W. Ha, S. Kim *et al.,* "StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 8789–8797, 2018.

[12] J. Y. Zhu, T. Park, P. Isola and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. of the IEEE Int. Conf. on Computer Vision (ICCV)*, Venice, Italy, pp. 2242–2251, 2017.

[13] Z. Yi, H. Zhang, P. Tan and M. Gong, "Dualgan: Unsupervised dual learning for image-to-image translation," in *Proc. of the IEEE Int. Conf. on Computer Vision (ICCV)*, Venice, Italy, pp. 2868–2876, 2017.

[14] K. Patel, D. Mehta, C. Mistry, R. Gupta, S. Tanwar *et al.,* "Facial sentiment analysis using AI techniques: State-of-the-art, taxonomies, and challenges," *IEEE Access*, vol. 8, pp. 90495–90519, 2020.

[15] P. Isola, J. Y. Zhu, T. Zhou and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, pp. 1125–1134, 2017.

[16] J. Bromley, I. Guyon, Y. LeCun, E. Sackinger, R. Shah *et al.,* "Signature verification using a siamese time delay neural network," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 7, no. 4, pp. 737–744, 1993.

[17] I. Melekhov, J. Kannala and E. Rahtu, "Siamese network features for image matching," in *2016 23rd Int. Conf. on Pattern Recognition (ICPR)*, Cancun, Mexico, pp. 378–383, 2016.

[18] X. X. Niu and C. Y. Suen, "A novel hybrid CNN-SVM classifier for recognizing handwritten digits," *Pattern Recognition*, vol. 45, no. 4, pp. 1318–1325, 2012.

[19] A. Shah, N. Sodhia, S. Saha, S. Banerjee and M. Chavan, "Blockchain enabled online-voting system," in *Int. Conf. on Automation, Computing and Communication 2020 (ICACC-2020)*, Mumbai, India, vol. 32, pp. 1–6, 2020.

[20] G. Kumar, R. Kaushik, K. Kumar and V. Birla, "Design and development to face recognition system by synthesizing aadhar enabled platform for online voting," *International Journal of Innovative Research in Computer Science & Technology (IJIRCST)*, vol. 8, no. 3, pp. 1–5, 2020.

[21] U. Jafar, M. J. A. Aziz and Z. Shukur, "Blockchain for electronic voting system review and open research challenges," *Sensors*, vol. 21, no. 17, pp. 1–22, 2021.

[22] S. Pooja, L. K. Raju, U. Chhapekar and C. B. Chandrakala, "Face detection using deep learning to ensure a coercion resistant blockchain-based electronic voting," *Engineered Science*, vol. 16, pp. 341–353, 2021.

[23] A. Parmar, S. Gada, T. Loke, Y. Jain, S. Pathak *et al.,* "Secure e-voting system using blockchain technology and authentication via face recognition and mobile OTP," in *2021 12th Int. Conf. on Computing Communication and Networking Technologies (ICCCNT)*, Kharagpur, India, pp. 1–5, 2021.

[24] S. T. Alvi, M. N. Uddin, L. Islam and S. Ahamed, "DVTChain: A blockchain-based decentralized mechanism to ensure the security of digital voting system voting system," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 9, pp. 6855–6871, 2022.

[25] A. Jabbar, X. Li and B. Omar, "A survey on generative adversarial networks: Variants, applications, and training," *ACM Computing Surveys*, vol. 54, no. 8, pp. 1–49, 2021.

[26] P. Isola, J. Y. Zhu, T. Zhou and A. A. Efros, "Image-to-image translation with conditional adversarial networks," *ACM Computing Surveys*, vol. 54, no. 8, pp. 1–49, 2018.

[27] Pix2pix Model Repository, 2023. [Online]. Available: https://github.com/phillipi/pix2pix

[28] Pix2pix Training Process, 2022. [Online]. Available: https://machinelearningmastery.com/how-to-develop-a-pix2pix-gan-for-image-to-image-translation/

[29] O. Ronneberger, P. Fischer and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention*, Munich, Germany, vol. 9351, pp. 234–241, 2015.

[30] Siamese Network, 2022. [Online]. Available: https://medium.com/wicds/face-recognition-using-siamese-networks-84d6f2e54ea4

[31] J. M. Joyce, "Kullback-Leibler divergence," in *International Encyclopedia of Statistical Science*, vol. 1. Heidelberg, Germany: Springer, pp. 720–722, 2014.