



---

*Research article*

## **Comparative analysis of GAN-based fusion deep neural models for fake face detection**

**Musiri Kailasanathan Nallakaruppan<sup>1,\*</sup>, Chiranji Lal Chowdhary<sup>1</sup>, SivaramaKrishnan Somayaji<sup>1</sup>, Himakshi Chaturvedi<sup>2</sup>, Sujatha. R<sup>3</sup>, Hafiz Tayyab Rauf<sup>4,\*</sup> and Mohamed Sharaf<sup>5</sup>**

<sup>1</sup> School of Computer Science Engineering and Information Systems, Vellore Institute of Technology, Vellore, Tamilnadu 632014, India

<sup>2</sup> School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamilnadu 632014, India

<sup>3</sup> School of Electronics and Communication Engineering, Vellore Institute of Technology, Vellore, Tamilnadu 632014, India

<sup>4</sup> Centre for Smart Systems, AI and Cybersecurity, Staffordshire University, Stoke-on-Trent ST4 2DE, UK

<sup>5</sup> Industrial Engineering Department, College of Engineering, King Saud University, P.O. Box 800, Riyadh 11421, Saudi Arabia

\* **Correspondence:** Email: [nallakaruppan.mk@vit.ac.in](mailto:nallakaruppan.mk@vit.ac.in), [hafiz-tayyabrauf093@gmail.com](mailto:hafiz-tayyabrauf093@gmail.com).

**Abstract:** Fake face identity is a serious, potentially fatal issue that affects every industry from the banking and finance industry to the military and mission-critical applications. This is where the proposed system offers artificial intelligence (AI)-based supported fake face detection. The models were trained on an extensive dataset of real and fake face images, incorporating steps like sampling, pre-processing, pooling, normalization, vectorization, batch processing and model training, testing-, and classification via output activation. The proposed work performs the comparative analysis of the three fusion models, which can be integrated with Generative Adversarial Networks (GAN) based on the performance evaluation. The Model-3, which contains the combination of DenseNet-201+ResNet-102+Xception, offers the highest accuracy of 0.9797, and the Model-2 with the combination of DenseNet-201+ResNet-50+Inception V3 offers the lowest loss value of 0.1146; both are suitable for the GAN integration. Additionally, the Model-1 performs admirably, with an accuracy of 0.9542 and a loss value of 0.1416. A second dataset was also tested where the proposed Model-3 provided maximum accuracy of 86.42% with a minimum loss of 0.4054.

**Keywords:** GAN; *Inception V3*; RESNET-50; Mobile Net V2; DenseNet-201; VGG-16; Xception

---

## 1. Introduction

The proliferation of fake images has become a significant challenge in the digital age, with deep-fake technology being used to create realistic fake faces that can be difficult to distinguish from genuine ones. Fake identity issues pose a significant threat across various sectors, with particularly severe implications in defence and finance. The ability to accurately detect fake faces is crucial in safeguarding sensitive information and maintaining the integrity of operations in these domains. The prevalence of fake identity issues underscores the urgency for robust solutions, especially as technologies advance.

The detection of fake faces remains a formidable challenge, with existing methods facing several hurdles. From the imperfections in traditional biometric systems to the increasing sophistication of fraudulent attempts, the need for a more reliable solution is evident. Existing methods often struggle with accuracy and are susceptible to manipulation. This study aims to address these challenges by proposing GAN based fusion models, which leverage the power of DL to overcome the limitations of conventional approaches. The challenges in fake face detection serve as a driving force behind the development and exploration of innovative solutions.

This research sets out with clear objectives aimed at revolutionizing fake face detection. The primary goal is to harness the capabilities of GANs and fusion models to significantly enhance accuracy and reduce vulnerabilities in identifying fake faces. The proposed models bring advancements by offering a more nuanced understanding of the loss function and accuracy of GANs, introducing the potential of federated learning for accuracy and loss estimation, and conducting a thorough comparative analysis between single DL models and fusion models. These contributions are poised to elevate the effectiveness of fake face detection, providing a cutting-edge solution to the pervasive challenges faced by existing methods.

GANs are deep neural network models that use a generator and a discriminator to generate realistic images and distinguish them from real ones, respectively. GAN are deep neural network models that were introduced by Goodfellow et al. in 2014 [1]. The network consists of two parts: The generator and the discriminator. The generator produces fake data, while the discriminator tries to identify the fake data from the real data.

GANs can be useful in detecting fake data by training a discriminator on a dataset of real and fake data. Once the discriminator is trained, it can be used to identify fake data with high accuracy. This can be particularly useful in detecting deepfakes, where a person's image or video is manipulated to create a fake representation of them. Previous research has shown the effectiveness of GANs in detecting deepfakes [2]. The authors [3] identified that GAN-generated fake images exhibit stronger nonlocal self-similarity. They introduced on-local Attention based Fake Image Detection network (NAFID), a detection network featuring a Non-local feature extraction (NFE) module capturing these nonlocal features. The authors in [4] presented a novel hybrid approach for detecting counterfeit luxury handbags by combining external attention generative adversarial networks (EAGANs) with deep convolutional neural networks (DCNNs). The proposed method incorporates four key enhancements. These include the utilization of transformers in EAGAN for local feature representation, the incorporation of an external attention mechanism for emphasizing global features, the addition of a CNN auxiliary classifier for efficient learning of feature influence, and the introduction of a three-stage weighted loss during training. DL is essential for GANs as they require multiple layers of neural networks to generate high-quality synthetic data. GANs have been applied to many different tasks including image generation,

video generation, and text generation. In GAN architectures, DL methods like recurrent neural networks (RNN) and CNN are frequently employed [5]. Although DL offers numerous advantages, there exists certain obstacles in its implementation [6]. The requirement for vast quantities of high-quality data to train the models is one of the main obstacles. Furthermore, DL models need powerful hardware to operate at their best and can be computationally expensive to train.

Fusion learning is useful for GANs as it can improve the quality of generated data by combining multiple models. Fusion learning can improve the model's robustness and reduce the possibility of overfitting. Additionally, fusion learning can help generate diverse and high-quality synthetic data. Previous research has shown the effectiveness of fusion learning in improving the performance of GANs [7]. However, the effectiveness of GAN-based fusion deep neural models for fake face detection has not been thoroughly explored. Fusion models combine multiple GAN-based models to improve their accuracy and robustness. Some recent studies have reported promising results with GAN-based fusion models for face recognition and detection tasks, but their effectiveness for fake face detection has not been fully investigated.

Therefore, this study proposes an artificial intelligence(AI)-supported fake face detection system and aims to conduct a comparative analysis of GAN-based fusion deep neural models for fake face detection. The effectiveness of various GAN-based models and their combinations in identifying false faces will be compared in this study. The effectiveness of different fusion methods, such as averaging, stacking, and boosting, will also be evaluated. The results of this study can provide insights into the effectiveness of different GAN-based models and their potential applications in combating the spread of fake images. They can also inform the development of more advanced GAN-based models and fusion methods for fake face detection.

The use of fake identities in mission-critical applications such as defence and financial sectors poses a significant threat to security. While smart sensors and internet of things (IoT) provide solutions for handling this problem in domestic and commercial applications, the accuracy and fake detection quality requirements for sensitive applications are extremely high. In this regard, AI-based supported fake face detection using Deep GANs can offer a promising solution.

Our goal is to use fusion DL models in this work to be integrated with GAN to provide high accuracy and lower loss function for fake face detection in mission-critical applications. Our contributions are summarized as follows.

- 1) To increase accuracy and decrease the loss function in fake face detection, the proposed research project will combine three DL models (Model-1, Model-2 and Model-3) with GAN integration.
- 2) The proposed work will perform a comparative analysis of the three fusion models based on their performance evaluation.
- 3) The fusion models will be trained on a large dataset of real and fake face images to improve the accuracy of the GAN-based system.

The structure of this paper is as follows. Section 2 describes the literature survey for our work. Sections 3–6, present the system model and architecture, results, and discussion, respectively. Finally, Section 7 concludes this paper.

## 2. Literature survey

Detecting fake faces is a complex challenge within the field of computer vision. Many strategies have been developed to solve this problem. A popular approach involves the use of DL models, specifically, CNNs, to distinguish between real and fake face images. However, relying solely on GANs is not sufficient for accurate fake face detection because GANs can generate visually convincing yet flawed images. To overcome this limitation, researchers have suggested the integration of GAN with other DL models, like CNN, to enhance the precision of fake face detection.

For security purposes, Closed Circuit Television(CCTV) cameras are positioned and monitored in both public and private spaces. Video and image footage are frequently used in home and commercial security for identity detection, object recognition, and fast replies. To identify dangerous objects from CCTV data, this paper [8] presents the attuned object detection scheme (AODS). CNN is used in the recommended approach for object detection and classification. Categorization is performed based on the features of the object that CNN has extracted and analyzed. The hidden layer approaches are classified into different feature-constraint-based studies to locate the item. This process is used to categorize the data to identify hazardous goods.

Pedestrian detection is helpful in the context of smart transportation systems and video surveillance because it provides baseline data for population tracking, people counting, and event detection. DL model development has attracted a lot of attention lately in computer vision techniques such as object recognition and object categorization. The foundation of this application is supervised learning, which requires labels. Because of the CNN feature's higher representational capacity, pedestrian recognition has significantly improved due to CNN. Pedestrian detection often encounters substantial difficulties. On hard negative examples, such as poles, tree leaves, traffic signals, etc., it is typically difficult to reduce erroneous negatives. For clever multimodal pedestrian recognition and categorization, this research [9] develops an intelligent multimodal pedestrian detection and classification using hybrid metaheuristic optimization with deep learning (IPDC-HMODL). It goes through three stages: Multimodal object recognition, categorization of pedestrians, and parameter tuning.

This study [10] explores a novel application of a Vision Transformer (ViT) as a reference measure for assessing the quality of reconstructed pictures after neural image compression. The Vision Transformer is a creative application of the Transformer attention mechanism for object recognition in digital images. Usually, language models use this mechanism. Against a set of training labels, the ViT architecture seeks to produce a classification likelihood distribution. The metric is called a ViT-Score. This method is supported by other similar measurement techniques based on structural comparison or the mean squared error (MSE) of differences between individual pixels (structural similarity index, SSIM).

Extensive training has led to the proliferation of powerful visual identification models. Generative models, such as GANs, have, however, usually been trained unsupervisedly from the beginning. In their pretrained model embeddings, the authors [11] propose an effective selection mechanism by analyzing the linear separability between real and fake samples. After that, the most accurate model is chosen and gradually added to the discriminator fusion.

Using parametric building information modeling (BIM), the authors [12] examine the automatic generation of training data for artificial neural networks (ANNs) to identify building items in photographs. Teaching AI computers to recognize building objects in images is the first step toward using AI to assist in the semantic 3D reconstruction of existing buildings. The challenge of obtaining training data that is typically human-annotated persists if a computer program is unable to supply high-quality data for self-training for a particular task. In keeping with this, the authors only used 3D BIM model images that were generated automatically and parametrically with the BIMGenE tool to train ANNs. In a test case and random images of buildings outside the training dataset, the ANN training result demonstrated strong semantic segmentation and generalizability, which is critical for the future of training AI with generated data to address real-world architectural challenges.

Evaluation from preexisting databases and adequate training data for data-hungry machine-learning algorithms are critical for facial forgery detection. Unfortunately, despite the widespread use of face recognition in near-infrared settings, there is currently no publicly accessible face forgery database that incorporates near-infrared modality. In their paper [13], the authors outline their attempts to build a substantial dataset for facial forgery detection in the near-infrared modality and present a novel knowledge-based forgery detection method known as cross-modality knowledge distillation. This method guides the student model with a sparse amount of near-infrared (NIR) data by using a teacher model that has been pretrained on big data based on visible light (VIS). A range of perturbations are used to mimic real world conditions. The suggested dataset closes the gap in the NIR modalities face forgery detection study.

While GAN-generated face forensic detectors based on deep neural networks (DNNs) have achieved remarkable performance, they are vulnerable to adversarial attacks. This study [14] proposes an effective local perturbation generation method to expose the shortcomings of state of the art forensic detectors. The primary objective is to determine which concerns are shared by fake faces in the decision-making process of multiple detectors. To enhance the visual appeal and transferability of fake faces, GANs are then utilized to generate local anti-forensic perturbations in these regions.

Low-light face identification is challenging but necessary for real-world uses such as autonomous driving at night and urban surveillance. The existing face identification techniques, which mainly rely on annotations, lack flexibility and generality. The authors in [15] explored face detector learning in this work without using low-light annotations. The authors propose to fully utilize the available normal light data to convert face detectors from normal light to low light. To solve this problem, the authors propose a combined high-low adaptation (HLA) strategy. The authors have designed both multitask high-level adaptation and bidirectional low-level adaptation. To achieve low-level convergence, the authors enhance the dark photographs while decreasing the quality of the normal lighting photographs.

In the age of cloud, there is increasing concern about a facial recognition system's susceptibility to cyber attacks, partly due to the system's intimate relationship to user privacy. One potential attack is the model inversion attack (MIA), which aims to identify a targeted user by generating the best data point input to the system with the highest associated confidence score at the output. The data created

by a registered user may be maliciously exploited, which would be a serious invasion of privacy. This study [16] assumes the MIA under a semi-white box scenario, in which the system model structure and parameters are available but no user data information is, despite the DL-based face recognition system's complex structure and the variety of registered user data. It then confirms that the MIA poses a significant risk. The integration of deep generative models into MIA, or "deep MIA," encourages the alert state. It is suggested to use a face-based seed to initiate a GAN-integrated MIA.

The most widely used method for verifying someone's identity these days is facial recognition technology. Concerningly, the method's growing popularity has led to an increase in face presentation assaults, which use a picture or video of an authorized person's face to gain access to services. A more efficient and trustworthy face presentation attack detection method is offered by the authors in [17], which is based on a combination of background subtraction (BS), CNNs, and a fusion of classifiers. This method uses a fully connected (FC) classifier with a majority vote (MV) approach in conjunction with multiple face presentation attack tools. The proposed method incorporates a MV to determine whether the input video is authentic, which significantly enhances the face anti-spoofing (FAS) system's performance.

In recent years, there has been a significant rise in the frequency and sophistication of deepfakes, raising concerns among forensic experts, decision-makers, and the general public regarding their potential to spread false information. The DL methods that are best at identifying and averting deepfakes are built on top of CNNs. CNN is used to extract the underlying patterns from the input frames. These patterns are fed into a fully linked binary classification network, which classifies them as trustworthy or untrustworthy. According to the authors [18], the detection algorithm is not robust enough to detect relatively tiny artifacts created by the generation algorithms, which makes this method less than ideal in a scenario where the generation algorithms are constantly evolving. This study suggests a forensics method that involves humans in the detection loop and is hierarchically explicable.

The authors in [19] examine methods to improve the reliability of CNN models used in facial recognition. They explore additional measures of uncertainty within the models' judgements, which can be applied in mission-critical scenarios to identify less trustworthy judgments and classify them as uncertain. This approach enhances the model's overall accuracy in trusted judgments, albeit at the expense of increasing the number of judgments classified as certain. The study involves partitioning the cosmetics and occlusions dataset to simulate and improve the training and test set consistency, reflecting real-life scenarios. This dataset is used to assess six CNN models that are currently in use. Simple A/B testing and meta-learning supervisor ANN solutions are used for experimentation to find error patterns within the underlying CNNs.

There is a provision that can switch faces due to deepfake, a cutting-edge DL system. Hyper-realistic photos and movies may be produced using new developments in neural network techniques, such as generative adversarial neural networks. These fake images and videos on various social media platforms pose a severe threat. To solve these issues, deepfakes can be categorized using a range of DL algorithms. Three neural network models—ResNext, Xception, and Fusion have been compared by the authors in this study [20,21]. According to the study, the Fusion technique performs better than others.

AI techniques such as GAN have made it possible to create convincing fake face photos. These photos are then used to create fraudulent accounts on different social media platforms. The authors of this paper [22] distinguish between camera-captured real face photos and GAN-generated false face photos using binary classification models based on DL. The classification models are developed by fine-tuning three state of the art, lightweight, pretrained CNNs: GoogleNet, ResNet-18 and MobileNet-v2. The process involves applying the transfer learning methodology. In this technique, the opponent color-local binary pattern (OC-LBP) pictures created using joint colour texture feature maps are employed as input to the CNN rather than Red Green Blue (RGB) images. The suggested approach performs remarkably well in terms of test precision, generalization ability, and Joint Photographic Experts Group (JPEG) compression resilience.

The authors in [23] proposed a novel approach for generating point clouds based on a deep adversarial local features-based generative model. Their proposed GAN reduces the computational load and improves the accuracy of three-dimensional (3D) acquisition, reconstruction, and rendering operations. To begin with, the authors employ an autoencoder to optimize the latent space and extract local features before training the GAN. Through the training process, a robust point cloud and an accurate estimation of the local context are achieved. This work's significant contribution lies in its revolutionary DL-based method for creating 3D point clouds, which greatly alleviates the computational demands associated with rendering augmented reality (AR) and mixed reality (MR) materials.

The development of prediction models is at the heart of consumer purchasing behavior analysis. In their study, Zhang et al. [24] introduced KmDNN and rDNN models. They use Python as an experimental tool to create algorithms, produce prediction results, and carry out comparative analysis. They also use the area under the region of convergence curve (AUC) value as an evaluation indicator. The study, which focuses on DNN models in DL, also suggests the random sampled deep neural networks (rDNN) and k-means deep neural network (KmDNN) models. From three angles—the underlying theory, model development and application, and model enhancement, it examines how well these models perform when analyzing consumer purchase behavior. The empirical analysis is supported by experimental results.

The method proposed in [25] tackles challenges in face anti-spoofing, a critical aspect of robust face recognition systems. Despite advancements in spatial and temporal aspects, the authors note persistent difficulties in utilizing facial depth as auxiliary information and developing lightweight temporal networks. To address these issues, they introduce a novel two-stream spatial-temporal network. This innovative design incorporates a temporal shift module for efficient temporal feature extraction and introduces a symmetry loss in the depth information estimation network to enhance auxiliary supervision accuracy. A scale-level attention module is then formulated based on estimated depth information, amalgamating data from the two-stream network to extract crucial discriminative features. The method concludes with a fully connected network for authenticating images. This approach presents a novel solution for exploring facial depth characteristics and improving discriminative feature extraction, contributing to the progression of anti-spoofing methodologies.

Addressing the persistent challenge of uncompleted face recognition, characterized by issues such as occlusion, a limited sample size, diverse expressions, and poses, the authors in [26] present a pragmatic solution. Their devised hybrid-supervision learning framework seamlessly integrates the

strengths of both supervised and unsupervised features. Within the supervised branch, they introduce hybrid multiple marginal fisher analysis (HMMFA), an effective feature learning method. Concurrently, in the unsupervised branch, enhancements are made to principle component analysis network (PCANet) to extract more robust local information. The fusion stage strategically extracts discriminant features from the hybrid features, utilizing support vector machine (SVM) as the final classifier. Notably, the proposed method, requiring no auxiliary set and featuring fewer parameters than DL methods, proves to be storage-efficient, presenting an economical and practical solution for smaller communities.

In response to the authenticity concerns arising from the massive growth of data, particularly on social media, the authors in [27] delve into digital forensics, focusing on the pervasive issue of face manipulation. The paper presents an innovative method to detect forged faces in light of the constantly evolving nature of image forgery tools and their challenge to current detection systems. The suggested method leverages image quality assessment (IQA) features, extracting data from both the frequency and spatial domains, in contrast to existing DL-based solutions. The generalization and explainability of the suggested model are also covered in the paper, providing a viable solution to the problems associated with fake face detection.

The study proposed by [28] presents an advanced method for automated fake news detection by introducing a deep normalized attention-based mechanism. This mechanism uses the bidirectional gated recurrent unit (BiGRU) to improve feature extraction, in contrast to conventional methods that only consider text features. To improve detection accuracy, the model also incorporates an adaptive genetic weight update-random forest (AGWu-RF) for classification, where hyperparameter values are optimized. The novel aspect of the mechanism is its capacity to take into account dual emotion features, utilizing both publisher and social emotions.

The authors in [29] present a novel model named graph contrastive learning with feature augmentation (FAGCL) to address the problem of rumors on social media. In contrast to recent DL methods that struggle with noisy and inadequate labeled data, FAGCL uses graph contrastive learning to improve rumor detection. The model uses user preference and news embedding for the rumor propagation tree, injects noise into the feature space and learns contrastively with asymmetric structures. Node representations are updated by a graph attention network, and various pooling techniques combine these to provide graph-level classification. Graph contrastive learning is used by FAGCL as an auxiliary task to improve representation consistency and build a stronger model.

As a result of technological advancements, there has been a noticeable increase in the daily use of bank credit cards. Consequently, the fraudulent use of credit cards by uninvited parties is one of the crimes that is quickly growing. As a result, there is currently active research being done to identify and stop these kinds of attacks. The authors [30] explore the difficulties in detecting fraudulent transactions in banking and suggests solutions based on DL methods. Transactions are assessed and compared against preestablished models to identify fraud. Based on aggregated data from a generative adversarial network, the study shows that the combined model of short-term memory and deep convolutional networks performs best. The paper's approach proves to be more effective in addressing the issue of imbalanced class compared to conventional techniques.

The application of generative networks based on DL has brought about a substantial change in the field of creating synthetic biometric data. Applications for synthetic biometric data can be found in many areas, such as training medical professionals through simulations and developing and testing



biometric systems under high load. Due to a lack of training samples, DL models frequently perform poorly when trained on biometric datasets. The authors [31] described DSB-GAN, a generative model that creates realistic synthetic biometrics for multiple modalities, including fingerprints, iris, and palmprints. It is based on a convolutional autoencoder (CAE) and GAN. Access to information that might not be easily available because of things like data corruption and distortion is made possible by the generated data.

The newest area of study in natural language processing is sentiment analysis (SA) of user-generated reviews. Due to its significance in understanding consumers' experiences with services or products, the extraction of consumer sentiment from the content of consumer evaluations is receiving a lot of attention nowadays. This paper suggested the BERT-GAN model with review aspect fusion for the consumer SA, which enhances the performance of the BERT model's fine-tuning by incorporating semi-supervised adversarial learning. The word sequences and numerous service-related elements were combined by the authors before being fed into the model. This facilitates the incorporation of aspect representation and positional information into the sentences' context.

In the emerging field of image style transfer (IST), combining CNN-based and GAN-based techniques has attracted a lot of interest. The authors of [32] present a novel IST technique that makes use of an augmented GAN framework with a previous circular local binary pattern (LBP), thereby addressing a common shortcoming seen in many existing approaches—namely, a reduction in the definition of image textures. The circular LBP is positioned inside the GAN generator and functions as a texture prior, thereby improving the fine-grained textures of the style images that are produced. The authors aim to improve the extraction of high-frequency features by incorporating an attention mechanism and a dense connection residual block into the generator in response to the need for more detailed information.

The authors of [33] highlight a noteworthy use of GANs in the context of vehicle re-identification (Re-ID) for intelligent transport systems, addressing the inherent difficulties of cross-domain vehicle matching. The GAN-Siamese network structure that the authors present was created especially for vehicle Re-ID. They highlight how well it works in both daytime and nighttime scenarios. A crucial part of their process is the integration of a GAN-based domain transformer, which skillfully modifies the domains of input car pictures. After this transformation, a four-branch Siamese network that is specifically designed to learn distance metrics between images within each domain is used. Most importantly, the integration of these learned metrics is made easier by the GAN-based domain transformer, which produces an all-encompassing similarity measurement for vehicle Re-ID.

Demand response modeling plays a crucial role in smart grids as it enables the assessment and modification of customer load profiles. This strategy enhances the system's effectiveness and improves energy management efficiency. In the residential sector, the scheduling method for user profiles is enhanced through demand response analysis during load profile calculations. Machine learning techniques are employed to analyze profile patterns and model user behavior. An incentive-based demand response approach is used to emphasize how important it is to protect residential users' privacy during interactions involving demand and load profile patterns. Real-time incentive-based demand response modeling for residential users is examined in the work by the authors [34]. The study focuses on setting up hubs in a cloud-fog smart grid environment so that demand response modeling can be combined

with home appliance scheduling. A cloud computing environment and a GAN quality (Q)learning model are used to address privacy concerns in the demand response model during strategy analysis.

Using GAN, the authors [35] present a DL technique for producing 8K ultra high definition in high dynamic range(UHD HDR) content from full high definition and 4K video. By employing a random downsampling method, a multilevel dense and residual structure and multiple levels of structure, this technique generates 8K UHD HDR footage that exhibits consistent color performance and is both aesthetically pleasing and natural.

### 3. System model and architecture

This section describes the dataset, the neural architecture of the fusion model and GAN along with the description of the GAN architecture in detail.

#### 3.1. Dataset

The dataset acquired from Kaggle source [36] with 70k real and 70k fake faces (generated by style GAN) for this work. More than a million faces are available in the dataset. The style GAN is an extension of the GAN, where the mapping points of the latent space are mapped to an intermediate latent space. Using this intermediate space to control the style of the GAN image, this results in the reproduction of high-quality images with varying degrees of detail by changing the noise and style quality. When the Fusion models are applied it increases the performance and reduces loss as well.

The second dataset is used also from the Kaggle source [37]. This dataset contains 2064 real and fake images, which is split based on 80% for training, 10% for validation and 10% for testing. The fusion models are large, and are capable of handling large datasets, but performed reasonably well with smaller datasets. This ensures the reliability of the models. More discussion is presented in the upcoming sections.

#### 3.2. Fusion DL model

A fusion DL model is an advanced model that combines several neural network architectures to improve a machine learning system's capabilities and performance. To process and analyze data from various sources or modalities, it essentially combines the strengths of various DL techniques, including recurrent neural networks (RNNs), CNNs, and more. The model can capture a wider range of features, patterns, and relationships within the data due to this fusion of disparate data and modeling strategies. Fusion DL is especially useful for tasks like audio-visual recognition and image and text analysis that require a thorough comprehension of complex, multimodal data. Through the combination of these various neural networks, the fusion DL model greatly enhances the system's capacity to make accurate predictions, classify data, and solve intricate problems, making it a valuable asset in various domains, including computer vision, natural language processing, and audio analysis.

Accuracy is a measure of the overall correctness of a classification model. Precision quantifies the accuracy of positive predictions made by the model. Recall, also known as sensitivity or true positive rate, measures the ability of a model to capture all the relevant instances. F1-Score is the harmonic mean of precision and recall. It provides a balanced measure of a model's performance.

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Predictions}} \quad (3.1)$$

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \quad (3.2)$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad (3.3)$$

$$\text{F1-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3.4)$$

### 3.3. Fusion DL architecture

Three models are fused to develop the hidden layer of the fusion model. There are three such fusion models designed with Densenet-201, ResNet-102 and Mobile Net V2 and the second fusion model is built with DenseNet-201, ResNet-50 and InceptionV3. The third fusion model is built with DenseNet-201, ResNet-50 and Xception. The ensemble output of these three models is applied as the input to the model-4, which activates the output layer with the target function, for the classification of the real and fake images. Three models were used as the convolution layers for the pooling (global 2D average pooling), batch processing and model training. The ensemble efforts of the three models as model-4 performs the activation of the output layer and image classification at the output. Thus the fusion models are interconnected with dropout and dense layers for connecting the other models with appropriate weight transfers at the output.

The fake face detection is a composite problem. It contains several parts such as sampling, pre-processing pooling, normalization, vectorization, batch processing, training the model. testing, and classification with the output activation. The various parts of the deep fake detection for the generic set of operations for the complete process are expressed as per the Algorithm 1. This algorithm illustrates the sequence of operations involved in the detection of the fake face.

The first step in the fake face detection is the convolution. The proposed system uses 2D convolution with average pooling with image dimensions of 128\*128\*3 pixels. The 2D convolution process with the extraction of the feature map can be mathematically expressed as per the Eq (3.5). This process converts the image into a feature map of size 128\*128\*3 through the kernel interface.

$$G[x, y] = i * o[x, y] = \sum_j \sum_k h[j, k] f[x - j, y - k] \quad (3.5)$$

The  $x$  and  $y$  are the rows and columns of the output image array,  $i$  is the input image and the  $o$  is the kernel layer. The next step is the feature selection of the image array. Here, we apply global average 2D pooling, which is applied for the selection of the image features through a feature map. This method is a replacement technique on a fully connected layer. This pooling allows the feature map for one classification process of the transfer layers. Instead of adding a fully connected layer over the feature map, this allows us to estimate of the average of each feature map, then connect the different classification models to the fully connected layer. This is mathematically expressed in Eq (3.6).

$$e = [(1/\bar{\omega} \sum_{u \subset \omega} x_{cu}^p)]^{1/2} \quad (3.6)$$

**Algorithm 1.** Algorithm for the overall fake-face classification problem.

**Data:**  $i = [\sum_{n=0}^s X_n]$

**Result:**  $j = wi + b$

$y \leftarrow y_{tr}, y_{tst};$

$x \leftarrow x_{tr}, x_{tst};$

$n \leftarrow imagesamples;$

$TRU_P \leftarrow TruePositive;$

$TRU_N \leftarrow TrueNegative;$

$FAL_P \leftarrow FalsePositive;$

$FAL_N \leftarrow FalseNegative;$

$Features \leftarrow x, y;$

**Pooling**  $max[x][y] = max(input[Z][K]);$

**Normalization**  $x^* = x_{tr} - E[x_{tr}] / \sqrt{var^2(x_{tr})};$

**Activation**  $f(y) = max(0, a);$

**Activation**  $tanh(a) = (e^a - e^{-a}) / (e^a + e^{-a});$

**while**  $x_{test} \neq 0$  **do**

**if**  $x_{test}$  is fake **then**

$acc \leftarrow TRU_P + TRU_N / (TRU_P + TRU_N + FAL_P + FAL_N);$

$precision \leftarrow TRU_P / (TRU_P + FAL_P);$

$recall \leftarrow TRU_P + TRU_N / (TRU_P + FAL_N);$

$f1 - score \leftarrow 2 * (pre * re - call) / (pre + re - call);$

$loss \leftarrow y(w^T, r) * 1/m \sum_1^m (b^{[\bar{y}]}, b^{(y)});$

**▷ \*Fake Face Detected**

**else**

**if**  $x_{test}$  is no-fake **then**

$acc \leftarrow T_P + T_N / (T_P + T_N + F_P + F_N);$

$precision \leftarrow T_P / (T_P + F_P);$

$recall \leftarrow T_P + T_N / (T_P + F_N);$

$f1 - score \leftarrow 2 * (precision * recall) / (precision + recall);$

$loss \leftarrow y(w^T, r) * 1/m \sum_1^m (b^{[\bar{y}]}, b^{(y)});$

**▷ \*Original Face Detected**

**end**

**end**

**end**

where  $p$  is a parameter that is generally greater than zero. Keeping this value greater than one, increases the contrast of the feature pool map and focuses only on the important features of the image. For average pooling, the value of  $p$  is one and for the max pooling, it is infinity. There are three fusion models used which is a combination of basic Densenet-201- ResNet 102/50 with a third layer of Inception V3, Xception or Mobilenet V2. These three transfer layers are ensembles at layer four before the output layer is connected to them. The batch normalization (BN) is applied in the proposed system. It contains four dimensions such as batch  $b$ , channel  $c$  and dimension vectors  $x$  and  $y$  for both input  $I$  and output  $O$ . The  $\mu$  is the mean value of activation and the  $\sigma$  is the standard deviation.  $\beta_c$  presents all the activation in the channel  $c$ . The BN can be mathematically expressed in Eq (3.7).

$$BN = \frac{I_{b,c,x,y} - \mu_c}{\sigma_c + c} + \beta_c \quad (3.7)$$

The next step is the activation of the output layer. The proposed system uses rectified linear activation function (ReLU), which activates the output layer between the values zero and one. This can be mathematically expressed in Eq (3.8).

$$RELU(y) = Max[0, y] \quad (3.8)$$

The next step is the application of the soft-max layer. This layer is required if the target is a multi-class classification problem. The soft-max layer is mathematically expressed as per Eq (3.9). The  $y_i$  and  $y_j$  are the activation of the previous and next stages.

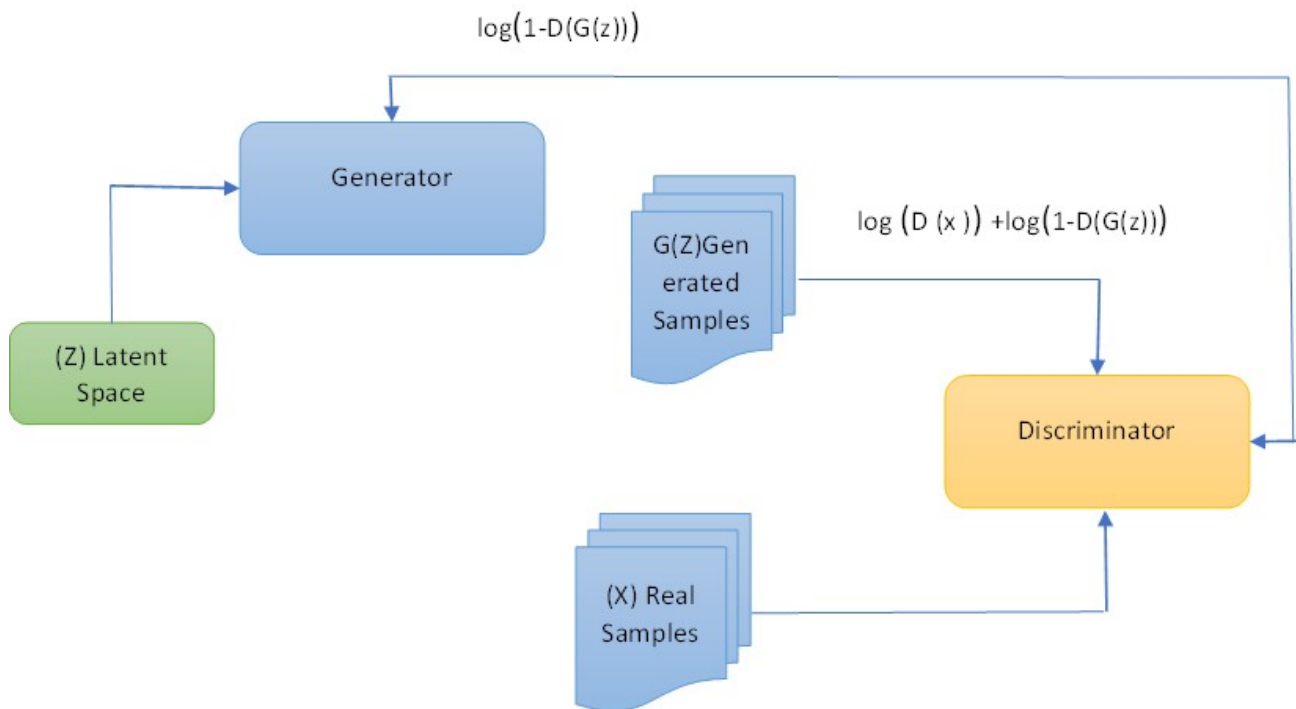
$$Softmax(y_j) = \frac{e^{y_i}}{\sum_j e^{y_j}} \quad (3.9)$$

Finally, the fully connected layer is established between input  $I$  and output  $O$  with bias  $b$  and weight  $w$  as per the Eq (3.10).

$$O = Iw + b \quad (3.10)$$

### 3.4. Deep GAN

Deep GAN in DL are emerging techniques nowadays and footprints in image segmentation that can produce photo-stylized images that are nearly identical to real images in terms of quality and realism. The deep GAN technique is devised for addressing the issues of threats in data scarcity, in which it is feasible to create a trained model using few objects. Microsoft common objects in context (MS COCO) and Pascal visual object classes (VOC) datasets have wide images of small objects and real interpretation, which reduces the dataset's concern to the researchers. However, generating high-quality images will consume much time and human labor, and it is rigorous and expensive. The location of minuscule objects in visual information is consequentially limited. Small objects can also be bendable or merged with larger objects. DL model performance is constrained by the small dataset and class imbalance issues caused by the lack of dataset. The aforementioned issues can be overcome by the advancement of DL techniques. There are two approaches to increasing the dataset. One is utilizing images from real datasets. Image data enhancement techniques have been developed to expand the dataset quantity at a cheaper cost. Basic image modification, which includes easy procedures like cropping or flipping photographs is one method of enhancing visual data. This method can increase



**Figure 1.** GAN deep neural network architecture.

the size of the dataset at a cheap computational cost, but if the dataset is too small, it may lead to overfitting issues. The second approach is the deep GAN(DGAN) technique for image augmentation in datasets. DGAN architecture has two networks namely, the generator and discriminator. The deep generator network produces several fake images of objects, and the deep discriminator network is to separate the actual objects from the artificial objects produced by the generator. The real and accurate data are produced by the generator and used to feed the discriminator as negative instances.

In Figure 1, the GAN generator and discriminator are explained. The generator output is described as per Eq (3.11).

$$GEN = \log(D(x)) + \log(1 - D(G(z)))_2 \quad (3.11)$$

where  $G(z)$  is the generated data and the  $D(X)$  is the discriminator output. The differential output of the discriminator and the training data is fed back to the generator through back-propagation with the value as per the Eq (3.12).

$$DIFFERENCE = \log(1 - D(G(z))) \quad (3.12)$$

The generator consists of eight closely packed blocks and a convolutional block. Convolutional layers, batch normalization, and ReLU activation are included in each block of the GAN network. The network's layers are fed with data from the preceding layers through back-propagation, except the last layer. This technique successfully lessens the vanishing gradient issue. The dense blocks remove the elevated features, whereas the convolutional layer removes the low-level information. The generating network can learn the remaining gap between the pixel level and the original data. The output images are produced by the last three-by-three convolutional layer. The final denoised image is visually pleasing because the discriminator network can distinguish between the false and the real image. Without

**Table 1.** Determination of GAN loss functions.

Epoch	Generator-loss	Discriminator-loss
1	0.531	1.529
2	0.351	1.723
3	0.622	1.39
4	0.393	1.34
5	1.325	1.492

**Table 2.** Description of various fusion models.

Model	Description	Optimizer	Norm
Model-1	Mobilenet201+Resnet 102+Mobilenet V2	Adam	Batch
Model-2	Mobilenet201+Resnet 50+ Inception V3	Adam	Batch
Model-3	Mobilenet201+Resnet 102+Xception	Adam	Batch

the need for human data gathering and labeling to enhance the dataset's size, GAN may expand an object detection dataset quickly and affordably.

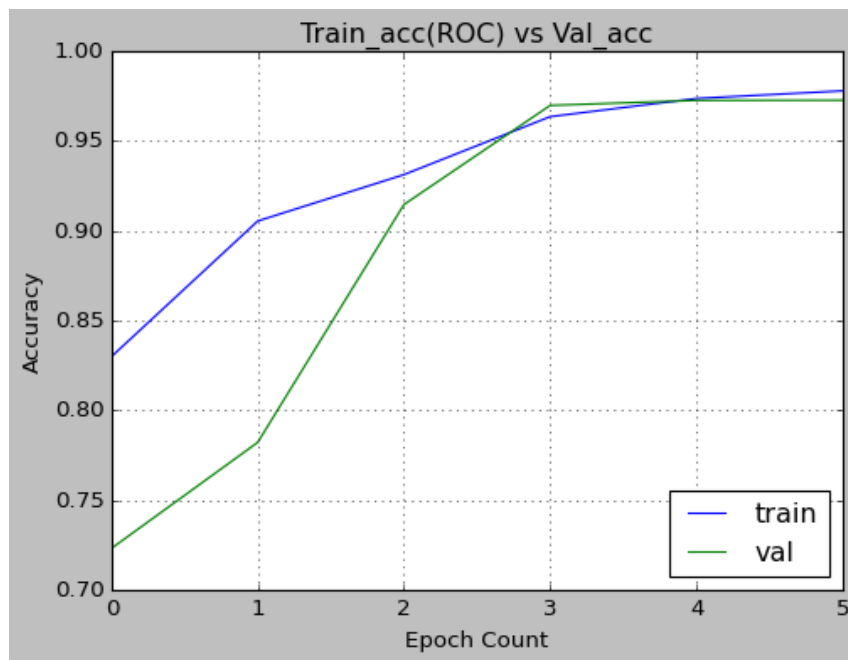
#### 4. Results

A simple GAN-based generator-discriminator pair is employed on the dataset to estimate the loss values of the generator and the discriminator. This loss value is estimated based on the generation of fake images and the comparison with the original training dataset. The results of the loss estimation are specified in the Table 1. Since the loss function of the generator and discriminator is huge for small-scale training, without the support of the external network fast fake face detection will be extremely difficult. The loss function also reduces the reliability, quality and stability of the reconstructed images. Hence, the proposed work integrates the efforts of GAN with fusion deep image neural network models to provide more stability, faster performance, high reliability, better reconstruction of the image, and global application on various datasets with similar or better accuracy. Thus, the proposed work is an integration of GAN with fusion DNNs. This work compares the three fusion models and integrates the one with GAN for the final framework. The model is selected based on the performance of loss and accuracy during training and validation.

The GAN fake-face detection is tested with three fusion models. The details of the models are presented in Table 2. The various models, their parameter evaluation performance evaluations and comparative analysis are described in detail in this result section.

The proposed work uses fusion DL models to handle the GAN generation and discrimination part. The proposed work contains a comparative analysis of the three fusion models. The first fusion model contains three models made out of three transfer functions such as Densenet201, Resnet-102 and Mobile net V2. These three models converged and combined by the fourth model which produces the output fully connected layer.

The combined layer uses MaxPooling2D to enhance the performance of the prediction. The output layer is activated by the ReLu function and since it needs to store and process the image, we used a dropout value of 0.5, and the output layer is activated by soft-max to handle the multi-class problem.



**Figure 2.** Performance convergence of the fusion Model-1.

**Table 3.** Fusion Model-1 performance description.

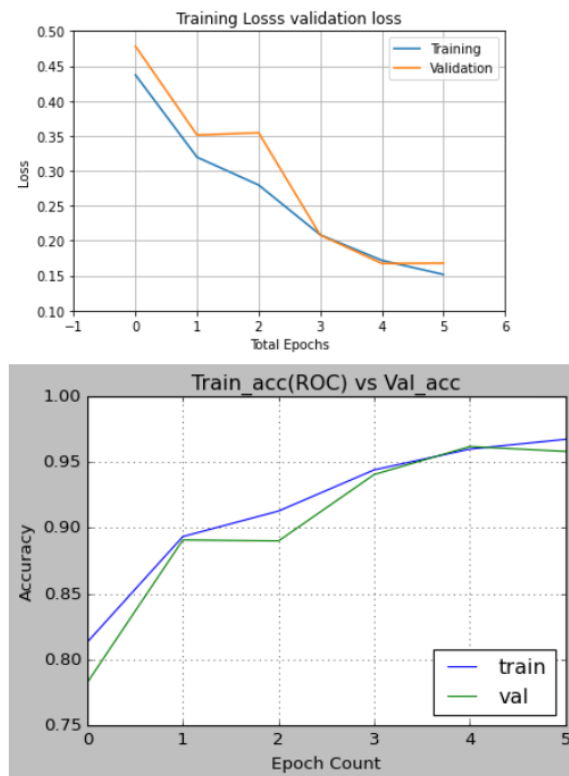
loss	Accuracy	Validation-loss	Validation-accuracy
0.4314	0.8189	0.5118	0.8021
0.3418	0.8836	0.4996	0.8206
0.2968	0.9146	0.4238	0.8593
0.2243	0.9591	0.2286	0.9609
0.2113	0.9651	0.2151	0.9577

The transfer models use ImageNet weights with the size of  $128 \times 128 \times 3$ . The batch size of the training data is 64 and the validation data is 32. The three models provide the feature mapping, which is done with GlobalAveragePooling 2D to improve the accuracy of the model. The performance of the model is good and produces an accuracy of about 0.9547 and a loss value of 0.2197. The model performance in terms of batch processing is listed below Table 3. The performance of the Fusion model based on training accuracy, validation accuracy, training loss and validation loss are expressed as per Figure 2. The response shows a gradual increase in accuracy and a decrease in the loss functions. Thus, this model provides stable performance during the time of the convergence.

The second fusion model uses Densenet201, Resnet-50 and Inception V3 DL model for transfer weights in the hidden layer. These three models converged and combined by the fourth model which produces the output fully connected layer. The BN is used for the normalization process. The combined layer uses MaxPooling2D to enhance the efficiency. The output layer is activated by the ReLu function. The dropout value is similar to 0.5 and the output layer is activated by soft-max to handle the multi-class problem.

The performance of the training accuracy, validation accuracy, training loss and validation loss for the Model-2 is expressed as per Figure 3. The response shows a gradual increase in accuracy and a





**Figure 3.** Performance convergence of the fusion Model-2.

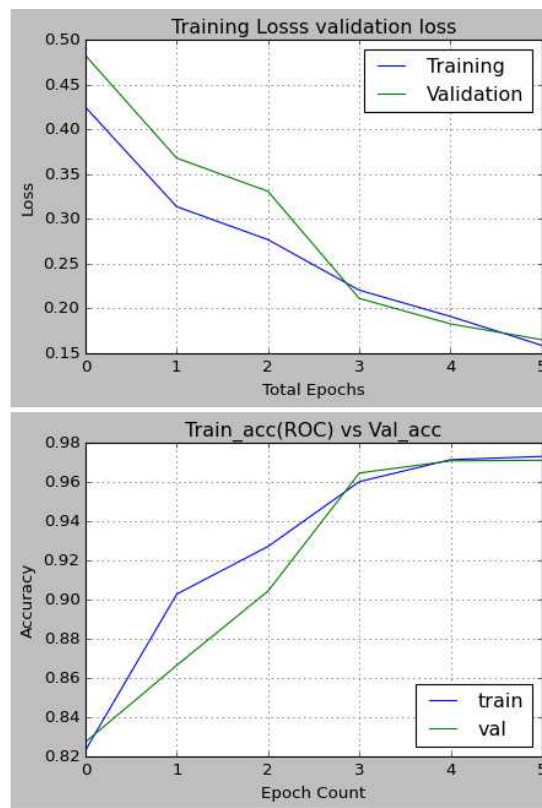
**Table 4.** Fusion Model-2 performance description.

loss	Accuracy	Validation-loss	Validation-accuracy
0.4372	0.8131	0.4783	0.7824
0.3194	0.8934	0.3512	0.8908
0.2796	0.9128	0.3544	0.8901
0.2084	0.9441	0.2081	0.9406
0.1718	0.9599	0.1674	0.9618
0.1517	0.9674	0.1678	0.9581

decrease in the loss functions. This model suffers small spikes in the middle due to a negligible amount of the positives, in validation parameters, but this model provides the lowest loss function hence the sensitivity of the model is higher.

The transfer models use ImageNet weights with the size of  $128 \times 128 \times 3$ . The three models provide the feature mapping with GlobalAveragePooling 2D to improve the accuracy of the model. The performance of the model is good and the model produces an accuracy of about 0.9578 and a loss value of 0.1146. The model performance in terms of batch processing is listed in Table 4.

The third fusion model uses Densenet201, Resnet-102 and Xception DL models for transfer weights in the hidden layer. The BN is used for the normalization process. The combined layer uses Max-Polling2D to enhance the efficiency. The output layer activation is done by the ReLu function. The dropout required for storing the preprocessed images and output layer is activated by soft-max to handle the multi-class problem. The transfer models use ImageNet weights with the size of  $128 \times 128 \times 3$ .



**Figure 4.** Performance convergence of the ensemble Model-3.

**Table 5.** Ensemble Model-3 performance description.

loss	Accuracy	Validation-loss	Validation-accuracy
0.4247	0.8231	0.4825	0.8273
0.3137	0.9028	0.3678	0.8665
0.2768	0.9271	0.3307	0.9043
0.2204	0.9602	0.2112	0.9645
0.1911	0.9714	0.1828	0.9708
0.1588	0.9730	0.1652	0.9711

The three models provide the feature mapping with GlobalAveragePooling 2D to improve the accuracy of the model. The performance of the model is good and produces an accuracy of about 0.9797 and a loss value of 0.1490. The model performance in terms of batch processing is listed in Table 5. The performance evaluation for Model-3 is described in Figure 4. This model shows no deviation in accuracy and receiver operating characteristics (ROC) performance. hence, the best-performing model of the three ensemble models is evident from the analysis.

The fusion model was experimented with for the second dataset. The GAN generation and discrimination were done with an 80-10-10% ratio for the training, validation and testing. The hyper-parameters for the training and validation had been configured with similar values such as the image size (128\*128\*3), learning rate of 0.001, batch size of 64 for two thousand and sixty four image samples, thousand six hundred and thirty two for training and two thousand and sixty four each for the

**Table 6.** Fusion Model-3 performance description for dataset-2.

loss	Accuracy	Validation-loss	Validation-accuracy
0.7096	0.6025	1.5485	0.4583
0.6081	0.6569	0.7703	0.5573
0.5372	0.7387	0.7641	0.4896
0.5029	0.7613	0.7088	0.6406
0.4582	0.8044	0.5976	0.6354
0.4054	0.8462	0.5655	0.6979

**Table 7.** Fusion Model-2 performance description for dataset-2.

loss	Accuracy	Validation-loss	Validation-accuracy
0.7078	0.5838	0.9179	0.4948
0.6175	0.6575	1.2108	0.5990
0.5820	0.7000	0.7193	0.5052
0.5576	0.7156	0.6327	0.6823
0.5123	0.7581	0.7538	0.5312
0.4975	0.7619	0.9799	0.5208

**Table 8.** Fusion Model-1 performance description for dataset-2.

loss	Accuracy	Validation-loss	Validation-accuracy
0.7178	0.5987	0.7762	0.5521
0.5895	0.7038	0.6770	0.5156
0.5403	0.7269	0.7619	0.5260
0.5234	0.7500	0.8734	0.5729
0.5101	0.7606	0.5990	0.5573
0.4412	0.8219	0.6720	0.6406

testing and validation. Model 3 performed better than Model 1 and Model 2 with the following performance measurements as per Table 6. The model, which was designed to handle the large dataset of about a million images, also works decently well with small datasets. The Model 1 and Model 2 performance metrics are tabulated in Tables 7 and 8. The fusion model 2 attained a maximum accuracy level of 76.19%. The fusion model -1 with ResNet-101 and MobileNet-V2 rendered 82.19% accuracy. The experimentation with the second dataset provides a test case for the reasonable performance of the proposed model, despite the application on a small dataset with only two thousand and sixty four image samples.

## 5. Discussion

A comparison of the three fusion models for GAN-based fake face detection and processing shows that the performance and accuracy are high for the third model that used Xception as one of the hidden layer transformations. This provided better accuracy compared with the ones that use Mobile Net V2 and Inception V3. The loss is minimal with the use of Inception V3, which provides more sensitivity

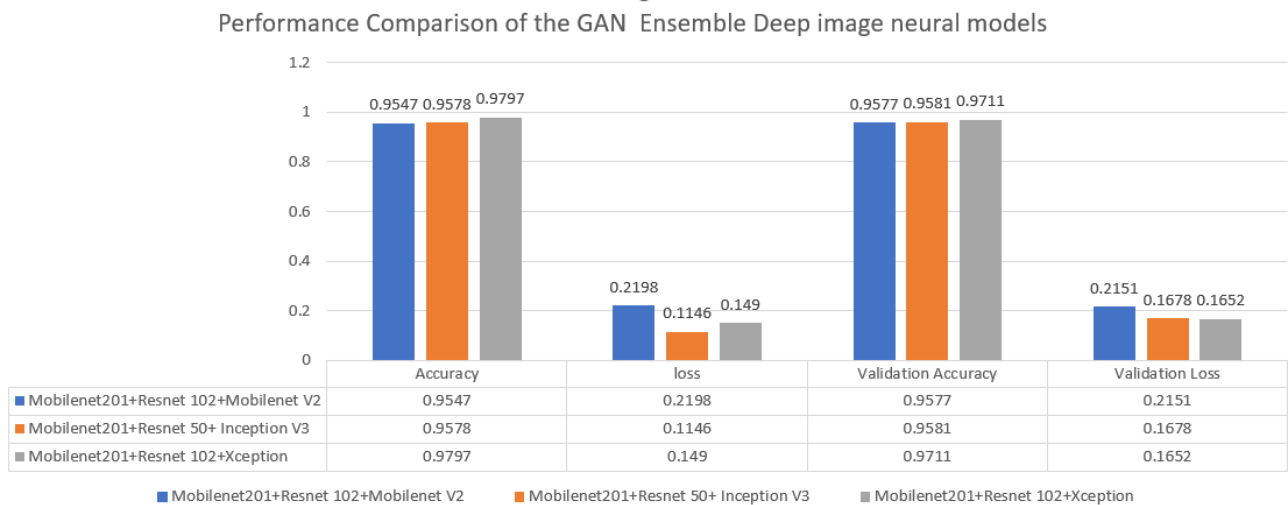
measures among the three.

Hence, Xception-based fusion model provides better accuracy among the three fusion models. The other two transfer functions are Densenet 201, ResNet 50 and ResNet 102, which are deeper convolution models to increase the accuracy of the input feature engineering and transfer mapping in the hidden layer. The comparative analysis of the three fusion models is shown in Figure 5 and Table 9. From the analysis, even though Model-2 provided the lowest loss and increased sensitivity, Model-3 provides stable performance with no false positives and offers the highest accuracy and specificity. The accuracy of the fusion Model-3 is 97.97% with a loss of 0.1490. This model also has a better validation accuracy of 97.11% with a validation loss of 0.16521. This is more than Model-1 and Model-2, where the accuracy is 95.47 and 95.78%, respectively. The loss performance in Model -1 and Model-2 are 0.2198 and 0.1146, respectively. The validation accuracy of Model-1 and Model-2 are also lesser, compared with Model-3 such as 95.77 and 95.81%. Hence, the Fusion Model-3 is the most suited out of the three designs, which are used in the proposed work for the classification of the fake face images. For lossless prediction, Model 2 is preferred, which can be trained to predict the fake images generated by GAN systems. For better accuracy, GAN integration can be used with Model 3.

GANs, lauded for their prowess in generating high-quality data, are not without their drawbacks. A notable challenge lies in the intricate and delicate process of training GANs. The inherent competition between the generator and discriminator networks can render the training phase extremely difficult, introducing instability and slowing down the convergence of the model. Furthermore, GANs demand a substantial volume of training data to yield satisfactory results. This dependency becomes a substantial hurdle when datasets are limited in size or accessibility, potentially hindering the applicability of GANs in scenarios where ample data is not readily available. Another significant concern is the susceptibility of GANs to mode collapse, wherein the generator tends to produce a restricted set of outputs, failing to manifest the desired diversity in generated data. This limitation raises questions about the reliability and robustness of GANs in achieving varied and representative outputs, emphasizing the need for careful consideration and mitigation strategies in their implementation.

The key findings of the proposed work are summarized as given below.

- 1) **Integration of Fusion and Ensemble Models:** The integration of a GAN with three fusion DL models demonstrates a significant benefit in fake face detection for mission-critical applications. Fusion and ensemble models, particularly Model-3, contribute to heightened accuracy (0.9797) and enhanced reliability, showcasing their efficacy in addressing the complexities of diverse face datasets.
- 2) **ResNet and DenseNet Integration:** The incorporation of ResNet and DenseNet architectures within the layers of the model proves advantageous, leading to improved accuracy and robust feature extraction. ResNet and DenseNet integration, as seen in Model-3, plays a pivotal role in achieving superior performance, showcasing the importance of diverse model architectures for heightened accuracy in fake face detection.
- 3) **Improved Accuracy for Complex Images:** The proposed system demonstrates a notable improvement in accuracy, particularly for large and complex datasets, emphasizing its adaptability to real-world scenarios. The integration of GAN, coupled with fusion models and the incorporation of ResNet and DenseNet, collectively contributes to enhanced accuracy in detecting fake faces within intricate and challenging image contexts.



**Figure 5.** Comparative analysis of GAN-based fusion models.

Table 10 shows the performance comparison of the various fusion models for the second dataset with two thousand and sixty four images. The performance of the Xception-based third fusion model provides the maximum accuracy of 84.62% for a smaller dataset with only two thousand and sixty four images. Model-1 and model-2 provide an accuracy of 82.19 and 76.19%, respectively. The loss is minimum in the fusion model-3 with values of 0.4054. The minimum validation loss is measured in Model-3 with 0.5655 and with a maximum validation accuracy of 69.79%. Hence, Model-3 is preferred for better performance compared with model-1 for the second dataset experimentation.

**Table 9.** Performance comparison of the GAN fusion models.

Model	Accuracy	Loss	Validation-Acc	Validation-loss
Model-1	0.9547	0.2198	0.9577	0.2151
Model-2	0.9578	0.1146	0.9581	0.1678
Model-3	0.9797	0.1490	0.9711	0.16521

**Table 10.** Performance comparison of the GAN fusion models for dataset-2.

Model	Loss	Accuracy	Validation-loss	<b>Validation-accuracy</b>
Model-1	0.4412	0.8219	0.6720	0.6406
Model-2	0.4975	0.7619	0.9799	0.5208
Model-3	0.4054	0.8462	0.5655	0.6979

## 6. Conclusions

Face recognition is a powerful biometric parameter that can serve in various applications, starting from research and e-governance to domestic home security applications. The proposed method used three fusion models integrated with GAN that have been trained on a sizable dataset of authentic and fraudulent face photos. Fake face detection is a complicated process that involves several steps, such

as batch processing, normalization, vectorization, sampling, preprocessing, pooling, training, testing, and output activation for the model. The proposed work enhances the trust and reliability of the GAN model, which is already used for fake face detection with reduced loss. Specifically the fusion Model-2 with Densenet-201, ResNet-50 and Inception V3 produced a loss value of 0.1146 with minimum training, which is the most desirable for the domestic applications. The accuracy of Model-3 with Densenet-201, Resnet-102 and Xception is 0.9797 which is almost equal to any advanced implementation on GAN with less significant datasets. As a result, the suggested approach recommends combining GAN with fusion DL algorithms to improve accuracy and reduce the loss of the generator and discriminator sections of the GAN.

Further research in the domain can offer valuable insights and advancements in several areas related to GANs. First, exploring the explainability of the loss function and accuracy of GANs can contribute to a deeper understanding of their inner workings, aiding in the interpretation of model outputs. Second, investigating the application of federated learning in estimating accuracy and loss for GAN networks presents an avenue for decentralized and collaborative model training, potentially enhancing robustness and privacy. Additionally, conducting a comparative analysis between single DL models and fusion models can provide a comprehensive performance evaluation, shedding light on the effectiveness of integrated approaches for tasks such as fake face detection. These avenues of research hold the potential to advance the field and address critical aspects of GAN-based systems.

### Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

### Acknowledgments

The authors extend their appreciation to King Saud University for funding this work through Researchers Supporting Project number (RSPD2023R704), King Saud University, Riyadh, Saudi Arabia.

### Conflict of interest

The authors declare there is no conflict of interest.

### References

1. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, et al., Generative adversarial nets, *Adv. Neural Inf. Process. Syst.*, (2014), 2672–2680. <https://doi.org/10.48550/arXiv.1406.2661>
2. M. S. Rana, M. N. Nobi, B. Murali, A. H. Sung, Deepfake detection: A systematic literature review, *IEEE Access*, 2022. <https://doi.org/10.1109/ACCESS.2022.3154404>
3. X. Deng, B. Zhao, Z. Guan, M. Xu, A new finding and unified framework for fake image detection, *IEEE Signal Process. Lett.*, **30** (2023), 90–94. <https://doi.org/10.1109/LSP.2023.3243770>

4. J. Peng, B. Zou, C. Zhu, Combining external attention gan with deep convolutional neural networks for real–fake identification of luxury handbags, *Vis. Comput.*, **39** (2023), 971–982. <https://doi.org/10.1007/s00371-021-02378-x>
5. M. Zhang, L. Zhao, B. Shi, Analysis and construction of the coal and rock cutting state identification system in coal mine intelligent mining, *Sci. Rep.*, **13** (2023), 3489. <https://doi.org/10.1038/s41598-023-30617-9>
6. Y. LeCun, Y. Bengio, G. Hinton, DL, *Nature*, **521** (2015), 436–444. <https://doi.org/10.1038/nature14539>
7. H. Ding, Y. Sun, Z. Wang, N. Huang, Z. Shen, X. Cui, et al., Rgan-el: A gan and ensemble learning-based hybrid approach for imbalanced data classification, *Inf. Process. Manage.*, **60** (2023), 103235. <https://doi.org/10.1016/j.ipm.2022.103235>
8. V. P. Manikandan, U. Rahamathunnisa, A neural network aided attuned scheme for gun detection in video surveillance images, *Image Vision Comput.*, **120** (2022), 104406. <https://doi.org/10.1016/j.imavis.2022.104406>
9. J. Kolluri, R. Das, Intelligent multimodal pedestrian detection using hybrid meta-heuristic optimization with DL model, *Image Vision Comput.*, (2023), 104628. <https://doi.org/10.1016/j.imavis.2023.104628>
10. K. Minche, *Vision Transformer-Assisted Analysis of Neural Image Compression and Generation*, PhD thesis, 2022. <http://dx.doi.org/10.26153/tsw/43656>
11. N. Kumari, R. Zhang, E. Shechtman, J. Y. Zhu, Ensembling off-the-shelf models for gan training, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2022), 10651–10662.
12. M. Alawadhi, W. Yan, DL from parametrically generated virtual buildings for real-world object recognition, preprint, arXiv:2302.05283. <https://doi.org/10.48550/arXiv.2302.05283>
13. Y. Wang, C. Peng, D. Liu, N. Wang, X. Gao, Forgerynir: Deep face forgery and detection in near-infrared scenario, *IEEE Trans. Inf. Forensics Secur.*, **17** (2022), 500–515. <https://doi.org/10.1109/TIFS.2022.3146766>
14. H. Zhang, B. Chen, J. Wang, G. Zhao, A local perturbation generation method for gan-generated face anti-forensics, *IEEE Trans. Circuits Syst. Video Technol.*, 2022. <https://doi.org/10.1109/TCSVT.2022.3207310>
15. W. Wang, X. Wang, W. Yang, J. Liu, Unsupervised face detection in the dark, *IEEE Trans. Pattern Anal. Mach. Intell.*, **45** (2022), 1250–1266. <https://doi.org/10.1109/TPAMI.2022.3152562>
16. M. Khosravy, K. Nakamura, Y. Hirose, N. Nitta, N. Babaguchi, Model inversion attack by integration of deep generative models: Privacy-sensitive face generation from a face recognition system, *IEEE Trans. Inf. Forensics Secur.*, **17** (2022), 357–372. <https://doi.org/10.1109/TIFS.2022.3140687>
17. A. Benlamoudi, S. E. Bekhouche, M. Korichi, K. Bensid, A. Ouahabi, A. Hadid, et al., Face presentation attack detection using deep background subtraction, *Sensors*, **22** (2022), 3760. <https://doi.org/10.3390/s22103760>

18. S. H. Silva, M. Bethany, A. M. Votto, I. H. Scarff, N. Beebe, P. Najafirad, et al., Deepfake forensics analysis: An explainable hierarchical ensemble of weakly supervised models, *Forensic Sci. Int.: Synergy*, **4** (2022), 100217. <https://doi.org/10.1016/j.fsisyn.2022.100217>
19. S. Selitskiy, N. Christou, N. Selitskaya, Using statistical and artificial neural networks meta-learning approaches for uncertainty isolation in face recognition by the established convolutional models, in *Machine Learning, Optimization, and Data Science: 7th International Conference, LOD 2021, Grasmere, UK, October 4–8, 2021, Revised Selected Papers, Part II*, (2022), 338–352.
20. A. Gowda, N. Thillaiarasu, Investigation of comparison on modified cnn techniques to classify fake face in deepfake videos, in *2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS)*, **1** (2022), 702–707. <https://doi.org/10.1109/ICACCS54159.2022.9785092>
21. S. Rao, N. A. Shelke, A. Goel, H. Bansal, Deepfake creation and detection using ensemble DL models, in *Proceedings of the 2022 Fourteenth International Conference on Contemporary Computing*, (2022), 313–319. <https://doi.org/10.1145/3549206.3549263>
22. K. R. Revi, M. M. Isaac, R. Antony, M. Wilscy, Gan-generated fake face image detection using opponent color local binary pattern and DL technique, in *2022 International Conference on Connected Systems & Intelligence (CSI)*, (2022), 1–7. <https://doi.org/10.1109/CSI54720.2022.9924077>
23. S. Lim, M. Shin, J. Paik, Point cloud generation using deep adversarial local features for augmented and mixed reality contents, *IEEE Trans. Consum. Electron.*, **68** (2022), 69–76. <https://doi.org/10.1109/TCE.2022.3141093>
24. Y. Zhang, A. Wang, W. Hu, DL-based consumer behavior analysis and application research, *Wireless Commun. Mobile Comput.*, **2022** (2022). <https://doi.org/10.1155/2022/4268982>
25. W. Zheng, M. Yue, S. Zhao, S. Liu, Attention-based spatial-temporal multi-scale network for face anti-spoofing, *IEEE Trans. Biom., Behav., Identity Sci.*, **3** (2021), 296–307. <https://doi.org/10.1109/TBIOM.2021.3066983>
26. S. Zhao, W. Liu, S. Liu, J. Ge, X. Liang, A hybrid-supervision learning algorithm for real-time un-completed face recognition, *Comput. Electr. Eng.*, **101** (2022), 108090. <https://doi.org/10.1016/j.compeleceng.2022.108090>
27. S. Kiruthika, V. Masilamani, Image quality assessment based fake face detection, *Multimed. Tools Appl.*, **82** (2023), 8691–8708. <https://doi.org/10.1007/s11042-021-11493-9>
28. A. M. Luvembe, W. Li, S. Li, F. Liu, G. Xu, Dual emotion based fake news detection: A deep attention-weight update approach, *Inf. Process. Manage.*, **60** (2023), 103354. <https://doi.org/10.1016/j.ipm.2023.103354>
29. S. Li, W. Li, A. M. Luvembe, W. Tong, Graph contrastive learning with feature augmentation for rumor detection, *IEEE Trans. Comput. Social Syst.*, 2023. <https://doi.org/10.1109/TCSS.2023.3269303>
30. F. Baratzadeh, S. M. H. Hasheminejad, Customer behavior analysis to improve detection of fraudulent transactions using DL, *J. AI Data Mining*, **10** (2022), 87–101. <https://doi.org/10.22044/jadm.2022.10124.2151>



31. P. Bamoriya, G. Siddhad, H. Kaur, P. Khanna, A. Ojha, Dsb-gan: Generation of DL based synthetic biometric data, *Displays*, **74** (2022), 102267. <https://doi.org/10.1016/j.displa.2022.102267>
32. W. Qian, H. Li, H. Mu, Circular lbp prior-based enhanced gan for image style transfer, *Int. J. Semantic Web Inf. Syst. (IJSWIS)*, **18** (2022), 1–15. <https://doi.org/10.4018/IJSWIS.315601>
33. Z. Zhou, Y. Li, J. Li, K. Yu, Guang Kou, M. Wang, et al., Gan-siamese network for cross-domain vehicle re-identification in intelligent transport systems, *IEEE Trans. Network Sci. Eng.*, 2022. <https://doi.org/10.1109/TNSE.2022.3199919>
34. S. S. Reka, P. Venugopal, V. Ravi, T. Dragicevic, Privacy-based demand response modeling for residential consumers using machine learning with a cloud–fog-based smart grid environment, *Energies*, **16** (2023), 1655. <https://doi.org/10.3390/en16041655>
35. Y. Wang, H. R. Tohidypour, M. T. Pourazad, P. Nasiopoulos, V. C. M. Leung, DL-based hdr image upscaling approach for 8k uhd applications, in *2022 IEEE International Conference on Consumer Electronics (ICCE)*, (2022), 1–2. <https://doi.org/10.1109/ICCE53296.2022.9730460>
36. Xhlulu, 140k real and fake faces, Feb 2020.
37. CIPLAB @ Yonsei University, Real and fake face detection, Jan 2019.



AIMS Press

© 2024 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)