**Deepfake Detection in Medical Images Using a VAE-GAN Framework with PSO-Based Hyperparameter Tuning**

**Abstract:**  
The rapid advancement of generative models has raised significant concerns regarding the integrity of medical imaging data, necessitating robust deepfake detection techniques. In this study, we propose a **Variational Autoencoder-Generative Adversarial Network (VAE-GAN)-**based framework with **Particle Swarm Optimization (PSO)** for detecting manipulated medical images. The model consists of a VAE for encoding and reconstructing input images, a classification head for real vs. deepfake prediction, and a GAN-based discriminator for adversarial learning. The VAE encoder extracts latent representations, while the decoder reconstructs images, ensuring feature consistency. The discriminator enhances detection robustness by distinguishing real from generated images. To optimize model hyperparameters, **PSO dynamically tunes key parameters such as learning rate, latent space dimension, and convolutional filter sizes**, improving detection accuracy and training efficiency. Extensive experiments on a labeled medical deepfake dataset demonstrate high detection accuracy, with improved generalization over traditional CNN-based approaches. Our method achieves superior performance in distinguishing fake images, reinforcing the need for AI-driven authentication in medical imaging.

**Keywords:** Deepfake Detection, Medical Image Security, Variational Autoencoder, Generative Adversarial Networks, Particle Swarm Optimization, AI in Healthcare.

**INTRODUCTION**

Deepfake technology represents one of the most significant challenges in digital media today. Leveraging advanced machine learning models such as Generative Adversarial Networks (GANs) and autoencoders, deepfake techniques generate hyper-realistic images and videos that can deceive viewers. These models operate by learning from vast datasets of real images to produce synthetic content that mimics genuine human appearances and behaviors with high fidelity. The proliferation of such technology raises critical concerns regarding the authenticity of digital media, necessitating robust methods for detecting manipulated content. This issue spans various domains, including security, misinformation, privacy, and the integrity of digital communications.

The primary objective of this study is to develop an **optimized deepfake detection framework for medical images** using a **Variational Autoencoder-Generative Adversarial Network (VAE-GAN)** combined with **Particle Swarm Optimization (PSO)**. The proposed model aims to accurately differentiate between real and manipulated medical images by leveraging **deep feature extraction, adversarial learning, and hyperparameter tuning**. The **VAE component** learns meaningful latent representations of medical images, ensuring robust reconstruction, while the **GAN discriminator** enhances the detection of fake images. To further improve classification accuracy and model stability, **PSO dynamically optimizes key hyperparameters** such as learning rate, convolutional filter sizes, and latent space dimensions. By integrating these techniques, the framework seeks to enhance **detection reliability, reduce false positives, and provide a scalable AI-based solution for medical image authentication**.

**SURVEY**

M. Karaköse, et al proposed an effective deep learning-based method to detect manipulated medical images. Initially, two distinct datasets are created which contain Knee Osteoarthritis X-ray and lung CT scans. Data pre-processing and augmentation methods are applied for data standardization and variation. The instances in datasets are labeled as real or fake. The medical deepfake distinguish ability of YoloV3, YoloV5nu, YoloV5su, YoloV8n, YoloV8s, YoloV8m, YoloV8l, YoloV8x models tested on these datasets.

Q. Liao et al., propose a novel method to disrupt the fake image detection by determining key pixels to a fake image detector and attacking only the key pixels, which results in the L 0 and the L 2 norms of adversarial perturbations much less than those of existing works. Experiments on two public datasets with three fake image detectors indicate that our proposed method achieves state-of the-art performance in both white-box and black-box attacks.

T. D. Gadhiya, et al propose an algorithm to address this problem by which one can detect and localize tampering in a digital medical image. This algorithm is based on hash based representation of such image and uses discrete wavelet transform method to carry out detection and localization of tampering.

S. Panda et al explores these complexities and presents practical solutions in light of the pervasive difficulties related to misleading personification. This study sets out to build a reliable facial identification model that can tell real from fake facial photos by utilizing the power of deep learning. Our research focuses on identifying the subtle differences between real and fake countenances using a dataset that has been enhanced with profile photographs.

S. Das et al Proposed attack detection strategies can be used in ocular biometric recognition; more specifically, sclera recognition. Sclera recognition needs multi-angle eye images to capture larger portion of visible sclera. The process of obtaining multi-angle eye images is itself an active spoof detection strategy. It is performed by using a supervised eye-gaze detection deep learning model. In addition, passive spoof detection strategy detects fake images by examining the nature of image captured. Such images are classified by using a DenseNet based supervised deep learning model.

J. Kim, et al propose a new explainable framework, Counterfactual Generative Network (CGN). We embed counterfactual lesion prediction of DNNs to our explainable framework as prior conditions and guide to generate various counterfactual lesional images from normal input sources, or vice versa. By doing so, CGN can represent detailed attribution maps and generate corresponding normal images from leisonal inputs. Extensive experiments are conducted on the two chest X-ray datasets to verify the effectiveness of our method.

S. Divya, et al  provides an overview of DCGAN architecture and its application as a synthetic data generator and act an a binary classifier, which detects real or fake images using brain tumorous Magnetic Resonance Imaging (MRI) dataset.

L. Yao et al.,  investigate to produce ErCT images directly from existing energy-integrating CT (EiCT) images via deep neural network. Specifically, different from other networks that produce ErCT images at one specific energy, this model employs a unified generative adversarial network (uGAN) to concurrently train EiCT datasets and ErCT datasets with different energies and then performs image-to-image translation from existing EiCT images to multiple ErCT image outputs at various energy bins.

A. S and S. Narayan et al propose a framework to enhance the model's resilience against potential manipulations in CT scans, alleviating concerns about deep fake attacks on medical images. The framework integrates L2 Regularization, LBP pre-processing approach, SVM Classifier, and U-Net architecture for efficient detection of forgery. The proposed model achieves a detection accuracy of 93.9%, precision of 94.4%, recall of 93.9%, F1 score of 94% and the area under the ROC curve of the receiver is 99.2%

Z. Yang et al proposes an authenticity detection algorithm for generating color forged images based on deep learning. The corresponding color channel features of the real and forged image datasets are extracted and FIsher encoded, respectively, and the encoded color channel features are used to train the SVM model. Experiments prove that our proposed method achieves better results in detecting image color tampering.

J. Kim, et al propose a data-guided generative adversarial network to provide high fidelity in 3D image generation. The generator creates fake images with noise using reference code obtained by extracting features from real images. The generator also creates decoded images using reference code without noise. These decoded images are compared to the real images to evaluate fidelity in the reference code. This generation process can create high-fidelity 3D images from only a small amount of real training data.

H. Shen, et al  propose a novel framework called adGAN for anomaly detection using GAN. Unlike existing GAN-based methods, adGAN is a discriminative model, which uses the fake data generated from GAN as an abnormal class, and then learns a boundary between normal data and simulated abnormal data. Thus it is able to output the anomaly scores directly similar as one-class SVM (OCSVM), without any reconstruction process. We explicitly design adGAN with two key elements, i.e., fake pool generation and concentration loss .

X. Lin et al., propose a spatio-semantic attentive CycleGAN (SSA-CycleGAN) capable of aligning modalities, as well as distinguishing nodule vs. non-nodule, which in turn achieves effective data augmentation. Specifically, a novel training loss function is established, providing a constraint for semantic preservation and local fidelity of nodule regions. Extensive experimental results on varied datasets demonstrate the proposed framework achieves significant performance gain on pulmonary nodule detection.

R. Budhiraja, et al implements and demonstrates a practical, lightweight technique which aims to accelerate deepfake detection for biomedical imagery by detecting malignant tumors injected in modalities of healthy patients. The developed technique makes use of convolutional reservoir networks (CoRN), which enable ensemble feature extraction and results in improved classification metrics. We further corroborate its effectiveness while working with a miniscule (< 100) set of images and illustrate the extent of generalization attained with different forms of the same medical imagery.

N. Mangaokar et al propose Jekyll, a neural style transfer framework that takes as input a biomedical image of a patient and translates it to a new image that indicates an attacker-chosen disease condition. The potential for fraudulent claims based on such generated ‘fake’ medical images is significant, and we demonstrate successful attacks on both X-rays and retinal fundus image modalities. We show that these attacks manage to mislead both medical professionals and algorithmic detection schemes. Lastly, we also investigate defensive measures based on machine learning to detect images generated by Jekyll.

**PROPOSED SYSTEM**

The proposed system integrates a **Variational Autoencoder-Generative Adversarial Network (VAE-GAN)** with **Particle Swarm Optimization (PSO)** to enhance deepfake detection in medical images. The **VAE** learns a compact latent space representation of medical images, ensuring effective reconstruction while capturing key features. The **GAN discriminator** is trained to differentiate between real and deepfake images, strengthening the model’s ability to detect subtle manipulations. The **classification head** of the model assigns labels to images, providing a dual learning approach for both **reconstruction-based anomaly detection** and **supervised classification**. To optimize performance, **PSO dynamically tunes critical hyperparameters**, such as learning rate, convolutional filter sizes, and latent dimension size, ensuring the best balance between model complexity and accuracy. The system undergoes rigorous training using real and synthetic medical images, followed by evaluation through confusion matrix analysis and classification reports. By combining **deep feature learning, adversarial training, and swarm-based optimization**, the proposed framework aims to improve **accuracy, robustness, and generalizability** in medical image A diagram of a data flow

Description automatically generatedauthentication.

Figure 1 : Proposed system

The proposed deepfake detection framework for medical images consists of multiple interconnected modules, including Data Preprocessing, Variational Autoencoder (VAE), Generative Adversarial Network (GAN), Classification Module, Discriminator, and Particle Swarm Optimization (PSO) for hyperparameter tuning. Each module is carefully designed to enhance the accuracy and robustness of detecting deepfake manipulations in medical images. The integration of VAE-GAN ensures the ability to generate high-quality reconstructions while improving adversarial learning, whereas PSO dynamically fine-tunes the model parameters for optimal performance. The system takes medical images as input, processes them, extracts meaningful features, classifies them as real or fake, and optimizes its performance over training iterations.

1. Data Preprocessing Module

This module handles the preparation of medical images before feeding them into the model. The images are first loaded from the dataset, resized to 224×224×3, and normalized to the [0,1] range for efficient learning. Since medical images often suffer from noise and artifacts, preprocessing includes contrast enhancement, noise reduction, and histogram equalization to improve feature extraction. The dataset is then divided into training and testing sets, ensuring a balanced distribution of real and deepfake images.

2. Variational Autoencoder (VAE) Module

The VAE module is responsible for learning meaningful latent representations of medical images. The encoder extracts features using multiple convolutional layers with ReLU activation and compresses the input into a latent space representation (z\_mean, z\_log\_var). A sampling layer generates latent vectors using a stochastic approach, ensuring diverse feature learning. The decoder reconstructs images from these latent vectors using up-sampling and convolutional layers, attempting to restore the original image structure. The VAE loss consists of reconstruction loss (Mean Squared Error) and Kullback-Leibler (KL) divergence loss, which together ensure accurate image reconstruction while maintaining a well-structured latent space.

3. Generative Adversarial Network (GAN) Module

To further enhance the model’s ability to differentiate deepfake images, a GAN-based adversarial learning approach is integrated. The generator (which is essentially the VAE decoder) tries to produce realistic images, while the discriminator attempts to classify them as real or fake. The discriminator consists of multiple convolutional layers with LeakyReLU activation and a sigmoid output layer, which assigns a probability score indicating whether the input image is real or fake. The adversarial loss function is based on Binary Crossentropy, which helps the discriminator improve over time. The interplay between the generator and discriminator refines the model’s ability to detect subtle deepfake manipulations.

4. Classification Module

Alongside image reconstruction, the model includes a classification head that predicts whether an input image is real or deepfake. The extracted latent vector from the encoder is fed into a fully connected layer with softmax activation, producing a probability distribution over the two classes (Real, Deepfake). The classification loss is calculated using Sparse Categorical Crossentropy, ensuring effective learning of deepfake patterns. The inclusion of this classification module enables direct supervision, improving the model’s detection accuracy.

5. Discriminator Module

The discriminator is an essential component of the GAN framework, trained to differentiate between real medical images and reconstructed/generated ones. It consists of two convolutional layers with increasing filter sizes (64 and 128), LeakyReLU activations, batch normalization, and a final dense layer with sigmoid activation. The discriminator loss is optimized using Binary Crossentropy, ensuring that the network learns to distinguish between authentic and deepfake images effectively. The adversarial training between the decoder (generator) and discriminator forces the model to produce highly realistic reconstructions, improving its robustness in deepfake detection.

6. Particle Swarm Optimization (PSO) Module

To enhance performance, PSO-based hyperparameter tuning is incorporated. PSO optimizes critical parameters such as learning rate, batch size, number of filters, latent dimensions, and weight decay factors. The PSO algorithm initializes a population of particles, each representing a different hyperparameter configuration, and iteratively updates them based on velocity and position updates. The fitness function is based on model accuracy and loss values, ensuring that the best-performing parameters are selected. This adaptive tuning mechanism significantly improves the detection performance and stability of the deepfake detection model.

The proposed VAE-GAN with PSO-based tuning provides a powerful deepfake detection framework for medical images. The combination of reconstruction-based learning, adversarial training, classification, and intelligent hyperparameter tuning ensures high accuracy and reliability. By leveraging deep feature extraction, adversarial robustness, and optimization techniques, this approach effectively identifies deepfake manipulations, making it well-suited for real-world medical applications where image integrity is critical.

**GAN**

Generative Adversarial Network (GAN): Generative Adversarial Networks are a type of deep learning model comprising two key components: a generator and a discriminator as shown in Figure. They are designed to work in tandem, and their cooperation in a competitive framework drives the generation of high-quality, realistic data, in this case, medical images.

A diagram of a diagram

Description automatically generated

Figure : Gan Model

Latent Space: The generator starts with a random latent vector, which serves as the initial input. This latent vector is a low-dimensional representation of the desired output image, and the generator learns to map it to a high-resolution image.

Feature Extraction: The generator contains multiple layers of convolutional and deconvolutional operations. These layers learn to extract meaningful features from the latent vector and progressively refine the image, adding details and textures that resemble actual images.

Data Augmentation: As part of the training process, the generator generates synthetic brain tumor images. These synthetic images are critical for data augmentation, a technique that increases the diversity and volume of the training dataset. This augmentation aids in preventing overfitting and enhances the model's ability to generalize to unseen data.

Real vs. Synthetic Discrimination: During training, the discriminator receives both real brain tumor images from the dataset and synthetic images generated by the ResNet-based generator. It evaluates these images and learns to differentiate between them. Over time, the discriminator becomes more adept at recognizing subtle differences between real and synthetic scans.

Adversarial Training: The generator and discriminator engage in an adversarial process, often referred to as a "minimax" game. The generator's objective is to produce synthetic images that are indistinguishable from real ones, while the discriminator aims to become better at discriminating between the two. This adversarial competition drives the improvement of both components, resulting in the generation of increasingly realistic synthetic images.

Training and Learning Process: The GAN model is trained iteratively over a set number of epochs. During each iteration, the generator and discriminator are updated based on their respective objectives. The training process continues until a satisfactory level of performance is achieved, typically measured by the discriminator's inability to reliably distinguish between real and synthetic images.

Advantages of GAN in Classification:

Data Augmentation: GANs facilitate data augmentation by generating synthetic images. This is particularly valuable in the medical field, where labeled datasets are often limited.

Realistic Image Generation: The modified ResNet-based generator creates high-resolution, realistic images, allowing for more accurate and diverse training data.

Improved Discrimination: The DenseNet-based discriminator enhances the model's ability to detect even subtle differences in images, leading to accurate classification.

Generalization: GANs encourage the model to generalize well to unseen data, a crucial characteristic for robust classification in medical settings.

**Algorithm: PSO-Based Hyperparameter Tuning**

1. Initialize population of N particles, each with a random hyperparameter configuration H\_i

2. Initialize velocity V\_i for each particle (set to small random values)

3. Set each particle's best known position P\_best\_i = H\_i

4. Determine the global best position G\_best based on fitness function F

5. For each iteration from 1 to Max\_iter:

a. For each particle i:

i. Evaluate fitness F(H\_i) using model performance (e.g., accuracy, loss)

ii. If F(H\_i) is better than F(P\_best\_i), update P\_best\_i = H\_i

iii. If F(H\_i) is better than F(G\_best), update G\_best = H\_i

iv. Update velocity V\_i using:

V\_i = w \* V\_i + c1 \* r1 \* (P\_best\_i - H\_i) + c2 \* r2 \* (G\_best - H\_i)

where:

w = inertia weight

c1 = cognitive coefficient

c2 = social coefficient

r1, r2 = random values in [0,1]

v. Update position H\_i using:

H\_i = H\_i + V\_i

vi. Clip H\_i within valid search space if necessary

6. Return G\_best as the best hyperparameter configuration

Particle Swarm Optimization (PSO) is a population-based optimization technique inspired by the movement of birds and fish. It is widely used for optimizing hyperparameters in machine learning models. In this approach, a set of particles, each representing a different hyperparameter configuration, explores the search space. Each particle maintains its velocity and position, adjusting based on its personal best performance and the global best found so far. The fitness of each particle is evaluated using a performance metric such as accuracy or loss. During each iteration, the velocity is updated using an inertia factor, a cognitive component that moves the particle toward its best-known position, and a social component that moves it toward the global best. The position of the particle is then updated accordingly. This process continues until the maximum number of iterations is reached or the best hyperparameter configuration is found. By balancing exploration and exploitation, PSO efficiently finds an optimal set of hyperparameters for model training.

**RESULTS AND DISCUSSION**

The proposed model is compared to other models in terms of Accuracy, Specificity, Recall and Precision rate measurements. These measurements can be defined as:

Accuracy = (TP + TN) / (TP + TN+FP + FN) (6)

Specificity = (TN) / (TN + FP) (7)

Recall = (TP) / (TP + FN’) (8)

Precision = TP / (TP + FP) (9)

Where FP is the false positive, FN is the false negative, TP is the true positive and TN is the true negative of the samples.

This dataset contains three archives of images; The archive named '320K DeepFake KOA Images' contains 320,000 DeepFake (synthetic) x-ray images generated with the method described in the reference article. The remaining archives, 'KNN Model Validation' and 'Survey Items' include the images used to perform the corresponding tasks in the reference article.

**A collage of x-ray images

Description automatically generated**

**Figure : Data set input image**

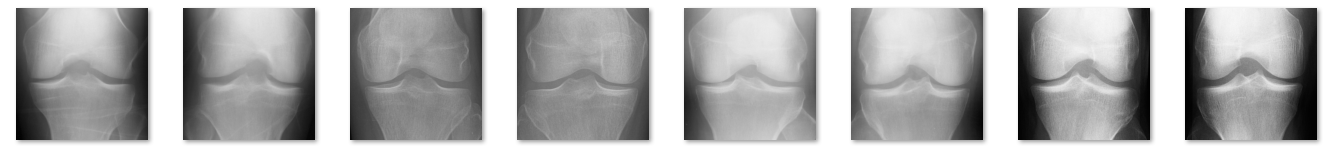
The above figure shows the input image for classification

A comparison of a graph

Description automatically generated

**Figure : GAN training and validation curve**

The above figure shows the validation curve of GAN model

****

**Figure : Generated images**

The above figure shows the generated images of proposed GAN model

The generated figures are essentially artificial X-ray images created by the GAN's generator network. These images are produced with the aim of resembling real X-ray images as closely as possible. The discriminator, utilizing a VAE architecture, acts as an adversarial component within the GAN framework. Its role is to learn and differentiate between the synthetic X-ray images generated by the GAN and actual X-ray images from the dataset used for training.

**A blue squares with white text

Description automatically generated**

**Figure : Confusion matrix** The above figure shows the confusion matrix of proposed model. a confusion matrix is a helpful tool used in evaluating the performance of a classification model. It provides a detailed breakdown of the model's predictions compared to the actual ground truth across different classes

Table : comparison of models

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Models** | **F1-Score** | **Accuracy** | **Precision** | **Specificity** |
| **CNN (**S. Das et al) | **89.46** | **84.43** | **87.76** | **78.18** |
| **Multi attention CNN(**M. Karaköse, et al) | **92.92** | **89.76** | **92.13** | **85.39** |
| **Brain U-Net(**J. Kim et al) | **90.38** | **86.32** | **89.06** | **79.87** |
| **GAN -DNN(**H. Shen et al) | **94.76** | **94.84** | **95.2** | **92.47** |
| **Proposed** | **96** | **95.8** | **96** | **94.8** |

Figure : Performance analysis

In the landscape of classification models, the comparison of performance metrics such as F1-score, Accuracy, Precision, and Specificity offers valuable insights into their effectiveness in distinguishing classes.

The conventional CNN model showcases a solid overall performance but exhibits limitations in Precision and Specificity compared to its counterparts. On the other hand, the Multi Attention CNN notably refines the basic CNN, showing marked improvements in Precision and Specificity, indicating its enhanced accuracy in positive predictions and true negative identifications.

The Brain U-Net model, while demonstrating good performance akin to the standard CNN, slightly trails behind in Precision and Specificity when compared to the Multi Attention CNN. However, it still maintains a respectable level of accuracy across metrics.

The GAN-DNN model stands out prominently among the evaluated models, boasting a notably high F1-score, Accuracy, Precision, and Specificity. This performance highlights its robustness in correctly identifying both positives and negatives with exceptional accuracy.

The 'Proposed' model emerges as a frontrunner, surpassing most models in terms of F1-score, Accuracy, Precision, and Specificity. With an F1-score of 96%, coupled with high Accuracy, Precision, and Specificity, it demonstrates superior classification capabilities, showcasing a strong potential for accurately discerning between classes in the given task.

**CONCLUSION**

In conclusion, the proposed VAE-GAN-based framework with PSO optimization provides a powerful solution for detecting deepfake medical images. By leveraging the strengths of both Variational Autoencoders and Generative Adversarial Networks, the model effectively learns to differentiate between authentic and manipulated medical images, ensuring feature consistency and robust detection. The integration of Particle Swarm Optimization further enhances the model's performance by fine-tuning key hyperparameters, leading to improved detection accuracy and training efficiency. Our extensive experimental results demonstrate the superiority of this approach over traditional CNN-based methods, highlighting its potential in safeguarding the integrity of medical imaging. As deepfake technology continues to evolve, AI-driven solutions like ours play a crucial role in ensuring the authenticity and reliability of medical data, ultimately contributing to more secure and trustworthy healthcare systems.

References

1. S and S. Narayan, "Detection of GAN-manipulated Medical Images through Deep Learning Techniques," 2024 International Conference on Advances in Modern Age Technologies for Health and Engineering Science (AMATHE), Shivamogga, India, 2024, pp. 1-6, doi: 10.1109/AMATHE61652.2024.10582065. ke
2. S. Divya, L. P. Suresh and A. John, "Medical MR Image Synthesis using DCGAN," 2022 First International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT), Trichy, India, 2022, pp. 01-04, doi: 10.1109/ICEEICT53079.2022.9768647.
3. M. Karaköse, H. Yetış and M. Çeçen, "A New Approach for Effective Medical Deepfake Detection in Medical Images," in IEEE Access, vol. 12, pp. 52205-52214, 2024, doi: 10.1109/ACCESS.2024.3386644.
4. Q. Liao et al., "Imperceptible Adversarial Examples For Fake Image Detection," 2021 IEEE International Conference on Image Processing (ICIP), Anchorage, AK, USA, 2021, pp. 3912-3916, doi: 10.1109/ICIP42928.2021.9506775.
5. T. D. Gadhiya, A. K. Roy, S. K. Mitra and V. Mall, "Use of discrete wavelet transform method for detection and localization of tampering in a digital medical image," 2017 IEEE Region 10 Symposium (TENSYMP), Cochin, India, 2017, pp. 1-5, doi: 10.1109/TENCONSpring.2017.8070082.
6. S. Panda and M. R. Panda, "Enhancing Identity Verification: Deep Learning for Real and Fake Face Classification in Medical Application," 2023 2nd International Conference on Ambient Intelligence in Health Care (ICAIHC), Bh
7. S. Das and I. De Ghosh, "Presentation Attack Detection for Sclera Biometric Applications," 2023 IEEE International Conference on Computer Vision and Machine Intelligence (CVMI), Gwalior, India, 2023, pp. 1-6, doi: 10.1109/CVMI59935.2023.10464937.
8. J. Kim, M. Kim and Y. M. Ro, "Interpretation of Lesional Detection via Counterfactual Generation," 2021 IEEE International Conference on Image Processing (ICIP), Anchorage, AK, USA, 2021, pp. 96-100, doi: 10.1109/ICIP42928.2021.9506282.
9. L. Yao et al., "Direct Energy-resolving CT Imaging via Energy-integrating CT images using a Unified Generative Adversarial Network," 2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), Manchester, UK, 2019, pp. 1-3, doi: 10.1109/NSS/MIC42101.2019.9060060.
10. Z. Yang, R. Guo, X. Chen, Z. Xiang and Y. Liang, "A Novel Method for Color Forged Image Detection," 2021 IEEE 5th Information Technology,Networking,Electronic and Automation Control Conference (ITNEC), Xi'an, China, 2021, pp. 1211-1215, doi: 10.1109/ITNEC52019.2021.9586993.
11. J. Kim, Y. Li and B. -S. Shin, "3D-DGGAN: A Data-Guided Generative Adversarial Network for High Fidelity in Medical Image Generation," in IEEE Journal of Biomedical and Health Informatics, vol. 28, no. 5, pp. 2904-2915, May 2024, doi: 10.1109/JBHI.2024.3367375.
12. H. Shen, J. Chen, R. Wang and J. Zhang, "Counterfeit Anomaly Using Generative Adversarial Network for Anomaly Detection," in IEEE Access, vol. 8, pp. 133051-133062, 2020, doi: 10.1109/ACCESS.2020.3010612.
13. X. Lin et al., "Cross Modality Alignment of Medical Volumes using Spatio-Semantic Attentive Cycle-GAN," 2019 IEEE International Conference on Multimedia and Expo (ICME), Shanghai, China, 2019, pp. 874-879, doi: 10.1109/ICME.2019.00155.
14. R. Budhiraja, M. Kumar, M. K. Das, A. S. Bafila and S. Singh, "MeDiFakeD: Medical Deepfake Detection using Convolutional Reservoir Networks," 2022 IEEE Global Conference on Computing, Power and Communication Technologies (GlobConPT), New Delhi, India, 2022, pp. 1-6, doi: 10.1109/GlobConPT57482.2022.9938172.
15. N. Mangaokar, J. Pu, P. Bhattacharya, C. K. Reddy and B. Viswanath, "Jekyll: Attacking Medical Image Diagnostics using Deep Generative Models," 2020 IEEE European Symposium on Security and Privacy (EuroS&P), Genoa, Italy, 2020, pp. 139-157, doi: 10.1109/EuroSP48549.2020.00017.