IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 27 November 2023, accepted 6 January 2024, date of publication 15 January 2024, date of current version 23 September 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3354724

METHODS

An Empirical Analysis of Diffusion, Autoencoders, and Adversarial Deep Learning Models for Predicting Dementia Using High-Fidelity MRI

PRANSHAV GAJJAR[1], MANAV GARG[2], SHIVANI DESAI[3,4], HITESH CHHINKANIWALA[5], HARSHAL A. SANGHVI[6,7], RIKI H. PATEL[8], SHAILESH GUPTA[7,9], AND ABHIJIT S. PANDYA[10]

[1]College of Engineering and Computing, George Mason University, Fairfax, VA 22030, USA
[2]School of Computing and Augmented Intelligence, Arizona State University, Tempe, AZ 85287, USA
[3]Gujarat Technological University, Ahmedabad 382424, India
[4]Department of Computer Science and Engineering, Institute of Technology, Nirma University, Ahmedabad 382481, India
[5]Department of Information and Communication Technology, Adani Institute of Infrastructure Engineering (AIIE), Ahmedabad 382421, India
[6]Florida Atlantic University, Boca Raton, FL 33431, USA
[7]Department of Technology and Clinical Trials, Advanced Research, USA
[8]Canopy Security, Boca Raton, FL 33431, USA
[9]Broward Health, Fort Lauderdale, FL 33316, USA
[10]Department of Computer and Electrical Engineering and Computer Science (CEECS), Florida Atlantic University, Boca Raton, FL 33431, USA

Corresponding authors: Pranshav Gajjar (pgajjar@gmu.edu) and Riki H. Patel (rikipatel26@gmail.com)

ABSTRACT This study explores cutting-edge computational technologies and intelligent methods to create realistic synthetic data, focusing on dementia-centric Magnetic Resonance Imaging (MRI) scans related to Alzheimer's and Parkinson's diseases. The research delves into Generative Adversarial Networks (GANs), Variational Autoencoders, and Diffusion Models, comparing their efficacy in generating synthetic MRI scans. Using datasets from Alzheimer's and Parkinson's patients, the study reveals intriguing findings. In the Alzheimer dataset, diffusion models produced non-dementia images with the lowest Frechet Inception Distance (FID) score at 92.46, while data-efficient GANs excelled in generating dementia images with an FID score of 178.53. In the Parkinson dataset, data-efficient GANs achieved remarkable FID scores of 102.71 for dementia images and 129.77 for non-dementia images. The study also introduces a novel aspect by incorporating a classification study, validating the generative metrics. DenseNets, a deep learning architecture, exhibited superior performance in disease detection compared to ResNets. Training both models on images generated by diffusion models further improved results, with DenseNet achieving accuracies of 80.84% and 92.42% in Alzheimer's and Parkinson's disease detection, respectively. The research not only presents innovative generative architectures but also emphasizes the importance of classification metrics, providing valuable insights into the synthesis and detection of neurodegenerative diseases through advanced computational techniques.

INDEX TERMS Diffusion models, data augmentation, biomedical deep learning, dementia, generative adversarial networks.

I. INTRODUCTION

Early identification of Parkinson's and accurate predictive analysis in the associated domain is a crucial task and depicts substantial societal value [1]. Through the advent of computational methodologies and the availability of scalable computing sources, deep learning architectures and methodologies have become highly prominent, and any tasks that were considered difficult to deploy have become more accessible [2]. The related set of approaches can be

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

understood as a subset of machine learning, and artificial intelligence, by an extended analogy. These approaches or architectures have been used for a diverse set of domains and have depicted a very positive utility [3], [4]. An interesting and sought-after problem in the field of predictive models is generating synthetic realistic data and developing intelligent architectures that are feasible to deploy and can remedy the scarcity of datasets and the need for higher data samples for training and testing [5].

The current standard architectures or baselines being utilized in the modality of domains are Generative Adversarial Networks (GAN) [5]. The famous paper presented a novel training mechanism, through a Generator and a Discriminator, where after substantial training, the generator could be used as a standalone model for predicting newer data samples from noise [5]. Many different versions and subsidiary algorithms of GANs have been developed [6], [7], [8], and the associated utility can be considered the most widespread across all the other generative pipelines. Another approach prominent in the recent work is Variational Autoencoders (VAE) [9], this methodology provides a probabilistic manner for describing an observation in latent space. Thus, rather than building an encoder that outputs a single value to describe each latent state attribute, this approach formulates our encoder to describe a probability distribution for each latent attribute [10]. This implementation can be leveraged for generating synthetic data and the methodology has been used for the aforementioned purpose for a multitude of research statements.

The current literature also proposes the use of Diffusion Models [11], the methodology is inspired by non-equilibrium thermodynamics. A Markov chain is a mathematical structure used to model a system of diffusion steps, which are designed to add random noise to data. This noise is then used to assist in the process of learning how to reverse the diffusion and create the desired data samples from the noise. [11]. Unlike VAE, diffusion models are learned with a fixed procedure and the latent variable has high dimensionality [11]. Diffusion Models are relatively recent, however, they represent a strong intuition and substantial uses. This paper offers a thorough comparison between these three generative pipelines by considering various encoder options and experimental settings for generating usable MRI scans, which are related to dementia with an emphasis on Alzheimer's and Parkinson's [12]. The paper is further divided into multiple sections, where the Introduction is followed by the Related and modern literature, the methodology, the obtained empirical result, and the analysis. This is further followed by the conclusion and the future implications and societal impact of this research. The gist of the paper is depicted in Figure 1, wherein three generative pipelines were employed. In the topmost pipeline, various deep learning models such as GAN, Data Efficient GAN, Diffusion models, and VAE were trained on the Alzheimer dataset to generate synthetic images. These generated images were then utilized to train DenseNet and ResNet models to detect Alzheimer's disease. Similarly, in the bottommost pipeline, the models were trained on the Parkinson dataset for Parkinson's disease detection. Notably, in the middle pipeline, the deep learning models originally trained on the Alzheimer's dataset underwent fine-tuning on the Parkinson's dataset due to limited data availability for Parkinson's disease.

II. RELATED LITERATURE

This section extensively elaborates upon the current standards, the existing use cases, and other related work in modern literature and recent articles. The research presented in the article [13] offered a novel deep learning pipeline centric on the Wasserstein GAN-based algorithm for augmenting Chest X-ray images and enhancing the automated COVID-19 detection pipeline. The paper [14] also developed a generative pipeline about the Wasserstein GAN paradigm system and altered it toward speech enhancement and robust speech synthesis. The article [15] depicted the role of generative models, specifically the GANs for dimensionality reduction, and associated the practice with the Principal Component Analysis (PCA) methodology. GAN-based approaches have also been used for creating Image translation techniques, with the primarily used architecture being the Pix2Pix model [16]. The paper [11] depicted the utility of a Cascaded Diffusion model specifically for high-fidelity image generation, showed thorough experimentation, and proposed the diffusion model paradigm for the aforementioned statement. The article [17] introduces a new approach to medical image synthesis called the denoising diffusion model for medical image synthesis (DDMM-Synth). The primary goal of this framework is to decrease radiation exposure during computed tomography (CT) scans while addressing several issues such as incomplete information, limitations of learned prior models, and errors in translating medical modalities from magnetic resonance imaging (MRI) to CT. This approach [17] can be used to optimize the projection number of CT for specific clinical applications and can significantly enhance results for cases with noise. In the paper [18], a novel method named Light and Effective Generative Adversarial Network (LEGAN) is introduced to generate high-quality medical images in a lightweight manner. The primary aim of this approach is to improve clinical diagnosis by providing additional pathological information. Conventional methods for medical image synthesis have certain limitations such as a lack of sensitivity towards local tissue details and the requirement of significant computing resources. To overcome these challenges, LEGAN employs a two-stage generative adversarial network with a coarse-to-fine paradigm, inspired by the painting process of humans [18]. This approach ensures the sensitivity of the model towards local information of medical images. Additionally, a low-rank convolutional layer is used to reduce model redundancy, which involves utilizing principal components of full-rank convolutional kernels [18]. The paper [19] illustrates the advantages and applications of GAN models in medical image fusion

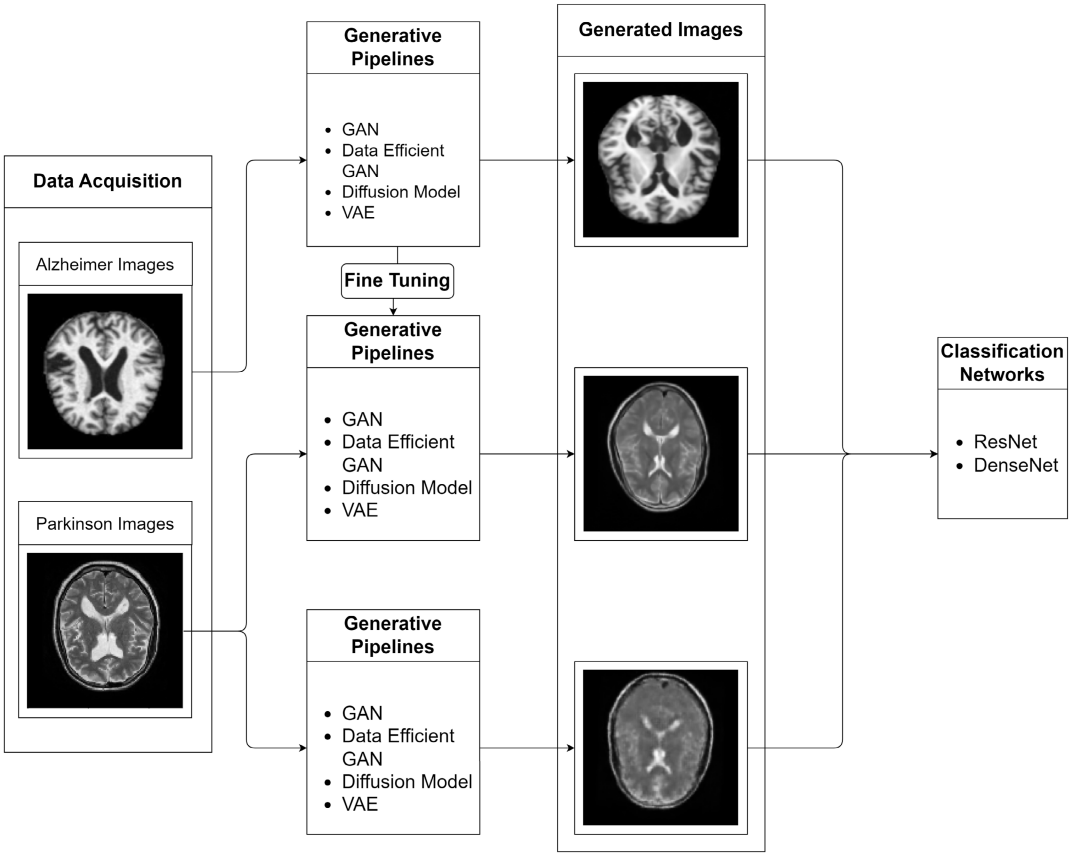

FIGURE 1. The Taxonomy associated with this paper.

and discusses the challenges faced by GANs and their applications in this field. The paper [20] proposes a novel Quantized Evolutionary Gradient Aware Multiobjective Cyclic GAN (QEMCGAN) to address issues faced by GANs in medical image-to-image translation. QEMCGAN utilizes evolutionary computation, multiobjective optimization, and an intelligent selection scheme, achieving visually realistic images and preserving crucial features efficiently, even with reduced model size. The study [21] described in the paper proposes an end-to-end GAN-based model (U-Patch GAN) for the self-supervised fusion of multimodal brain images, resulting in improved fusion quality. The study [22] proposes FS-GAN, an unpaired learning approach with a fuzzy discriminator and self-guided modules for medical image enhancement, yielding improved texture structure, uniform illumination, and superior performance in downstream tasks. The research presented in the paper [23] showcases "ImUnity," an innovative 2.5D deep-learning model for efficient MR image harmonization, outperforming existing methods and enabling cross-center population studies. The paper [24] introduces a diffusion model-based approach that leverages stochastic sampling to generate a distribution of segmentation masks, outperforming existing state-of-the-art networks in accuracy and preserving natural variation in medical image segmentation. The study [25] investigated diffusion models' performance in generating brain tumor images revealing their tendency to memorize training images, particularly for small datasets. The study [26] provides a comprehensive review of deep generative models, including variational autoencoders, generative adversarial networks, and diffusion models, for medical image augmentation, highlighting their potential for enhancing deep learning algorithms in medical image analysis. The paper [27] introduces a 2D medical image synthesis framework based on a diffusion model using a Swin-transformer-based network, addressing limited training datasets for AI models. It demonstrates the potential of generating high-quality medical images for various modalities to supplement existing training sets for AI model deployment. The paper [28] introduces FreMAE, a new self-supervised Masked Image Modeling (MIM) framework, which leverages frequency domain perspective to incorporate global structural information and local details for medical image segmentation. FreMAE outperforms previous MIM methods on benchmark datasets, demonstrating consistent improvements over baselines. This work marks the first exploration of MIM with Fourier Transform in medical image segmentation. The research [29] proposes the Attri-VAE, a novel VAE-based approach with attribute regularization to improve interpretability and disentanglement of clinical and medical imaging attributes in the generated latent space. The

TABLE 1. Current literature pertaining to medical imaging and unconditional synthesis [30].

Methods	Modalities	Description
MRI	SCGAN	A Semi-supervised method involving two stages is used for the detection of missing features from cardiac MR images, here, SC stands for Semi-Coupled [31].
CT	PGGAN	Segmenting, mapping, and augmenting brain images using joint learning [32].
MRI	LAPGAN	Generating synthetic MR brain image slices [33].
MRI	DCGAN	Leveraging manifold learning methodology for image synthesis and denoising [34].
CT	DCGAN	The mentioned approach synthesizes liver lesions by using DCGAN for each class, then classifying different classes of lesions [35].
X-ray	DCGAN	Pathology classification and augmentation of Chest X-Rays [36].
Dermo	LAPGAN	High-resolution skin lesion image generation [37].
X-ray	DCGAN	Abnormal cardiac classification using Semi-supervised learning [38].

model outperforms state-of-the-art VAE approaches, providing a valuable trade-off between reconstruction fidelity and interpretability. To better understand the generative potential of these methodologies in this domain the below-mentioned table offers a condensed description for the same and it also accurately depicts the varied data sources and use cases that have been leveraged.

The research article [39] presented the use of the diffusion model paradigm for symbolic music generation, with a superlative and high-fidelity result. The paper [40] functioned on creating a deep learning pipeline for super-resolution in magnetic resonance images of the brain under clinical protocols and utilized the attention paradigm. The research article [41] proposed a GAN-based framework for information encoding in acoustic data for modeling lexical learning. The article [42] showcased a novel GAN-based methodology for augmenting data pertaining to rare liver cancers and depicted an exemplary performance across various evaluation metrics. The paper [43] presented a novel Variational Auto-encoder based architecture for generating 3-D brain MRI scans and the model can be understood as a Multi-scale Metamorphic VAE. The article [44] proposed a diffusion model-based deep learning pipeline for generating realistic histopathology images and achieved a stellar performance across multiple performance metrics. The paper [45] proposed a novel generative pipeline based on the GAN methodology for 3D image transformation in the domain of medicine and biology and achieved state-of-the-art results. The research presented in the paper [46] presented a novel learning pipeline for the Variational Autoencoder methodology by involving sparse encoding, the proposed methodology provided a superlative performance and sufficient utility. The article [47] presented a novel Variational Autoencoder termed the Radon Sobolev Variational Autoencoder. The famous paper [48] extended the VAE approach to the paradigm of graph-centric computation and a potential generative algorithm for graph-based data samples.

III. METHODOLOGY

This section offers thorough information considering the tested methodologies and the conducted experiments. The paper leverages the famous generative pipelines and also utilizes the ResNet architecture [49] primarily for further exploiting the generated synthetic images. The section is further divided into multiple sub-sections that pertain to the generative pipelines and the classification network. This is followed by the experimental setting and the obtained empirical results.

A. GENERATIVE PIPELINES

The paper explores three main augmentative pipelines that fall under the umbrella term of deep learning. These are learning architectures that after being trained on a multitude of images, are capable of successfully generating novel and unseen synthetic images that fall under the specified data distributions while depicting high fidelity [5]. The experimented deep learning pipelines include Generative Adversarial Networks, Diffusion models, and Variational Autoencoders. For the former category, two different GAN pipelines (abbreviated as GAN1 throughout the paper), the standard architecture, and the relatively recent data-efficient GANs [50] (abbreviated as GAN2 throughout the paper) are further analyzed.

1) GENERATIVE ADVERSARIAL NETWORKS

A game-theoretic scenario in which the generator network must compete with an attacker serves as the foundation for generative adversarial networks. Samples are generated directly by the generator network [35]. The discriminator network, which is its rival, makes an effort to differentiate between samples taken from the training data and those taken from the generator. The generator model takes a fixed-length random vector as input, which is generated using a Gaussian distribution, and produces a sample in the domain. This vector is used as a seed for the generative process [35].

After training the data distribution can be compressed into a multidimensional vector space. This type of vector space is called a latent space and consists of latent variables, which are important to the domain but cannot be observed directly. The discriminator model then classifies an example from the domain as real or fake based on its input, whether it is real or synthetic (created artificially) [51]. The real data samples originate from the training dataset, while

the generator model is responsible for the production of generated data. The discriminator model performs the task of classifying real and generated data. After the training phase is complete, the discriminator is no longer required and only the generator is used for producing images. The generator model produces images that are indistinguishable from real data, accurately capturing the nuances of the data it is trained on. The generated images are of high quality, containing all essential details, making them indistinguishable from the real data. This concept of a Generative Adversarial Network (GAN) is illustrated in the following Figure 2. The inherent functioning of the GAN methodology can be understood better by the below-mentioned equations [52]. In these equations, the variables represent different components and loss functions used in the GAN framework. Equation 1 defines the conditional GAN loss, denoted as $\mathcal{L}_{cGAN}(G, D)$. In this equation, G represents the generator, responsible for generating synthetic images, while D represents the discriminator, which tries to differentiate between real and generated images. The loss function comprises two parts. The first term seeks to maximize the probability that the discriminator correctly classifies real image pairs (a and b), and the second term aims to maximize the probability that the discriminator correctly classifies pairs of real images a and the images generated by G from some input c. Equation 2 introduces the L1 loss function, denoted as $\mathcal{L}_{L1}(G)$. The purpose of this loss function is to measure the difference between the real target image (y) and the image generated by the generator G using input a and c. The absolute difference (L1 norm) between these images is calculated and used to quantify the dissimilarity between them. Equation 3 presents the combined objective of the GAN framework. The goal here is to optimize the generator G to minimize the conditional GAN loss $\mathcal{L}_{cGAN}(G, D)$ along with the L1 loss $\mathcal{L}_{L1}(G)$. The parameter λ allows you to control the importance of the L1 loss relative to the GAN loss during the optimization process. Finding the optimal G that minimizes this combined function while simultaneously maximizing the performance of the discriminator D results in an effective GAN model capable of producing high-quality generated images.

$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{a,b}[\log D(a, b)] + \mathbb{E}_{a,c}[\log(1 - D(a, G(a, c))] \quad (1)$$

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{a,b,c}[\|y - G(a, c)\|_1] \quad (2)$$

$$G^* = \arg\min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G) \quad (3)$$

The second GAN pipeline used in this paper can be summarised as a Data-efficient GAN with Adaptive Discriminator Augmentation. The original article proposed Differentiable Augmentation (DiffAugment) is a relatively simple methodology that functioned to improve the data efficiency of GANs by imposing various types of differentiable augmentations on real and fake samples [53]. The conventional attempts to directly augment the training data and manipulate the distribution of real images, yield substantially little benefit. The algorithm, which is abbreviated as DiffAugment enables the adoption of the differentiable augmentation pipeline for the generated samples, further stabilizing effective training, and an overall better convergence [54]. The experiments available in the original paper demonstrated a consistent gain of the aforementioned method over a variety of baseline GAN architectures and loss functions for both unconditional and class-conditional generation.

2) VARIATIONAL AUTOENCODERS

Variational Autoencoder (VAE) are a type of generative model used in unsupervised learning. They consist of two main parts: an encoder and a decoder. The encoder takes input data and compresses it into a latent space representation (lower-dimensional representation). The decoder then reconstructs the original data from this representation. VAEs are trained to generate new data points that resemble the input data distribution. They are known for their ability to generate diverse and realistic outputs [55].

An encoder creates a single value to represent each dimension in the latent space of a VAE. This generates a probability distribution for each latent attribute, allowing for a probabilistic representation of the observation [56]. By constructing the encoder model to generate a range of potential values (a statistical distribution) from which to randomly sample the decoder model, a continuous and uniform latent space representation is enforced [57]. When sampling from the latent distribution, the decoder must be able to accurately reconstruct the input with values in latent space that are similar to one another. The process is illustrated in the following Figure 3.

A continuous, smooth latent space representation is essentially enforced by building the encoder model to generate a range of potential values that can also be perceived as a statistical distribution from which random picks are fed into the decoder model. It can be anticipated that the decoder model will be able to precisely reconstruct the input for any sampling of the latent distributions. This results in values that are close to one another in latent space that corresponds to reconstructions that are quite similar [56]. The characteristic equations for this methodology are mentioned below where $p_\theta(a)$ is the probability of observing data a given model parameters θ, $\log p_\theta(a)$ is the logarithm of this probability, $q_\phi(c|a)$ is the conditional distribution of latent variables c given the data a parameterized by ϕ, $\mathbb{E}_{q_\phi(c|a)}[\log[\frac{p_\theta(a,c)}{q_\phi(c|a)}]]$ is the expected log-likelihood term representing the expected log-ratio between joint and posterior distributions over c, and $D_{KL}(q_\phi(c|a)||p_\theta(c|a))$ is the Kullback-Leibler divergence measuring the difference between the posterior and prior distributions over c. The equation is used in variational inference to approximate complex posterior distributions by optimizing model parameters θ and ϕ to maximize the log-likelihood of observed data a while minimizing the divergence between the posterior and prior distributions over

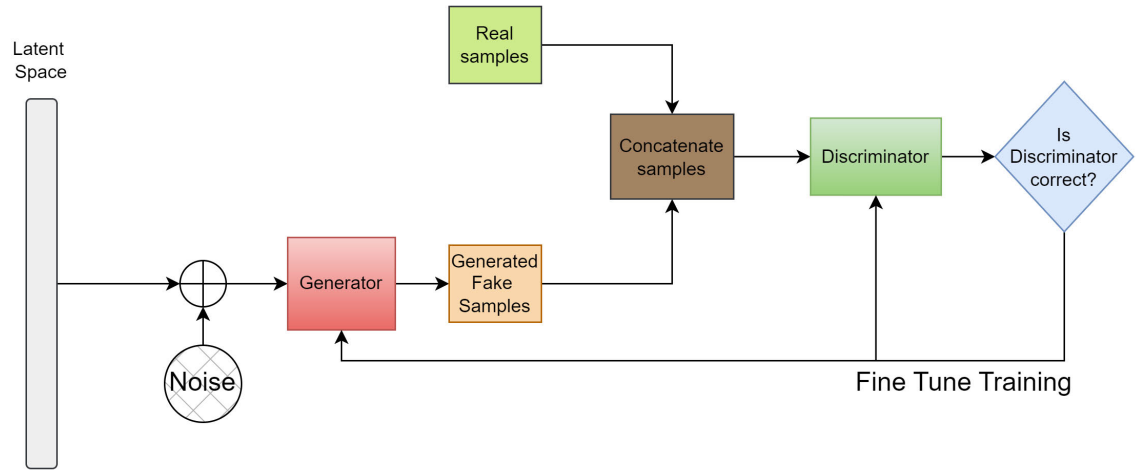

FIGURE 2. The inherent working of Generative Adversarial Networks [54].

latent variables c [56]:

$$\log p_\theta(a) == \mathbb{E}_{q_\phi(c|a)}[\log[\frac{p_\theta(a,c)}{q_\phi(c|a)}]] + D_{KL}(q_\phi(c|a)||p_\theta(c|a)) \tag{4}$$

The encoder component of a Variational Autoencoder (VAE) will not directly output values for the latent state, as is the case with a traditional autoencoder. Instead, it will generate parameters describing the distribution of each dimension of the latent space [56]. Two vectors characterizing the mean and variance of the latent state distributions will be outputted because it is presumed that the prior has a normal distribution.

3) DIFFUSION MODELS

Diffusion models are probabilistic generative models that aim to model complex data distributions. They work by iteratively refining a probability distribution. Unlike many traditional generative models that sample data directly from a simple distribution (like a Gaussian distribution), diffusion models start with a simple distribution and iteratively transition it into a more complex distribution using a diffusion process. They are particularly effective for generating high-quality, diverse samples from complex data distributions [58]. The paradigm is divided into many sub-architectures and after a thorough literature survey, the authors have leveraged the Denoising Diffusion Implicit Model (DDIM) methodology [59]. DDIM serves as an upgrade to the relatively conventional Denoising Diffusion Probabilistic Models (DDPM).

Without adversarial training, DDPMs have generated high-quality images. To produce a sample, they must, however, simulate a Markov chain for a lengthy period. With the same training process as DDPMs, DDIMs use a more effective class of iterative implicit probabilistic models to speed up sampling [59]. The generating process in DDPMs is described as the Markovian diffusion process inverted. The technique creates a class of non-Markovian diffusion processes that achieve the same training goal but whose reverse process may be sampled considerably more quickly [59]. Further allowing a trade-off computation for sample quality, the paper empirically showed that DDIMs can produce high-quality samples far faster than DDPMs in terms of wall-clock time. They can also interpolate semantically relevant images directly in the latent space [59]. The inherent functioning of the diffusion model methodology as a graphical description is available below in Figure 4.

B. CLASSIFICATION ARCHITECTURES

This subsection elaborates on the various classification encoders or architectures used throughout this study to better understand or validate the generated samples of the tested generative methodologies. This paper primarily uses the ResNet architecture and also tests the DenseNet methodology [60] to have a thorough and unbiased study.

1) RESNET

This subsection provides a summarised understanding of ResNets as defined in the original paper [49]. The Residual Networks more commonly known as ResNets is one of the proposed encoders that we are going to use. The uniqueness that ResNets offer, over plain networks, is that they address the degradation problem that was exposed when the overly deep networks started to converge. The ResNets introduced identity mapping which meant that the input from a previous layer was taken and passed to another layer as a shortcut. Mostly the 34-layer and 18-layer ResNets are used as they fetch less error and better accuracy as compared to their plain competitors. The 34-layer ResNet displays fairly decreased training error and handles the degradation problem that is observed in its plain compeer, thus high accuracy is gained from increased depths. Not to overlook the fact that the 18-layer ResNet also fetched a better accuracy than its plain compeer, the 18-layer ResNet was able to achieve convergence faster and obtain good solutions on smaller

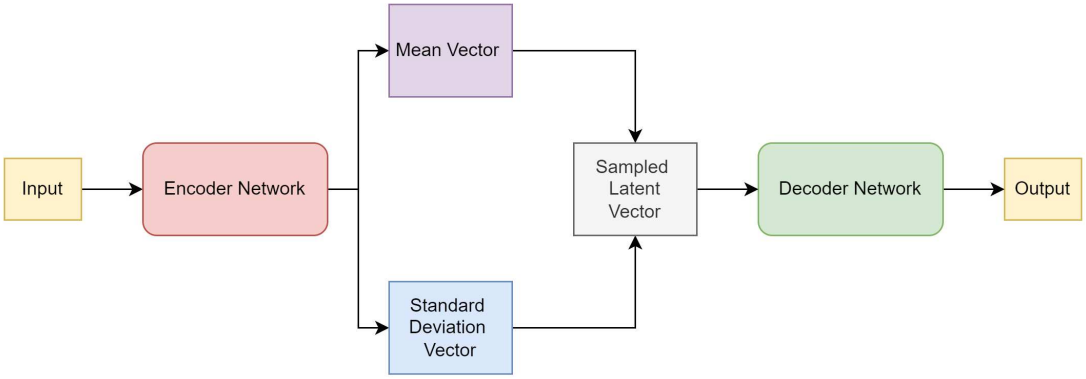

FIGURE 3. Pictorial representation corresponding to the VAE computing paradigm [56].

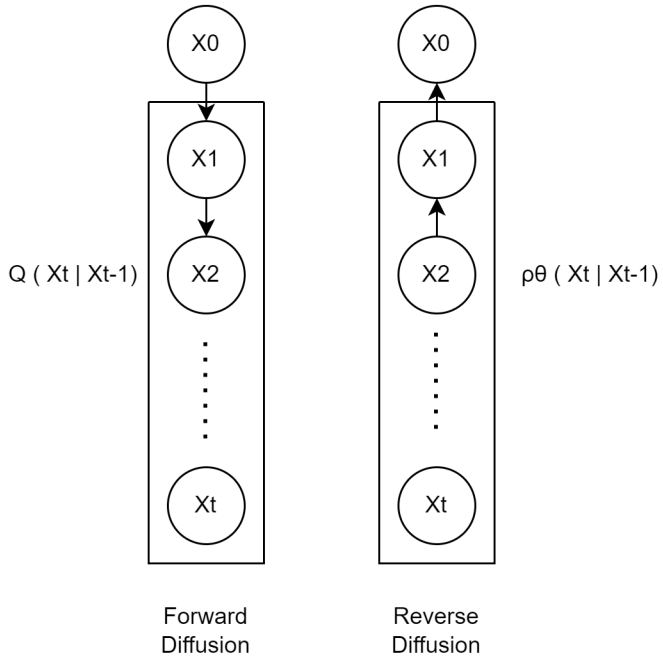

FIGURE 4. Information flow associated with the Denoising Diffusion Models [59].

datasets [49]. For the experiments in this paper, the ResNet18 variant of residual networks is used.

2) DENSENET

The Dense Convolution Network, also known as DenseNet, is another variety of encoders used in the proposed architecture. The advantage of using this particular type of network is that each of its layers gathers supplementary inputs from all of the layers before it. The data is concatenated so essentially every single layer obtains the cumulative intelligence of all the antecedent layers [60]. Therefore when each layer obtains feature maps from the previous layers it makes the complete network more compressed, which means that the total channels will be fewer. The contrasting detail that separates the DenseNets from the ResNets is that they use the parameters more dexterously. Outwardly both networks are quite similar the only major difference is that the DenseNets concatenate the inputs while the process of summation is what happens in ResNets. Although this seems like a small adjustment it brings out a rather considerable change in behavior between them both [60]. Adding to that fact DenseNets crave extensively less number of parameters and computational power to obtain highly accurate and cutting-edge performances and results with better accuracy can be achieved when the hyperparameters are tuned with attention to detail. In this regard we will be using the DenseNet-161 model, it's one of the high-accuracy models of the DenseNet group, and the size of this model is considerably larger than its other variants at 100MB. The model used in this paper is a DenseNet-161 [60].

C. DATASET

This paper primarily leverages a publicly available Alzheimer's MRI dataset comprising a total of 6400 MRI images as depicted in paper [61]. The dataset consists of 4 classes, Mild Demented, Moderate Demented, Very Mild Demented, and Non-Demented [61]. The former three categories are combined to generate the dementia class, and an 80-20 stratified train/test split is created. A subset of the PPMI (Parkinson's Progression Markers Initiative) dataset consisting of 128 scans is used for analyzing the proposed algorithms' impact on predicting a Parkinson's class [62]. The paper primarily emphasizes Alzheimer's due to the significantly higher data samples, and experiments on Parkinson's are also presented to assess the robustness and general implications of the proposed approaches. Due to images from different machines and data acquisition technologies, there were a lot of image biases and a usable generative model was not obtained. To have images from the same domain and to have a feasible image generation experiment a single category was focused on and a total of 43 Parkinsons-based images (PD) and 18 images based on healthy patients (HC) were obtained from the ''Axial Dual Turbo Spin Echo'' (AX DUAL TSE) category of the PPMI dataset. AX DUAL TSE is an imaging sequence that provides detailed images of the brain in the axial plane using turbo spin echo techniques, which are commonly used in clinical MRI for T2-weighted imaging.

IV. RESULTS AND DISCUSSION

This section elaborates on the experimental setting and obtained results of the proposed architectures. Each model permutation is trained in an identical experimental setting to maintain the unbiased nature of this study. Every associated hyperparameter and attribute is thoroughly assessed to maintain the scientific integrity of this study. For each generative pipeline, 500 images are generated for the normal and dementia classes which are tested for their FID scores (Frechet Inception Distance) [63]. FID scores are a measure used to evaluate the quality of generated images compared to real images. They quantify the similarity between two sets of images: the set of real images and the set of generated images. FID utilizes features extracted from a pre-trained deep neural network, often InceptionV3, to capture the statistics of real and generated images [63]. The features extracted from the intermediate layers of the neural network are used to represent both real and generated images. FID calculates the Fréchet distance between the multivariate Gaussian distributions of feature representations from real and generated images. A lower FID score indicates higher similarity between the distributions of real and generated images, implying better quality and diversity in the generated images [63]. A lower score suggests that the images have more similar or comparable statistics [63]. The inception score helps in estimating the variety and the ability of the model to generate newer images that distinctly look like a possible entity [63]. The trained models are fine-tuned on the Parkinson's dataset and compared with a standalone training setting to further assess the tested pipelines' capabilities. The characteristic equations that pertain to these performance measures are mentioned below. The equation 5 and 6 demonstrate the FID (Fréchet Inception Distance) calculation. Here, μ and μ_w represent specific mean values, and Σ and Σ_w are covariance matrices. The equation 7 demonstrates the Inception score, which involves KL-Divergence (KL). In this context, $\hat{A}$ denotes the estimated probability distribution, A represents the true probability distribution, and the sum is taken over the 'M' categories. These equations serve as integral components for assessing the respective FID score and Inception score in performance evaluation [63].

$$\text{FD} = |\mu - \mu_w|^2 + \text{tr}(\Sigma + \Sigma_w - 2(\Sigma\Sigma_w)^{1/2}) \quad (5)$$

$$FD = |\mu - \mu_w|^2 + \text{tr}(\Sigma + \Sigma_w - 2(\Sigma\Sigma_w)^{\frac{1}{2}}) \quad (6)$$

$$KL(\hat{A}||A) = sum_{c=1}^{M}\hat{A}_c \log\frac{\hat{A}_c}{A_c} \quad (7)$$

TABLE 2. Models' training and testing time on a single image.

Model	Training Time (s)	Testing Time (s)
GAN1	0.003	0.080
GAN2	0.001	0.017
DM	0.014	0.005
VAE	0.0007	0.0035

The below-mentioned tables showcase the performance of the generative pipelines and their temporal characteristics for the Alzheimer's dataset.

From the above-mentioned tables 2 and 3, it can be observed that the best-performing model for the Non-Dementia related sub experiments was the Diffusion model which showcased the lowest FID scores of 92.46. In the context of the Dementia-related sub-experiments, the GAN2 model demonstrated the lowest Frechet Inception Distance (FID) scores, achieving a value of 177.53. However, when considering the Inception Score (IS) value, the diffusion model outperformed other models. It showed an increase of 4.55 units more than the baseline GAN for the dementia images and an increase of 2.50 units more than the baseline GAN for the non-dementia images. In terms of FID, the VAE model produced the poorest results, with FID scores of 333.27 for dementia images and 336.88 for non-dementia images compared to other models. Regarding the temporal characteristics, the VAE methodology performed best, demonstrating a substantial decrease in both training and testing times by 0.0023 seconds and 0.0765 seconds, respectively, compared to the baseline GAN. On the other hand, diffusion models showcased the highest training time of 0.014 seconds, while the baseline GAN exhibited the highest testing time of 0.080 seconds. All of these models were used to generate a data bank of generated images which was coupled with the already existing training set to gauge the classification enhancements of these methodologies. This paper leverages four primary performance metrics for

TABLE 3. Performance measures pertaining to generative pipelines and the Alzheimer's dataset.

Model	Category	FID Score	Inception Score	
			Average	Standard Deviation
GAN1	Dementia	212.48	4.07	0.074
	Non-Dementia	200.42	4.28	0.136
GAN2	Dementia	178.53	7.33	0.374
	Non-Dementia	243.51	4.93	0.008
Diffusion	Dementia	186.36	8.62	0.095
	Non-Dementia	92.46	6.68	0.082
VAE	Dementia	333.27	4.87	0.024
	Non-Dementia	336.88	3.76	0.022

TABLE 4. Experiments pertaining to the classification architectures and the Alzheimers centric generative pipelines.

Model	Accuracy	Precision	Recall	F1-Score
ResNet	74.35	78.63	66.82	72.25
ResNet + GAN1	74.74	74.15	75.89	75.01
ResNet + GAN2	78.34	78.01	78.88	78.44
ResNet + DM	77.01	73.40	84.66	78.63
ResNet + VAE	75.21	72.93	80.12	76.36
DenseNet	79.35	80.38	77.62	78.98
DenseNet + GAN1	75.52	71.39	85.13	77.65
DenseNet + GAN2	79.59	**86.06**	70.57	77.55
DenseNet + DM	**80.84**	77.82	**86.22**	**81.81**
DenseNet + VAE	78.42	75.31	84.50	79.64

TABLE 5. Performance measures pertaining to generative pipelines and the Parkinson's dataset.

Model	Category	FID Score	Inception Score	
			Average	Standard Deviation
Gan1	PD	124.59	10.57	0.101
	HC	184.79	15.62	0.483
Gan2	PD	102.71	5.32	0.148
	HC	129.77	1.50	0.001
Diffusion	PD	115.02	10.46	0.0001
	HC	139.34	10.76	0.271
VAE	PD	184.04	25.43	0.73
	HC	195.66	18.89	1.08

TABLE 6. PPMI classification Result.

Model	Accuracy	Precision	Recall	F1-Score
ResNet	69.23	69.24	**99.13**	81.53
ResNet + GAN1	68.99	69.29	99.10	81.55
ResNet + GAN2	76.20	76.92	92.36	84.31
ResNet + DM	84.97	90.33	87.67	88.98
ResNet + VAE	82.09	84.60	90.62	87.51
DenseNet	77.16	77.89	93.57	85.01
DenseNet + GAN1	76.33	78.06	93.57	85.01
DenseNet + GAN2	79.80	81.19	92.18	86.34
DenseNet + DM	**92.42**	**96.21**	92.70	**94.42**
DenseNet + VAE	84.01	88.25	88.71	88.48

TABLE 7. Performance measures pertaining to generative pipelines and finetuning for Parkinson's.

Model	Category	FID Score	Inception Score	
			Average	Standard Deviation
Gan1	PD	166.64	12.91	0.191
	HC	260.08	8.62	0.013
Gan2	PD	121.81	6.66	0.083
	HC	166.61	4.81	0.116
Diffusion	PD	109.48	12.84	0.035
	HC	106.35	9.17	0.220
VAE	PD	213.36	11.64	0.297
	HC	212.75	1.06	3.4987977e-05

the classification networks, percentage accuracy, precision, recall, and the F-1 score. All of these metrics can be understood better by the below-mentioned equations, here, TP stands for True Positive, FP for False Positive, TN for

TABLE 8. PPMI + Fine Tuning Classification Results.

Model	Accuracy	Precision	Recall	F1-Score
ResNet + GAN1 + FT	79.68	81.94	90.62	86.06
ResNet + GAN2 + FT	76.80	83.07	83.50	83.29
ResNet + DM + FT	85.21	83.65	**97.74**	90.15
ResNet + VAE+ FT	68.99	69.23	72.46	70.65
DenseNet + GAN1+ FT	80.16	80.53	94.09	86.78
DenseNet + GAN2+ FT	87.01	93.01	87.84	90.35
DenseNet + DM+ FT	**90.50**	**94.74**	91.31	**93.01**
DenseNet + VAE+ FT	78.84	82.15	88.71	85.30

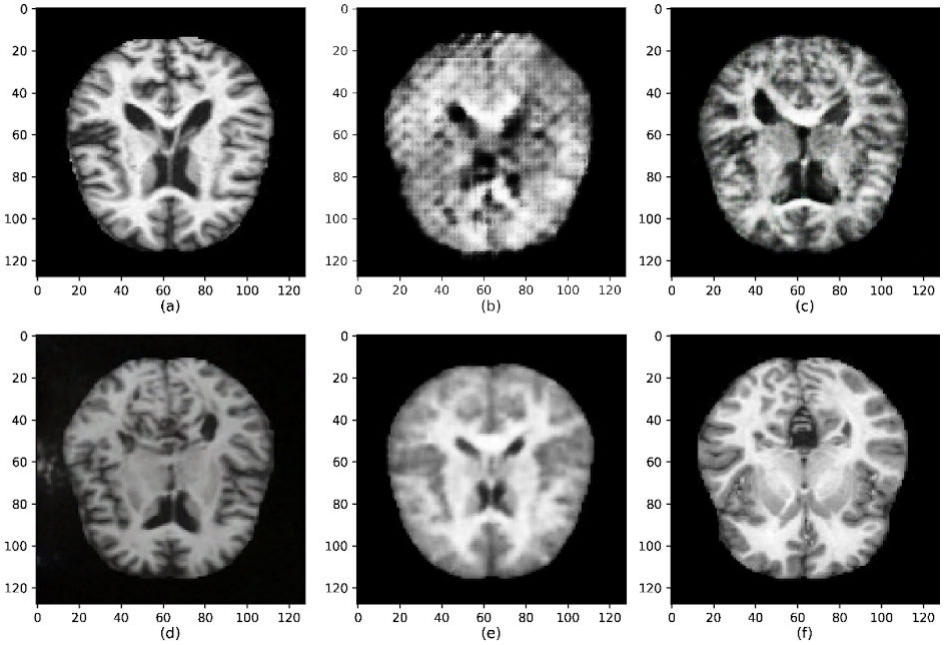

FIGURE 5. The generated images and the ground truth, here (a) and (f) represent the actual scans of a non-demented person. The images (b)-(e) are the GAN1, GAN2, DDIM, and VAE respectively.

True Negative, and FN for False Negative [64].

$$Accuracy = \frac{TN + TP}{TP + TN + FP + FN} \tag{8}$$

$$Precision = \frac{TP}{FP + TP} \tag{9}$$

$$Recall = \frac{TP}{FN + TP} \tag{10}$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FN + FP} \tag{11}$$

From table 4, we can successfully infer that the use of generative pipelines is justified and a clear increase in performance metrics can be seen for the majority of experiments. By considering accuracy as the primary performance metric the best-performing generative pipeline was obtained as the diffusion model functionality by achieving an accuracy of 80.84% when DenseNet was used as the classifier. The next best-performing generative pipeline would be the data-efficient GANs as they have shown s similar results for the ResNet set of experiments and have also showcased a better precision score of 86.06% for the experiments pertaining to the DenseNet as a backbone. Due to a relatively smaller dataset for the Parkinson's experiment, we have also employed a fine-tuning regiment where the already trained generative models on the Alzheimer's dataset were fine-tuned on the Parkinson's dataset. For both of these generative experiments, the classification models were trained to accurately understand the utility of finetuning in generative models and also the applicability of the tested strategies in the domain of Parkinson's and Alzheimer's.

A similar inference was observed for the Parkinson's set of experiments from tables 5 and 6 with Diffusion Models giving the best overall performance followed by the data-efficient GAN paradigm. The data-efficient Generative Adversarial Network (GAN2) paradigm exhibits the lowest FID scores for both the dementia (PD) and non-dementia (HC) categories, with FID scores of 102.71 and 129.77, respectively. However, upon subjecting the generated images to classifier training, the diffusion models set of experiments achieved the highest classification accuracy of 92.42%. On the other hand, the Variational Autoencoder

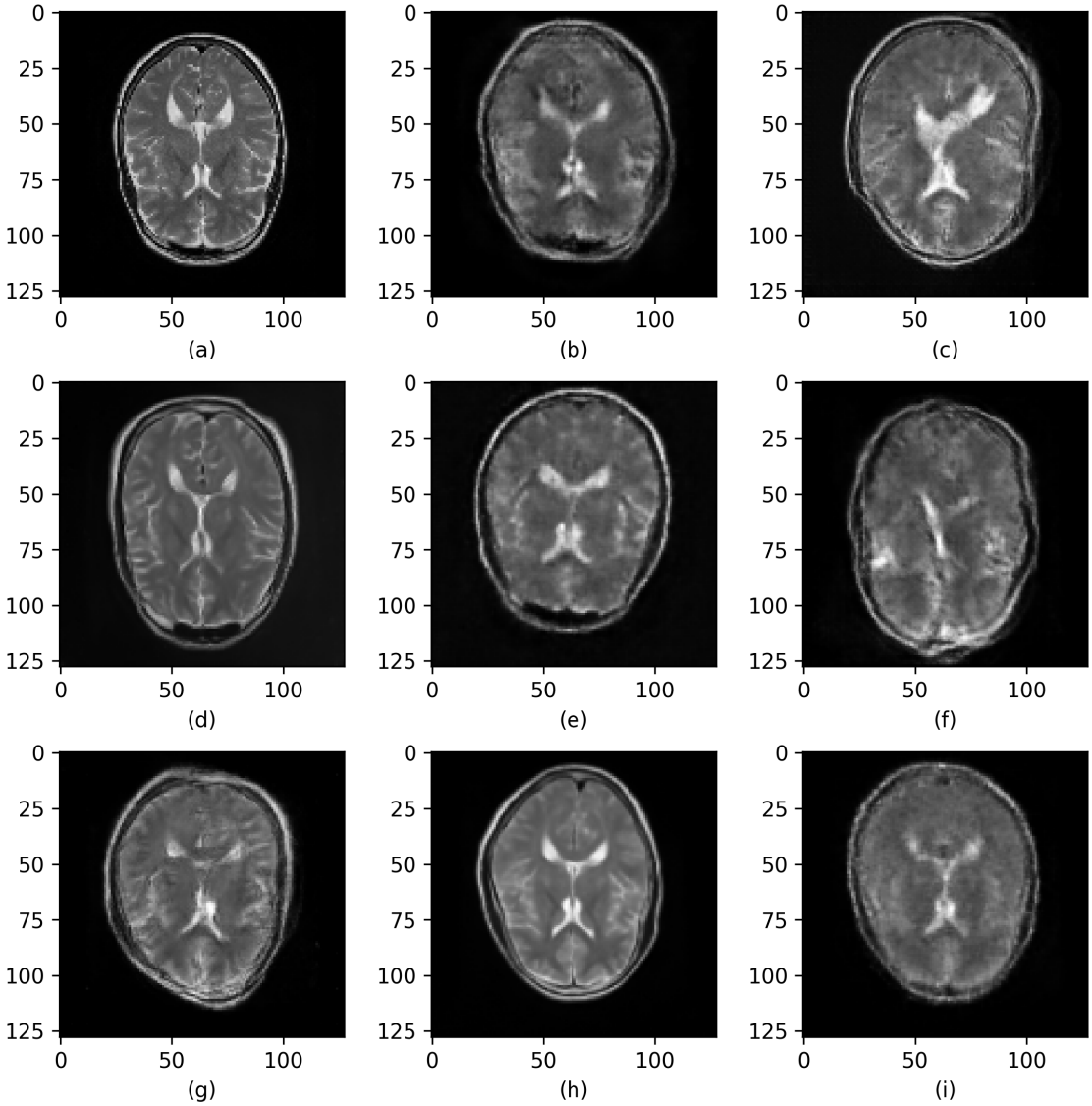

FIGURE 6. The generated images and the ground truth, here (a) represent the actual scans of a dementia person. The images (b)-(e) are the GAN1, GAN2, DDIM, and VAE respectively. The images produced by fine-tuning GAN1, GAN2, DDIM, and VAE are represented as (f) - (i).

(VAE) approach yielded a relatively poor FID score of 180.04, yet demonstrated an impressive Inception Score (IS) of 25.43 for the PD category. Consequently, it emerged as the second-best performing model when utilized for classification in conjunction with the ResNet and DenseNet sets of experiments, achieving classification accuracies of 82.09% and 84.01%, respectively. The following tables extend upon the aforementioned experiments and indicate the classification results associated with the generative pipelines. Experiments have been conducted for both the fine-tuned models and the standalone classification architectures.

From the tables 7 and 8 it can be inferred that a performance boost was obtained for the classification architectures for the fine-tuning strategy. It can be confidently said that the use or utility of augmentative approaches was thoroughly validated. A clear enhancement in the overall performance metrics was observed for the classification architectures and the methodology can be confidently proposed. The sample-generated images for each architecture and their contrast with real-world samples are mentioned below in Figure 5 for experiments pertaining to Alzheimer's and Figure 6 for the Parkinson's equivalent.

V. CONCLUSION AND FUTURE WORK

This paper aims to offer a thorough comparative study between the various available generative pipelines to remedy the scarcity of data and related predicaments. The study performed experiments on Parkinson's and Alzheimer's datasets using four deep-learning models to generate dementia and non-dementia images. GAN2 demonstrated superior

performance in generating Alzheimer's and Parkinson's dementia images with FID scores of 178.53 and 102.71, respectively. Fine-tuning models on the Parkinson's dataset produced poorer images, increasing the baseline GAN1 FID score by 42.51. The utilization of generated images in a classification task resulted in notable enhancements across various models. Images generated by diffusion models when trained for classification outperformed GANs and VAE by yielding an accuracy of 80.84% for Alzheimer's and 92.42% for Parkinson's. Training on images generated by fine-tuning has been found to increase classification accuracy in the majority of cases. However, there are instances where poor results were observed, particularly with images generated by VAE during the fine-tuning process. When the classifier used was DenseNet, the accuracy decreased by 5.17%, and with ResNet, the accuracy decreased by 13.1%. The experiments also indicate that specialized provisions for Data efficiency can be introduced in the denoising diffusion paradigm which should result in a substantial performance increase. For a specialized associated task emphasis can also be given to the temporal characteristics for obtaining a usable computational power and efficiency tradeoff. In the future, the authors would like to emphasize the different domains associated with data acquisition of MRI scans and work towards creating a robust processing pipeline.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

[1] K. R. Varshney, "Proceedings of the 2016 ICML workshop on #data4good: Machine learning in social good applications," Jul. 2016, *arXiv:1607.02450*.

[2] A. L. Tarca, V. J. Carey, X.-W. Chen, R. Romero, and S. Drăghici, "Machine learning and its applications to biology," *PLoS Comput. Biol.*, vol. 3, no. 6, p. e116, Art. no. 2007.

[3] J. Parmar, S. Chouhan, V. Raychoudhury, and S. Rathore, "Open-world machine learning: Applications, challenges, and opportunities," *ACM Comput. Surv.*, vol. 55, no. 10, pp. 1–37, Oct. 2023.

[4] R. Fu, A. Kundu, N. Mitsakakis, T. Elton-Marshall, W. Wang, S. Hill, S. J. Bondy, H. Hamilton, P. Selby, R. Schwartz, and M. O. Chaiton, "Machine learning applications in tobacco research: A scoping review," *Tobacco Control*, vol. 32, no. 1, pp. 99–109, Jan. 2023.

[5] B. Ghojogh, A. Ghodsi, F. Karray, and M. Crowley, "Generative adversarial networks and adversarial autoencoders: Tutorial and survey," Nov. 2021, *arXiv:2111.13282*.

[6] J. Jam, C. Kendrick, V. Drouard, K. Walker, G.-S. Hsu, and M. Yap, "Symmetric skip connection Wasserstein GAN for high-resolution facial image inpainting," in *Proc. 16th Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl.*, 2021, pp. 1–10.

[7] S.-A. Chen, C.-L. Li, and H.-T. Lin, "A unified view of cGANs with and without classifiers," in *Proc. Adv. Neural Inf. Process. Syst.*, Nov. 2021, pp. 27566–27579.

[8] D. Utyamishev and I. Partin-Vaisband, "Late breaking results: Parallelizing net routing with cGANs," in *Proc. 58th ACM/IEEE Design Autom. Conf. (DAC)*, Dec. 2021, pp. 1372–1373.

[9] R. Zemouri, "Semi-supervised adversarial variational autoencoder," *Mach. Learn. Knowl. Extraction*, vol. 2, no. 3, p. 20, Aug. 2020.

[10] S. N. Dean and S. A. Walper, "Variational autoencoder for generation of antimicrobial peptides," *ACS Omega*, vol. 5, no. 33, pp. 20746–20754, 2020.

[11] J. Ho, C. Saharia, W. Chan, D. J. Fleet, M. Norouzi, and T. Salimans, "Cascaded diffusion models for high fidelity image generation," *J. Mach. Learn. Res.*, vol. 23, no. 47, pp. 1–33, May 2021.

[12] A. B. Tufail, Y.-K. Ma, Q.-N. Zhang, A. Khan, L. Zhao, Q. Yang, M. Adeel, R. Khan, and I. Ullah, "3D convolutional neural networks-based multiclass classification of Alzheimer's and Parkinson's diseases using PET and SPECT neuroimaging modalities," *Brain Informat.*, vol. 8, no. 1, pp. 1–9, Dec. 2021.

[13] B. Z. Hussain, I. Andleeb, M. S. Ansari, A. M. Joshi, and N. Kanwal, "Wasserstein GAN based chest X-ray dataset augmentation for deep learning models: COVID-19 detection use-case," in *Proc. 44th Annu. Int. Conf. IEEE Eng. Med. Biol.*, Jul. 2022, pp. 2058–2061.

[14] N. Adiga, Y. Pantazis, V. Tsiaras, and Y. Stylianou, "Speech enhancement for noise-robust speech synthesis using Wasserstein GAN," in *Proc. Interspeech*, 2019, pp. 1821–1825.

[15] J. Cho and C. Suh, "Wasserstein GAN can perform PCA," in *Proc. 57th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Sep. 2019, pp. 895–901.

[16] K. Yanagida, K. Gyohten, H. Ohki, and T. Takami, "Automatic characteristic line drawing generation using pix2pix," in *Proc. 11th Int. Conf. Pattern Recognit. Appl. Methods*, 2022, pp. 155–162.

[17] X. Li, K. Shang, G. Wang, and M. D. Butala, "DDMM-synth: A denoising diffusion model for cross-modal medical image synthesis with sparse-view measurement embedding," 2023, *arXiv:2303.15770*.

[18] J. Gao, W. Zhao, P. Li, W. Huang, and Z. Chen, "LEGAN: A light and effective generative adversarial network for medical image synthesis," *Comput. Biol. Med.*, vol. 148, Sep. 2022, Art. no. 105878.

[19] T. Zhou, Q. Li, H. Lu, Q. Cheng, and X. Zhang, "GAN review: Models and medical image fusion applications," *Inf. Fusion*, vol. 91, pp. 134–148, Mar. 2023.

[20] V. Bharti, B. Biswas, and K. K. Shukla, "QEMCGAN: Quantized evolutionary gradient aware multiobjective cyclic GAN for medical image translation," *IEEE J. Biomed. Health Inform.*, vol. 28, no. 3, pp. 1240–1251, 2023.

[21] C. Fan, H. Lin, and Y. Qiu, "U-patch GAN: A medical image fusion method based on GAN," *J. Digit. Imag.*, vol. 36, no. 1, pp. 339–355, Aug. 2022.

[22] Y.-F. Yu, G. Zhong, Y. Zhou, and L. Chen, "FS-GAN: Fuzzy self-guided structure retention generative adversarial network for medical image enhancement," *Inf. Sci.*, vol. 642, Sep. 2023, Art. no. 119114.

[23] S. Cackowski, E. L. Barbier, M. Dojat, and T. Christen, "ImUnity: A generalizable VAE-GAN solution for multicenter MR image harmonization," *Med. Image Anal.*, vol. 88, Aug. 2023, Art. no. 102799.

[24] A. Rahman, J. M. J. Valanarasu, I. Hacihaliloglu, and V. M. Patel, "Ambiguous medical image segmentation using diffusion models," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 11536–11546.

[25] M. U. Akbar, W. Wang, and A. Eklund, "Beware of diffusion models for synthesizing medical images—A comparison with GANs in terms of memorizing brain tumor images," 2023, *arXiv:2305.07644*.

[26] A. Kebaili, J. Lapuyade-Lahorgue, and S. Ruan, "Deep learning approaches for data augmentation in medical imaging: A review," *J. Imag.*, vol. 9, no. 4, p. 81, Apr. 2023.

[27] S. Pan, T. Wang, R. L. J. Qiu, M. Axente, C.-W. Chang, J. Peng, A. B. Patel, J. Shelton, S. A. Patel, J. Roper, and X. Yang, "2D medical image synthesis using transformer-based denoising diffusion probabilistic model," *Phys. Med. Biol.*, vol. 68, no. 10, May 2023, Art. no. 105004.

[28] W. Wang, J. Wang, C. Chen, J. Jiao, Y. Cai, S. Song, and J. Li, "FreMIM: Fourier transform meets masked image modeling for medical image segmentation," 2023, *arXiv:2304.10864*.

[29] I. Cetin, M. Stephens, O. Camara, and M. A. G. Ballester, "Attri-VAE: Attribute-based interpretable representations of medical images with variational autoencoders," *Computerized Med. Imag. Graph.*, vol. 104, Mar. 2023, Art. no. 102158.

[30] N. K. Singh and K. Raza, "Medical image generation using generative adversarial networks," May 2020, *arXiv:2005.10687*.

[31] L. Zhang, A. Gooya, and A. F. Frangi, "Semi-supervised assessment of incomplete LV coverage in cardiac MRI using generative adversarial nets," in *Simulation and Synthesis in Medical Imaging*. Berlin, Germany: Springer, 2017, pp. 61–68.

[32] C. Bowles, R. Gunn, A. Hammers, and D. Rueckert, "Modelling the progression of Alzheimer's disease in MRI using generative adversarial networks," *Proc. SPIE*, vol. 10574, pp. 397–407, Mar. 2018.

[33] F. Calimeri, A. Marzullo, C. Stamile, and G. Terracina, "Biomedical data augmentation using generative adversarial neural networks," in *Proc. Int. Conf. Artif. Neural Netw.* Cham, Switzerland: Springer, 2017, pp. 626–634.

[34] A. J. Plassard, L. T. Davis, A. T. Newton, S. M. Resnick, B. A. Landman, and C. Bermudez, "Learning implicit brain MRI manifolds with deep learning," *Proc. SPIE*, vol. 10574, pp. 408–414, Mar. 2018.

[35] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification," *Neurocomputing*, vol. 321, pp. 321–331, Dec. 2018.

[36] H. Salehinejad, S. Valaee, T. Dowdell, E. Colak, and J. Barfett, "Generalization of deep neural networks for chest pathology classification in X-rays using generative adversarial networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 990–994.

[37] C. Baur, S. Albarqouni, and N. Navab, "Generating highly realistic images of skin lesions with GANs," Sep. 2018, *arXiv:1809.01410*.

[38] A. Madani, M. Moradi, A. Karargyris, and T. Syeda-Mahmood, "Semi-supervised learning with generative adversarial networks for chest X-ray classification with ability of data domain adaptation," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 1038–1042.

[39] G. Mittal, J. Engel, C. Hawthorne, and I. Simon, "Symbolic music generation with diffusion models," Mar. 2021, *arXiv:2103.16091*.

[40] B. M. Li, L. V. Castorina, M. D. C. V. Hernández, U. Clancy, S. J. Wiseman, E. Sakka, A. J. Storkey, D. J. Garcia, Y. Cheng, F. Doubal, M. T. Thrippleton, M. Stringer, and J. M. Wardlaw, "Deep attention super-resolution of brain magnetic resonance images acquired under clinical protocols," *Frontiers Comput. Neurosci.*, vol. 16, Aug. 2022, Art. no. 887633.

[41] G. Begus, "Ciwgan and fiwgan: Encoding information in acoustic data to model lexical learning with generative adversarial networks," *Neural Netw.*, vol. 139, pp. 305–325, Jul. 2020.

[42] S. Mulé, L. Lawrance, Y. Belkouchi, V. Vilgrain, M. Lewin, H. Trillaud, C. Hoeffel, V. Laurent, S. Ammari, E. Morand, O. Faucoz, A. Tenenhaus, A. Cotten, J.-F. Meder, H. Talbot, A. Luciani, and N. Lassau, "Generative adversarial networks (GAN)-based data augmentation of rare liver cancers: The SFR 2021 artificial intelligence data challenge," *Diagnostic Interventional Imag.*, vol. 104, no. 1, pp. 43–48, Jan. 2023.

[43] J. Kapoor, J. H. Macke, and C. F. Baumgartner, "Multiscale metamorphic VAE for 3D brain MRI synthesis," 2023, *arXiv:2301.03588*.

[44] P. A. Moghadam, S. Van Dalen, K. C. Martin, J. Lennerz, S. Yip, H. Farahani, and A. Bashashati, "A morphology focused diffusion probabilistic model for synthesis of histopathology images," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2023, pp. 1999–2008.

[45] S. Poonkodi and M. Kanchana, "3D-MedTranCSGAN: 3D medical image transformation using CSGAN," *Comput. Biol. Med.*, vol. 153, Feb. 2023, Art. no. 106541.

[46] V. Geadah, G. Barello, D. Greenidge, A. S. Charles, and J. W. Pillow, "Sparse-coding variational auto-encoders," *BioRxiv*, 2018, Art. no. 399246.

[47] G. Turinici, "Radon Sobolev variational auto-encoders," *Neural Netw.*, vol. 141, pp. 294–305, Sep. 2021.

[48] T. N. Kipf and M. Welling, "Variational graph auto-encoders," Nov. 2016, *arXiv:1611.07308*.

[49] M. Safari, M. Beiki, A. Ameri, S. H. Toudeshki, A. Fatemi, and L. Archambault, "Shuffle-ResNet: Deep learning for predicting LGG IDH1 mutation from multicenter anatomical MRI sequences," *Biomed. Phys. Eng. Exp.*, vol. 8, no. 6, Nov. 2022, Art. no. 065036.

[50] Y. Eizuka, K. Hara, and I. Suzuki, "Impact of duplicating small training data on GANs," in *Proc. 10th Int. Conf. Data Sci., Technol. Appl.*, 2021, pp. 308–315.

[51] T. Ganokratanaa and S. Aramvith, "Generative adversarial network for video anomaly detection," in *Generative Adversarial Networks for Image-to-Image Translation*. Amsterdam, The Netherlands: Elsevier, 2021, pp. 377–420.

[52] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5967–5976.

[53] S. Hirose, N. Wada, J. Katto, and H. Sun, "ViT-GAN: Using vision transformer as discriminator with adaptive data augmentation," in *Proc. 3rd Int. Conf. Comput. Commun. Internet (ICCCI)*, Jun. 2021, pp. 185–189.

[54] S. Zhao, Z. Liu, J. Lin, J.-Y. Zhu, and S. Han, "Differentiable augmentation for data-efficient GAN training," in *Proc. Adv. Neural Inf. Process. Syst.*, Jun. 2020, pp. 7559–7570.

[55] N. Miolane and S. Holmes, "Learning weighted submanifolds with variational autoencoders and Riemannian variational autoencoders," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 14491–14499.

[56] M. Connor, G. Canal, and C. Rozell, "Variational autoencoder with learned latent structure," in *Proc. Int. Conf. Artif. Intell. Statist.* PMLR, 2021, pp. 2359–2367.

[57] G. Carbajal, J. Richter, and T. Gerkmann, "Disentanglement learning for variational autoencoders applied to audio-visual speech enhancement," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, May 2021, pp. 126–130.

[58] Q. Zhang, M. Tao, and Y. Chen, "gddim: Generalized denoising diffusion implicit models," Jun. 2022, *arXiv:2206.05564*.

[59] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," Oct. 2020, *arXiv:2010.02502*.

[60] G. Pleiss, D. Chen, G. Huang, T. Li, L. van der Maaten, and K. Q. Weinberger, "Memory-efficient implementation of densenets," Jul. 2017, *arXiv:1707.06990*.

[61] Y. N. Fu'adah, I. Wijayanto, N. K. C. Pratiwi, F. F. Taliningsih, S. Rizal, and M. A. Pramudito, "Automated classification of Alzheimer's disease based on MRI image processing using convolutional neural network (CNN) with AlexNet architecture," *J. Phys., Conf.*, vol. 1844, no. 1, Mar. 2021, Art. no. 012020.

[62] M. A. Nalls, M. F. Keller, D. G. Hernandez, L. Chen, D. J. Stone, and A. B. Singleton, "Baseline genetic associations in the Parkinson's progression markers initiative (PPMI)," *Movement Disorders*, vol. 31, no. 1, pp. 79–85, Jan. 2016.

[63] A. Borji, "Pros and cons of GAN evaluation measures: New developments," *Comput. Vis. Image Understand.*, vol. 215, Jan. 2022, Art. no. 103329.

[64] R. Yacouby and D. Axman, "Probabilistic extension of precision, recall, and F1 score for more thorough evaluation of classification models," in *Proc. 1st Workshop Eval. Comparison NLP Syst.*, 2020, pp. 79–91.