Leveraging GANs, Autoencoders and their Combined Deep Learning Architecture for Anomaly Detection in Multi-Domain Image Data

by

Ashim Chandra Das
20201042
Anjishnu Banik
20201065
Noshin Sayara
20201127
Sazzad Hossain Reaz
20301342

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
Brac University
June 2025

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Ashim Chandra Das	Anjishnu Banik
20201042	20201065
Noshin Sayara	Sazzad Hossain Reaz
20201127	20301342

Approval

The thesis titled "Leveraging GANs, Autoencoders and their Combined Deep Learning Architectures for Anomaly Detection in Multi-Domain Image Data" submitted by

1. Ashim Chandra Das (20201042)
2. Anjishnu Banik (20201065)
3. Noshin Sayara (20201127)
4. Sazzad Hossain Reaz (20301342)

Of Spring, 2025 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on June 20, 2025.

Examining Committee:

Supervisor:
(Member)

Dr. Amitabha Chakrabarty

Professor
Department of Computer Science and Engineering
Brac University

Program Coordinator:
(Member)

Md. Golam Rabiul Alam, PhD

Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Sadia Hamid Kazi, PhD

Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Abstract

The study aims to identify anomalies in multiple domains through deep learning and it evaluates the performance of 9 state-of-the-art models for anomaly detection in industrial quality control, plant disease diagnosis, and medical imaging. It compares them and proposes a modification to the Bi-GAN model because this modification is expected to improve the model's performance. The 9 models are: FCAE, CAE, VAE, CVAE, Beta-VAE, VQ-VAE, AnoGAN, GANomaly, and Bi-GAN, and some are based on autoencoders (mirror) and others are based on GANs (adversarial competition between two agents). The authors modified the Bi-GAN model and they employed 3 datasets: brain MRI scans to detect abnormalities, industrial components to detect defects, and plant leaves to detect diseases, thus allowing them to test the models in different domains. The modified Bi-GAN model was the most effective model for identifying anomalies through the comparison of images, and this makes the model a good fit for anomaly detection in a variety of domains. All nine models showed promise, but each model also had its limitations, because by modifying the Bi-GAN model, the researchers were able to create a more robust and reliable approach.The study contributes and an evaluation and comparison of nine anomaly detection types. An introduction of a novel anomaly detection type that performs well in general, and a comparison of all anomaly detection types. This highlights the strengths and weaknesses of each type, and this provides good insights. The study contributes to our understanding of anomaly detection, and it also provides a basis for the design of novel anomaly detection types.

Keywords: Image Processing, Deep Learning, Generative Adversarial Networks, Autoencoder Neural Networks, Image Reconstruction, Anomaly Detection.

Acknowledgement

We acknowledge the blessing of God for the strength, endurance, and persistence he has given us to complete this thesis, and we acknowledge our supervisor, DR. Amitabha Chakrabarty for his mentorship, assistance, and input on this research.

We acknowledge Fahim-Ul-Islam, Research Assistant, BRAC University, for his assistance and providing a conducive learning and research environment, because he has played a significant role in our academic journey, and thus we acknowledge our parents and family for their care, prayers, and support throughout our studies.

We acknowledge all individuals, groups, and resources that provided data, tools, or assistance that contributed to our ability to complete this thesis, therefore we are grateful for their contributions.

Table of Contents

Declaration i

Approval ii

Abstract iii

Acknowledgment iv

Table of Contents v

List of Figures vii

List of Tables ix

1 Introduction 1
- 1.1 Background Information . . . 1
- 1.2 Problem Statement . . . 2
- 1.3 Research Objective . . . 3

2 Literature Review 4
- 2.1 Review . . . 4
- 2.2 Summary . . . 17
- 2.3 Research Gap . . . 20

3 Dataset Description 22
- 3.1 Medical Sector: Brain MRI . . . 22
- 3.2 Industrial Sector: MVTec AD . . . 23
- 3.3 Agricultural Sector: Plant Village Dataset . . . 24

4 Model Architectures 26
- 4.1 Fully Connected Autoencoder . . . 27
- 4.2 Convolutional Autoencoder . . . 29
- 4.3 Variational Autoencoder . . . 31
- 4.4 Beta-VAE . . . 32
- 4.5 Conditional Variational Autoencoder . . . 33
- 4.6 VQ-VAE . . . 34
- 4.7 AnoGAN . . . 35
- 4.8 GANomaly . . . 37
- 4.9 Bi-GAN . . . 39
- 4.10 Proposed Model . . . 41

5 Methodology **46**
5.1 Data Preprocessing 50
5.1.1 Preprocessing the Brain MRI Dataset 50
5.1.2 Preprocessing the MVTec Dataset 50
5.1.3 Preprocessing the Plant Disease Dataset 51
5.2 Experimental Setup 51
5.2.1 Training Setup 51
5.2.2 Evaluation Metrics 52
5.2.3 Computational Resources 53

6 Result **54**
6.1 Evaluation of Model Architectures (Brain MRI Dataset) 55
6.2 Evaluation of Model Architectures (MVTec AD Dataset) 63
6.3 Evaluation of Model Architectures (Plant Disease Dataset) 67

7 Discussion **72**
7.1 Result Analysis 72
7.2 Computational Efficiency 73
7.3 Real World Applicability 75
7.4 Future direction 76

8 Conclusion **77**

Bibliography **80**

List of Figures

3.1 Brain MRI dataset . . . 23
3.2 Distribution of Normal vs Abnormal images . . . 23
3.3 Normal and Abnormal Instances Across Various Industrial Categories 24
3.4 Samples of Healthy and Diseased Leaves from the Plant Pathology Dataset . . . 25

4.1 Architecture of Fully Connected Autoencoder . . . 28
4.2 Architecture of Convolutional Autoencoder . . . 30
4.3 Architecture of Conditional Variational Autoencoder . . . 33
4.4 Architecture of VQ-VAE . . . 34
4.5 Architecture of GANomoly . . . 38
4.6 Architecture of Bi-GAN . . . 39
4.7 Squeeze-and-Excitation (SE) Attention Block . . . 41
4.8 Architecture of Encoder . . . 42
4.9 Architecture of Generator . . . 43
4.10 Overview of Training Process . . . 44

5.1 Methodology for Brain MRI Dataset . . . 47
5.2 Methodology for MVTec AD Dataset . . . 48
5.3 Methodology for Plant Disease Dataset . . . 49

6.1 Confusion matrix of FC Autoencoder (Brain MRI) . . . 56
6.2 Confusion matrix of CNN Autoencoder (Brain MRI) . . . 57
6.3 Confusion matrix of VAE (Brain MRI) . . . 57
6.4 Confusion matrix of Beta-VAE (Brain MRI) . . . 58
6.5 Confusion matrix of CVAE (Brain MRI) . . . 59
6.6 Confusion matrix of VQ-VAE (Brain MRI) . . . 59
6.7 Confusion matrix of Bi-GAN (Brain MRI) . . . 60
6.8 Confusion matrix of GANomoly (Brain MRI) . . . 61
6.9 Confusion matrix of AnoGAN (Brain MRI) . . . 61
6.10 Confusion matrix of Improved Bi-GAN (Brain MRI) . . . 62
6.11 Confusion matrix of Autoencoder-Based Model (MVTec AD) . . . 64
6.12 Confusion matrix of AnoGAN (MVTec AD) . . . 65
6.13 Confusion matrix of Hybrid Model (MVTec AD) . . . 66
6.14 Confusion matrix of our proposed model (MVTec AD) . . . 66
6.15 Confusion matrix of Autoencoder-Based Model (Plant disease) . . . 68
6.16 Confusion matrix of AnoGAN (Plant disease) . . . 69
6.17 Confusion matrix of Hybrid Model (Plant disease) . . . 70
6.18 Confusion matrix of our proposed model (Plant disease) . . . 70

7.1 Error mapping for Healthy Samples 72
7.2 Error mapping for Diseased Samples 73

List of Tables

2.1 Summary of Papers 20

6.1 Performance Comparison of Different Models on Brain MRI Dataset . 55
6.2 Performance Comparison of Different Models on MVTec AD Dataset 63
6.3 Performance Comparison of Different Models on Plant Disease Dataset 67

7.1 Computational Efficiency 74

Chapter 1

Introduction

1.1 Background Information

To build a robust and precise model, a substantial amount of data for analysis is consistently necessary by Machine Learning. Heart disease, a challenging diagnosis that must be performed accurately and promptly. It is crucial to anticipate such ailments in advance by utilizing a substantial amount of data. Generating such vast quantities of data would be unfeasible without the use of synthetic data. Recent advancements in deep generative models, like GANs and VAEs, enable the creation of high-quality synthetic data, essential for machine learning algorithms in classification tasks[18]. In the medical domain, certain conditions necessitate the utilization of deep generative models to synthesize ASL images (Arterial spin labeling), which represents a novel MRI technique. It is specially tailored for quantifying cerebral blood flow. In dementia-related conditions, this tool is crucial to diagnosis. According to the World Health Organization, over the age of 65 one of each 20 people suffers from dementia disease[15]. Cutting-edge deep learning architectures are observed to be utilized in forthcoming wireless network generations. With the constraints of existing DL in meeting the demands of 5G and post-5G networks, advanced DL tackles this issue through the implementation of transformer-masked autoencoders (TMAE)[21]. As life goes on, the need for high-quality industrial products is growing at a constant pace. Image anomaly detection is crucial for ensuring product quality and operational efficiency in various sectors such as finance, cyber security, and manufacturing. However, challenges such as imbalanced datasets and complex annotation hinder its efficacy, especially in manufacturing. Based on GANs, a new theme that is scrutinized to detect anomaly images accurately without relying on anomaly samples is known as attention feature fusion (AFF)[22]. To enhance efficiency, the integration of GAN and Autoencoder is imperative. Leveraging GAN-based techniques for anomaly detection, CBiGAN merges the advantages of current methods in identifying texture anomalies while maintaining computational efficiency. Notably, this model enhances alignment between original and reconstructed images[13]

By introducing the Generative model, GANs revolutionized the field of Artificial Intelligence. With the new advanced approach, GANs represent a groundbreaking advancement that allows machines to create data besides analyzing it. Consisting of two types of neural networks - a generator and a discriminator engaged in concurrent training within a competitive environment. The discriminator attempts to

distinguish between fake and real samples once the generator creates a fake sample. Both networks increase their capacities as they compete with one another. The discriminator improves its capacity to discern between real and artificial data points, while the generator continuously improves its accuracy to create samples that mimic real data. In unsupervised machine learning a class of generative models is used that is denoted as Variational Autoencoders (VAE). In a lower-dimensional latent space, this autoencoder enables the representation of high-dimensional data. By imposing a probabilistic structure on the latent space, VAEs, unlike traditional autoencoders, allow for the formation of new data samples. Like a traditional model, the VAE model consists of an encoder and a decoder. In the latent space, using a Gaussian distribution, the encoder transfers the input data to a distribution. To produce output data, the decoder uses samples taken from the distribution. During training, VAEs attempt to reconstruct both the input data and adjust the latent space distribution in order to ensure to a pre-established prior distribution.

Neural networks are inspired by the human brain also made up of interconnected nodes, or neurons. By using neural networks with multiple layers, a branch of machine learning known as "deep learning" is concerned with modeling complex patterns in data. Every neuron receives input data, processes it, and transmits the output to the subsequent layer. Deep learning uses these multi-layered networks to automatically learn representations from data, which helps in image and speech recognition, natural language processing, and many other tasks. This hierarchical learning approach enables deep learning models to handle large amounts of data, capture complex patterns, and achieve state-of-the-art performance in a variety of applications.

GANs (Generative Adversarial Networks) and VAEs (Variational Autoencoders) significantly improve deep learning and neural networks in various ways. GANs are adept at generating high-quality synthetic data, which is crucial for augmenting datasets, especially when real data is limited or expensive to collect. This feature is helpful in many domains where realistic picture synthesis is crucial, like computer graphics and virtual reality. Also, GANs are helpful in anomaly detection, identifying anomalous data when they fail to produce a realistic version of the input. GANs make it possible for models to function successfully across domains by producing target domain data from source domains. However, for tasks like picture production, data imputation, and anomaly detection, VAEs offer an organized method to learn representations of data in a lower-dimensional latent space. VAEs are particularly good at capturing the underlying structure of the data and modeling complex data distributions. It enables a smooth interpolation between the data point and the generated one. Combining both, GANs and VAEs, enhances the performance of deep learning models and makes them stronger to handle a wide variety of applications.

1.2 Problem Statement

Image reconstruction and anomaly detection are crucial and critical tasks in several areas such as security surveillance, medical imaging and industrial inspection, which expects high precision and accuracy. Traditional methods for image reconstruction and anomaly detection often fail to provide expected output with hyperspectral

, complex data, resulting in deficient performance in precisely reconstructing images and detecting anomalies. Failure to reconstruction of images and detection of anomalies might lead to misdiagnosis and inappropriate treatment, which will affect the patients adversely. If the anomaly detection is not achieved accurately in security footage, undetected criminal activities might happen. Inaccurate anomaly detection might lead to failure to detect fault in manufactured products, which might result in financial losses, product failures and safety hazards. To deal with the mentioned problems, generative adversarial networks and variational autoencoders have shown an optimistic approach. However, their optimal configurations and comparative effectiveness still remains unexplored. Our goal is to leverage generative adversarial networks and variational autoencoder models to achieve high precision and accuracy in image reconstruction and anomaly detection.

1.3 Research Objective

1. Investigate several generative adversarial networks (GANs) architectures and its loss functions for high precision image reconstruction.

2. Explore the use of autoencoders as an alternative model to increase the accuracy of generative adversarial networks (GANs) for image reconstruction.

3. Examine different autoencoder architectures such as convolutional autoencoder, variational autoencoder and their accuracy on different types of image datasets and anomalies or outliers.

4. Build techniques for understanding and visualising the learned representations of generative adversarial networks and autoencoders to detect the outliers and anomalies.

5. Inspect the performance of offered models on several image datasets, benchmark datasets, medical image analysis and security surveillance and compare the performance of it with traditional methods.

Chapter 2

Literature Review

2.1 Review

Hyperspectral anomaly detection is crucial for identifying anomaly objects without prior knowledge, but with this strategy there are many challenges like limited labelled data and finding anomaly objects in complex distribution of data where it is difficult to distinguish between normal data and abnormal data. In paper [7], the authors introduce a framework for hyperspectral anomaly detection where GAN based semi-supervised spectral learning is used. Here GAN is used to estimate background distributions and to improve anomaly detection semi-supervised spectral learning is used with limited labelled data. To find anomalies in the training data, training strategies like limiting model capacity and dropout are employed, which ensures the improved hyperspectral anomaly detection performance and reliable background estimation. In this paper, the spatial features are generated by a morphological attribute filter and feature fusion technique to deal with the combine and background information from multiple features for a fused detection map. This paper uses 4 hyperspectral images to evaluate the proposed method from different scenes, including Texas Coast, San Diego, Los Angeles and El Segundo datasets. The images captured by the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) sensor. The Texas Coast images were captured with 100 by 100 pixels and 207 spectral bands and also corrupted with strip noise. In these images buildings were considered as anomalies in the datasets. The Los Angeles images were captured with 100 by 100 pixels and 205 spectral bands. In these images different scales in the urban area are considered as anomalies in the datasets. The proposed method was compared with the existing methods of hyperspectral anomaly detection methods like RX, LRX, CRD, AED, SAFL and AAE. The experimental result shows the superiority of the proposed method in terms of visual comparison, quantitative evaluation and computation effectiveness.

In the paper [1], the authors introduce an anomaly detection method by utilising reconstruction probability from variational autoencoder, which is more efficient and superior than existing methods like principal components and autoencoders. In the context of Variational Autoencoders, reconstruction probability is a measure which is used to evaluate how variational encoders can reconstruct data from a given input data by encoding and decoding it. This paper uses Variational Autoencoder based training algorithm to train data, which involves reparameterization to ensure the

distribution of latent variables must follow a particular form, which will enhance stability in training and convergence. The training process in that algorithm forces posterior distribution of data to be similar to the prior distribution of data. Reconstruction of data is processed through the posterior distribution of data and its likelihood, here the likelihood of individual data is calculated by Multivariate Gaussian for continuous data. The paper used a Variational Autoencoder based anomaly detection algorithm to detect anomalies, which takes samples from a distribution of data to measure reconstruction probability for each and every data point. The data points which have high reconstruction probability are classified as anomalies in the datasets. The MNIST dataset and KDD cup dataset are used here for anomaly detection. Datasets are divided into two classes one is normal data and another is anomaly data based on their classes, which enables semi-supervised learning. In the MNIST dataset, two models are trained where each digit of one model is trained as normal data, and each digit of another model is trained as anomaly data. KDD cup dataset, which consists of five main classes of data. The five classes are DoS, R2L, U2R, Probe and normal. Here DoS, R2L, U2R and Probe are in the anomaly class and the fifth one is in normal class. The experimental result demonstrates how superior the proposed method in this paper is than the existing method of anomaly detection like principle component based method and autoencoder based method, where principle component based method use dimension reduction technique and autoencoder based method uses reconstruction error for anomaly detection.

The instances of heart disease are increasing swiftly, but the research to reduce it is not making progress because of lack of data. So the authors proposed this paper to mitigate the problem by emphasizing the importance of swift and accurate prediction from synthetic data. Generative Adversarial Networks, Convolutional Generative Adversarial Networks and Variational Autoencoders have been used for augmenting the original dataset by generating synthetic data. The aim of this augmentation is to enhance the accuracy of the machine learning models which are used here for predicting heart disease. Machine learning models such as decision trees, logistic regression, knn and random forest have been used to compare the performance of both original and synthetic data in this paper. The experiment result shows that the accuracy of the models have been increased significantly which showcases the effectiveness of synthetic data in heart disease. The paper utilizes a dataset which contains 304 rows of heart disease original data from the Irvine's repository of University of California. The dataset consists of 13 different medical characteristics of 304 individuals.

In this paper[12], the authors introduce an advanced generative model called conditional variational autoencoders with adversarial training which enhances the performance in spectral synthesis. The experimental result demonstrates the superiority of the proposed model in spectral synthesis based on the analysis of three different real spectral datasets. The proposed method combines the adversarial training and variational inference processes. It mitigates the limitations of existing models such as GANs, VAEs by incorporating the labels with the data during training. The loss function of this model introduces spectral angle distance measurement and vector angle measurement to improve the spectral synthesis performance. The three datasets used in this paper are the Xuzhou dataset (HYSPEX), the University of

Pavia dataset (ROSIS) and the Indian pines dataset (AVIRIS). The Xuzhou dataset contains 368 spectral bands which is high spatial resolution and also provides a challenging dataset because of its noise bands and spectral range.

The authors of this paper [11] introduces an innovative approach abnormal to normal translation generative adversarial networks called ANT-GAN for image synthesis, which generates abnormal looking images to normal looking images without paired training data. The ANT-GAN model shows the potential for data augmentation in medical image synthesis by producing realistic lesions containing images from real ones. The architecture of ANT-GAN contains three main components, each component addressing specific aspects of the problem. The GA2N generator is the main deep network in ANT-GAN which inputs images with abnormality or without abnormality and outputs the images without abnormality. GA2N generator and DA discriminator both used simultaneously to ensure that abnormal images can be converted to normal looking images and normal looking images can be converted to abnormal looking images. Hence, ANT-GAN provides the ability to generate realistic lesions containing images from real images which provides data augmentation capabilities to medical imaging tasks.

Anomaly detection in images often relies on reconstruction-based techniques using deep generative models like Autoencoders and GANs. Autoencoders aim to minimize reconstruction errors, but struggle with pixel dependencies. In the paper [13] the authors delve into various approaches for anomaly detection in images using generative model that focuses on reconstruction-based techniques. These types of models encompass Autoencoder (AEs) and Generative Adversarial Networks (GANs). In various fields such as manufacturing and medicine, detecting defective items in large datasets is often both challenging and crucial. Conventional techniques frequently have difficulty handling high-dimensional data. By focusing on GANs based methods for anomaly detection the authors formulated a new model, CBiGAN, which combines the strengths of existing methods for detecting texture anomalies, while remaining computationally efficient. This model ensures better alignment between the original and reconstructed images. The proposed approach underwent testing using real-world data, showcasing substantial enhancements in identifying anomalies, particularly within texture elements, all the while exhibiting superior speed compared to existing advanced methods. For one class anomaly detection in images CBiGAN also combines the modeling power of BGiANs (Bi-Directional GAN model) with consistency regularization on both encoder and decoder modules. Focusing on industrial applications, the authors in paper [13] evaluated CBiGANs on the MVTec AD benchmark. In conclusion, the findings highlighted a remarkable enhancement in reconstruction ability, especially noticeable in texture images, establishing a new benchmark in this domain. However, performance varied across different object categories. CBIGAN exhibited performance on par with other efficient methods.

In ensuring product quality and operational efficiency of industrial products, image anomaly detection made a crucial area of research. Image anomaly detection has seen widespread application across various domains such as finance, cyber security and manufacturing. Yet, within the manufacturing sector, obstacles like imbalanced datasets, the complexity of annotating anomaly samples, and the varied nature of

anomalies impede the efficacy of anomaly detection. In response to these challenges, this paper presents a novel method aimed at overcoming the aforementioned hurdles. In paper [22] authors scrunitize a new theme attention feature fusion (AFF) based on GANs. Without relying on anomaly samples this approach aims to detect anomaly images accurately. For that purpose authors leverage the correlation learning ability of AFF and reconstruction ability of an auto-encoder. To enhance dataset diversity paper [22] introduces an image cut-paste module and image augmentation techniques. The paper delineates three principal contributions: firstly, the inception of a pioneering encoder-decoder GAN module augmented with AFF. Secondly, the integration of an attention feature fusion mechanism and lastly, the demonstration of enhanced anomaly classification proficiency in comparison to established methodologies.

Dementia denotes a disease of forgetfulness, dull speech, and loss of self-care ability in later stages. The World Health Organization reports that 1 in 20 individuals over the age of 65 suffers from dementia, ranking it as the fourth leading cause of death among the elderly. Accurate diagnosis is vital and often relies on biomarkers like cortical atrophy and hippocampal reduction, along with imaging tools such as MRI, CT, SPECT, and PET. Arterial Spin Labeling (ASL) is a pivotal MRI technique for diagnosing dementia, as it measures cerebral blood flow in a non-invasive manner. Unlike many other imaging methods that require the injection of contrast agents or radiation exposure, in paper [15] ASL uses endogenous tracers, making it a safer and more patient-friendly option. This advantage makes ASL particularly suitable for elderly patients who are often at higher risk from invasive procedures. However, Many major dementia research datasets lack ASL data. To tackle this issue, in paper [15] researchers have developed a new approach utilizing a type of artificial intelligence known as a Generative Adversarial Network (GAN) to generate ASL images. By creating a VAE-GAN, this new approach combines VAE (Variational Auto Encoder) with a GAN. In generating ASL images, extensive testing has demonstrated that this model outperforms other methods. In diagnosing dementia, it can enhance accuracy by up to 42.41%.

Due to the scarcity of large datasets, machine learning, and deep learning methods often face challenges that are essential for reliable and robust results. Because of privacy regulations that limit data sharing and the small number of patients available, this scarcity is primarily. To tackle this issue, Generative Adversarial Networks (GANs) have emerged as a highly promising and innovative solution in the field of artificial intelligence. GANs can produce synthetic data that closely resembles actual patient data without even referencing actual patients. In this paper[9][19] for training binary classifiers that distinguish between malignant and benign the researcher explores the use of several GAN models to create synthetic data. The objective is to explore how incorporating synthetic data can improve classification accuracy, especially in situations where there is a scarcity of real data. To accomplish this, an evaluation framework was created in which binary classifiers are trained using datasets that integrate both real and synthetic data. The findings reveal that classifiers trained on datasets enhanced with synthetic data produced by more advanced GAN models exhibit higher accuracy compared to those trained exclusively on real data. This improvement is particularly notable when the initial dataset is limited

in size. The study highlights the potential of using GAN-generated synthetic data to boost the performance of machine learning models in the medical field, proving to be a valuable approach in scenarios where data is sparse.

Brain Tumor , a notable aggressive and life-threatening disease including types such as pituitary tumors, meningiomas, and gliomas, with gliomas being the most aggressive and exhibiting the highest mortality rates. Symptoms vary from headaches and seizures to neuropsychiatric issues, making diagnosis more challenging. CT and MRI both are widely used in conventional techniques for its high resolutional images and MRI is preferred most. MRI is limited in its ability to differentiate neoplastic tissue from post-treatment changes and to reveal the full extent of tumors. So, the authors in this paper delves deeply into brain tumors, their diagnosis, and the integration of advanced medical imaging and AI to bolster diagnostic precision. For diagonostic support AI and Deep Learning are often promising to tackle this issue. Traditional machine learning relies on manual feature extraction. In paper[10] authors describes a different efficient method (GANs) that will learn directly from extensive datasets. Nevertheless, the scarcity of medical image dataset remains a challenge. To tackle this issue, the authors in paper [20] suggest a novel approach merging Variational Autoencoders (VAEs) and GANs to create synthetic MRI images. Initially, using existing MRI images an encoder-decoder network undergoes training.This network generates a noise vector encapsulating vital image features. The GAN then leverages this vector to produce realistic MRI images, successfully avoiding issues like mode collapse. A classifier was trained with both real and generated images ResNet50 that shows increase in classification accuracy from 72.63% to 96.25% . The model demonstrates elevated recall, specificity, precision, and F1-score values for gliomas, which are recognized as the most severe brain tumor type. Moreover, this method shows potential for wider utilization in diverse medical imaging disciplines, serving as a valuable tool to boost diagnostic accuracy and optimize workflow for healthcare professionals.

The paper "Attention and Autoencoder Hybrid Model for Unsupervised Online Anomaly Detection"[23] introduced an innovative method to detect unsupervised anomalies in time series datasets by integrating attention mechanisms into autoencoders architecture. Unveiling these unlabelled anomalies is vital as they pose serious threats in forms of cyber attacks, sensor malfunctions as well as operational inefficiencies. Conventional models either demand labeled data or are not capable of functioning in real-time, raising difficulties in dynamic scenarios. This paper tackles these difficulties by presenting an unsupervised and online approach that detects abnormalities in real-time as data is being generated. To reduce reconstruction error, they first convert the data into a latent space and then rebuild it. Without the requirement for recurrent or convolutional networks, transformers are able to effectively represent sequences through the use of attention processes. This enables effective parallel processing. For the purpose of maintaining the chronological sequence of time series data, positional embeddings are absolutely necessary.The proposed hybrid model dynamically combines autoencoders and attention mechanisms to maximize their own strengths. The distinctive design of the hybrid model allows it to identify various anomalies by taking advantage of both short and long-term features. Anomalies are identified based on reconstruction errors. Significant

deviations in error across consecutive time steps signal anomalies.For the purpose of evaluating the model, six real-world datasets acquired from the Numenta Anomaly Benchmark (NAB) were utilized. The data for these statistics originated from a wide range of sources, such as traffic data, cloud services provided by Amazon Web Services (AWS), temperatures of industrial machines, and activity on social media. Furthermore, it demonstrated a high level of accuracy and comprehensiveness in spotting abnormalities, and it typically beat models that were already in existence. i) Both point and collective abnormalities, as well as contextual anomalies, were effectively identified using the hybrid model. ii) Both recall and precision were high, with a minimum precision of over 93%. iii)It was able to defeat competitor models, such as VAE-LSTM, particularly when applied to datasets that had intricate patterns. In summary, The hybrid technique efficiently detects anomalies by including both short- and long-term connections within time series data. The real-time functionality and unsupervised nature make it suitable for a diverse variety of applications and provide a dependable solution for dynamic and high-frequency data environments.

In this research paper[21], it is discussed about the potentials of a distinctive deep learning architecture applied in the wireless communication sector of the future. In particular, the potential of using one of the newest attention-based models, Transformer Masked Autoencoders. The research presents TMAEs as a potential remedy for addressing a range of issues in these networks, particularly those pertaining to 5G and future generations.Transformer-based autoencoders (TMAEs) capitalize on the advantageous attributes of both transformers and autoencoders to reconstruct data from scattered observations. The following components comprise the architecture: Encoder: The encoder converts partial observation of input data through transformer-based neural network (NN) with masked segments to latent representations. Decoder: It helps to reconstruct the original data using the latent representation as well as placement of segments that have been encrypted. The paper describes a case study that shows how TMAEs can be used for picture compression in NG, and in the scenario where Unmanned Aerial Vehicles are involved. The text explains the way the exploitation of TMAEs, in cooperation with traditional techniques for picture compression like JPEG, can remarkably raise the compression rates with no loss of quality of the processed images.TMAEs can solve several of the most serious problems in next-generation mobile networks, such as: (i)Source and Channel Coding,(ii)Estimation, iii) Security TMAEs are promising but pose challenges in terms of high computing resources and requiring large amounts of labeled data to train. The report recommends further research into increasing the efficiency of these models for practical implementation in wireless networksTMAEs are a viable technique for improving wireless communication performance, efficiency, and security for future generations. They will be offering a new paradigm by blending transformer-based designs with autoencoding methods appropriate for complicated tasks such as signal processing, channel estimation, and resource allocation.

The article, titled "Synthesizing Multi-Contrast MR Images Via Novel 3D Conditional Variational Auto-Encoding GAN,"[17] introduces a method for generating multi-contrast MRI pictures from only a single CT scan, using a specially designed generative network called CAE-ACGAN. It proposes a network architecture called

CAE-ACGAN that integrates the principles of a Variational Auto-Encoder and a Generative Adversarial Network for the generation of multi-contrast MR pictures from single CT scans. It allows for the effective generation of Magnetic Resonance images that rely on only a small amount of Computed Tomography data. In order to create synthetic magnetic resonance pictures that are more accurate and of higher quality than those generated by alternative methods, the suggested network combines the capabilities of VAE and GAN, in addition to an extra discriminative classifier network.This study highlights the need of selecting proper loss functions when utilizing deep learning techniques. Additionally, it investigates the benefits and challenges that are related with the use of 3D networks for picture synthesis tasks. The combination of Variational Autoencoder and Generative Adversarial Network is a method of effectively overcoming the problems of blur images and mode collapse. The VAE employs variational inference to rebuild target pictures, although it frequently generates hazy images as a result of improper optimisation and noise.GAN-based approaches acquire the measurement by training a discriminative network without the need for explicit specification. However, these methods can encounter problems with convergence and mode collapse. By combining VAE and GAN, CAE-ACGAN compensates for their intrinsic shortcomings, enhancing image clarity. The network integrates an additional classifier to enhance its capacity to distinguish, while ensuring that synthetic and real MR images remain comparable and consistent. By employing adversarial losses and image reconstruction loss, CAE-ACGAN guarantees the high quality and clarity of the synthesized MR images. CAE-ACGAN effectively reduces image blurriness, resulting in higher-quality synthetic MR images, by utilizing the capabilities of VAE, GAN, and the auxiliary classifier, in contrast to other conventional networks. The architecture, benchmarked with tests on brain imaging datasets, performed well on generating multi-contrast MRI pictures better than methods developed previously. According to both quantitative evaluation and visual inspection, CAE-ACGAN was capable of generating synthetic images that were far better in quality and more faithful than images generated by other methods. The authors also bring to light the challenges that the training of GAN-based models, such as instability and mode collapse, which their network overcomes in a very satisfying way. The comparison of the CAE-ACGAN network with pix2pix and WGAN-GP for multi-contrast MR images showed better performance concerning picture quality and accuracy. The quantitative analysis using measures such as PSNR, SSIM, and MAE showed that MR images synthesized by CAE-ACGAN were more similar to the reference images and therefore had better quality compared to other networks. Upon visual examination, it was observed that the synthetic MR images generated by CAE-ACGAN exhibited superior preservation of intricate brain anatomical characteristics compared to pix2pix. Additionally, the picture quality of CAE-ACGAN was found to be greater than that of WGAN-GP. The length of a round of trials varied between 6 and 9 hours, with the longest runs taking up to 9 hours for CAE-ACGAN and WGAN-GP, since both of them had more parameters to be trained, but the testing to generate a whole synthetic MR image took about 20 seconds. The end of the study gives a summary of what the CAE-ACGAN network has done to improve the quality and accuracy of synthetic multi-contrast MRI images. The authors also note that their network is fairly complex and requires lots of computational resources, they also note that future work will concentrate on reducing the design without sacrificing speed. They are

of the opinion that their method may be utilized for a variety of different activities and circumstances, which might potentially increase its application in the field of medical imaging.

The paper[10] explains a model that combines Variational Autoencoder and Generative Adversarial Networks to synthesize highly realistic MRI pictures. These images include precise pixel-level information for cardiac segmentation. The VAE-GAN model synthesizes cardiac shapes and the associated MR images. These images are then used to train Convolutional Neural Networks (CNNs) for segmentation. The paper shows that Convolutional Neural Networks trained with such artificially generated pictures achieve comparable results. Moreover, the use of conventional data augmentation methods with synthetic data further improves the performance. Acquiring large, annotated medical data sets is costly in time and effort. In the presented paper, this work presents a VAE-GAN model that is used to generate realistic macro MR images as that they resemble real ones such that they can be employed in creating large annotated datasets for training segmentation models. Furthermore, the method makes use of a SPADE within the GAN generator in order to generate MR pictures that will match to the dimensions of the heart that have been specified. The VAE is able to acquire knowledge on the distribution of heart shapes, which enables it to generate several anatomical maps through sampling. These maps then generate Magnetic Resonance pictures from the training process by the GAN. These pictures generated are used to train a discriminator. In the test time, the VAE gives anatomical maps, from which the GAN generates corresponding MR pictures. The technique is applied for training and testing the two datasets: ACDC and Sunnybrook. This produced improved Dice scores. For instance, fine-tuning on the synthetic Sunnybrook images improved the Dice score from 0.776 to 0.874. The VAE-GAN model also enhances generalization capabilities across datasets. The ENet CNN's Dice scores when training on augmented VAE-GAN-generated datasets were much higher compared with training on real datasets between 2-10%. The VAE-GAN model performed better in terms of generalization. The model that was trained on the synthetic ACDC dataset earned a Dice score of 0.853 on the Sunnybrook dataset. On the other hand, when it was trained on the genuine ACDC dataset, it achieved a Dice score of 0.773. In short, the article introduced a novel VAE-GAN model meant for producing real cardiac MRI images along with their annotations which can be used for heart cine-MRI segmentation. Following the training and testing of the model on two datasets, namely ACDC and Sunnybrook, promising results were obtained, particularly when using data augmentation and fine-tuning techniques.

This research article [9] is about examining various approaches and strategies that deal with video anomalies particularly focusing on the improvements achieved through deep learning techniques. The proposed model is a combination of deep learning and anomaly detection. It uses a video feature extractor trained by Generative Adversarial Network (GAN) along with an anomaly detector that has been enhanced by transferring the extractor from one network to another which makes it perform better than most other algorithms used for detecting anomalies. Additionally, the topic of anomaly detection in video surveillance is highly demanded, as the use of surveillance devices experiences growth. The model being proposed uses some of the

recent advanced deep learning techniques like Generative Adversarial Networks and Autoencoders for improved performance in identifying anomalies in videos. This hybrid model's main goal is to be as good at detecting video anomalies as other models. It achieved 94.4% recall rate and 86.4% precision rate on UCSD pedestrian dataset. This shows that the hybrid deep learning technique used for training was effective. Quantitative analysis and visualizations showed that the model has the ability to generate real video clips, thus proving the efficiency of Generative Adversarial Network (GAN) component in producing data that look like real surveillance camera footage. The hybrid deep learning model that was developed attained an accuracy of 88.60% in the evaluations that were carried out, demonstrating its usefulness in accomplishing tasks involving the identification of video anomalies.

In paper [5], the authors examine the anomaly detection for skin Disease images. Through using VAEs that stands for Variational Autoencoder model authors conduct comprehensive analysis to detect anomaly (abnormalities) in skin images. To enhance the accuracy of detecting anomaly they propose a model architecture based on the Deep Convolutional Generative Adversarial Network in short (DCGAN). Utilizing fully convolutional layers within both the encoder and decoder components. Consisting of five convolutional blocks and five transposed convolutional blocks enable to make the model suitable for different image sizes that was conducted by decoder and encoder respectively. Images were allocated to 128x128 pixels for this experiment. The VAE model was conducted to an extensive training process, utilizing the ADAM optimization over the course of 40 epoched in order to ensure robust performance. Important hyperparameters incorporated into the model included a learning rate set at 0.0001, a batch size of 32, and a latent dimension of 300. The ISIC2018 Challenge Dataset, which encompasses images representing seven distinct types of skin disease was used. To ensure consistency, the images were normalised and resized. The Final outcome denotes that VAEs model's efficacy in accurately detecting anomalies in skin images, highlighting its robust capacity. With an overall AUC of 0.77, the proposed model achieved the best performance based on reconstruction scores. In AKIEC and Melanoma, the model exhibited notable proficiency in the detection task. With AUC scores 0.872 and 0.862 respectively. The effectiveness of scores derived from KL divergence were less, as the model assigned less importance to the KL term during training. Modifying the weight of the KL term failed to improve the performance. Contrasting against conventional approaches like PCA or Kernel-PCA demonstrated that VAEs handle high-dimensional image data better then not needing detailed feature engineering. Examination of reconstructed image revealed that they were somehow blurry while the VAE could recreate images, pointing towards potential areas for enhancement within the environment. In summation, the experiment provided comprehensive insight showing VAEs's usefulness in detecting abnormal skin images. This proposed method enables doctors to identify skin disease with almost highest accuracy and efficiently. The findings strongly indicate that with further improvement, VAEs could potentially serve as a valuable tool in dermatology for early and accurate disease detection. In future work, at augmenting VAE efficacy, Potential enhancements encompass the utilization of more sophisticated decoders and the exploration of diverse methodologies aimed Other anomaly detection techniques using VAE latents, such as fitting probability distributions or using metric learning, are proposed.

In order to detect abnormalities in time series data, a novel technique called "LSTM-Based VAE-GAN for Time-Series Anomaly Detection" [8] employs Generative Adversarial Networks (GAN), Variational Autoencoders (VAE), and Long Short-Term Memory (LSTM) networks. The large volume and poor periodicity of industrial data make traditional techniques like ARIMA and EWMA difficult to handle. The time series is encoded into a latent space and probabilistically reconstructed using VAE, LSTM networks are used to capture temporal relationships, and GAN is used to ensure that the produced data closely approaches the real distribution of normal data, hence improving the quality of the reconstruction. Three basic parts make up the model: an encoder (LSTM) that maps the input time series to a latent space; a generator (LSTM) that reconstructs the time series from this latent space; and a discriminator (LSTM) that divides the data into produced and actual.The model learns to use reconstruction error and Kullback-Leibler divergence to encode, produce, and discriminate time series data during the training phase. The discriminator and generator are trained in an adversarial fashion. In conclusion, by merging extensive neural network architectures and maximizing their combined training, the LSTM-based VAE-GAN approach successfully tackles the difficulties of identifying anomalies in complicated industrial time series data.By Replacing the LSTM networks with Transformer architectures to detect sequential anomalies, which may show superior performance in capturing long-range dependencies in sequential data. Hybrid Model Integration with LSTM-based VAE-GAN To create a model that can manage datasets with numerical and categorical data in an efficient manner, enhancing the categorical characteristics in a high-dimensional space while maintaining the semantic linkages. such as a CNC (Computer Numerical Control) machine, is kept under observation to guarantee peak performance and avert malfunctions. (Astronomical telemetry dataset, electric power plant or health care monitoring system, financial market analysis)

In this study[4], conditional variational autoencoders (CVAEs) are introduced as a tool for illness diagnosis in medical pictures. By using CVAE to take into account extra information, it seeks to address the drawbacks of supervised approaches by learning compact representations of input data. The primary obstacle in detecting diseases is their heterogeneous nature, which poses a barrier to the development of precise automated detection systems. The suggested approach allows CVAE to precisely rebuild medical pictures because it is trained solely on healthy ones. Because CVAE was trained on healthy pictures, it struggles to recreate sick areas while processing diseased images during testing. The sick regions can be more easily identified because of the variations between the original and reconstructed pictures. Both 2D and 3D medical picture datasets were used in the experiments, and while the 2D photos produced encouraging results, the 3D images exhibited performance constraints, highlighting the need for more model architecture refinement for complicated data. This method's applicability in picture registration, where inaccuracies might be caused by illnesses, is significant. picture registration accuracy can be increased by preprocessing the picture and applying CVAE-based disease diagnosis. Future work and limitations: The paper notes several shortcomings, including lower performance when compared to supervised approaches, difficulties with 3D picture reconstruction, and high computing demands during training. Improved segmenta-

tion techniques, CVAE architecture optimization, and method validation on various medical datasets should be the main goals of future development. Furthermore, improving the reconstruction quality and illness diagnosis accuracy may be achieved by integrating CVAE with additional models, such as Generative Adversarial Networks (GANs).

The goal of the article[14] is to address the deficiency of diverse MRI images of brain tumors for computer model training. With the use of Structural Similarity Index (SSIM) loss function and Generative Adversarial Networks (GANs), the authors created a model known as PGGAN-SSIM. Compared to conventional enhancement approaches, this model can provide realistic, high-quality brain tumor MRI pictures with better contrast. PGGAN-SSIM outperforms other GAN-based techniques, as evidenced by its higher Multi-Scale Structural Similarity (MS-SSIM) scores and lower Fréchette Inception Distance (FID) scores in the experimental data. On the other hand, issues including high processing costs and low image resolution have also been mentioned. The authors propose investigating ways to generate high-resolution photographs while lowering processing demands in order to tackle these issues. For improved performance, they advise enhancing GAN structures and adding multimodal data. All things considered, further investigation is required to get beyond current obstacles and completely realize the promise of GAN-based data augmentation in medical imaging, even though the suggested approach seems promising for enhancing medical image analysis.

According to the article [2], In the past few years, GANs have become a powerful framework for learning generative models that transform simple latent distributions into complicated data distributions. Essentially, Generative Adversarial Networks consist of a generator that produces data samples from latent variables and a discriminator that tries to distinguish between real and generated samples. The adversarial training forces the generator to produce data that are very realistic, especially for image domains. In simpler terms, the latent space acquired by GANs effectively encodes significant semantic variations within the data, for example, facial attributes in human face datasets. However, GANs typically have limited ability to learn features because they lack an explicit connection between data and the latent space. According to Donahue et al. (2017), Bidirectional Generative Adversarial Networks (BiGANs) were created to address this problem. In addition to the generator and discriminator, BiGANs extend the GAN architecture by jointly learning an encoder that maps input to latent representations. Hence, BiGANs do not need to be taught with labelled data; instead, this bidirectional mapping helps them learn feature representations useful for downstream supervised tasks such as detection and classification. BiGANs were tested on complicated datasets like ImageNet and permutation-invariant MNIST. BiGANs capture so much of the visual characteristics for the encoder to learn convolution filters with Gabor-like structures in a way almost parallel to those learned by fully supervised networks. Quantitative studies show that BiGAN features perform competitively in classification accuracies on the ImageNet and PASCAL VOC benchmarks, generally outperforming other unsupervised and some self-supervised feature learning methods. BiGANs learn directly from unlabelled static images in ways that are indifferent to domains whereas

self-supervised methods need certain domain-based pretext tasks or auxiliary inputs such as video or egomotion. Hence, this generality may prevent domain-shift problems usually experienced by self-supervised methods, where the pretext task, such as patch prediction or picture inpainting, alters the distribution of the data. To improve feature quality, the BiGAN architecture is generalized in order to allow the encoder to work on inputs with higher resolution than the output of the generator. In summary, BiGANs provide a strong unsupervised learning method for feature representations by utilizing an encoder network along with generative modelling. BiGANs have obtained competitive performance in visual identification without requiring labelled input or domain specific supervision and bring together two areas of research (generative adversarial training and learning feature representations).

The main objective of the article [6] denotes the Imaging biomarker detection, which is a clinical need for diagnosing disease, assessing treatment, and monitoring follow-up. State-of-the-art supervised deep learning methods have detected such biomarkers with expert-level performance in various imaging modalities, including OCT, CT, MRI, and X-ray. Such methods, nevertheless, depend significantly on expert-annotated datasets that are both time-consuming and costly to obtain. In addition, the ability of supervised methods to detect new or subtle disease symptoms is limited by their very nature of only being able to identify predetermined and known abnormalities. Unsupervised anomaly detection methods have been of specific interest as a means to overcome these limitations. By learning the distribution of normal data and expressing deviations as anomalies, these methods enable the detection of uncommon or new pathological patterns without training on labelled abnormal data. Generative adversarial networks (GANs) have been one of the most successful methods for learning complex data distributions. GANs, which were originally proposed by Goodfellow, can be constituted of an adversarially trained generator and discriminator for producing realistic samples. WGAN architecture, nevertheless, which achieves stable training through optimization of the Wasserstein distance, dispenses with the mode collapse and instability in training problems of the original GAN proposals. Schlegl built on these advances by proposing AnoGAN, an unsupervised anomaly detection with GANs trained on normal data. AnoGAN detects anomalies by iteratively projecting query images onto the GANs latent space and estimating reconstruction errors. Although promising results were reported, AnoGANs iterative mapping is computationally expensive, its application being limited due to this. By utilizing a fast encoder network that directly maps images to the latent space, the f-AnoGAN model overcomes this limitation and performs real-time anomaly detection. The approach learns healthy anatomical variation by training a WGAN on normal images. It then trains an encoder to invert the mapping created by the generator. For better detection accuracy and pixel-level anomaly localization, an anomaly is detected by utilizing a composite score of image reconstruction residuals and discriminator feature residuals. In short, f-AnoGAN leverages WGANs generative power with an learned encoder for efficient latent space mapping, which is a significant advancement in unsupervised anomaly detection. The technique has wide-ranging applications outside of ophthalmology and facilitates scalable annotation-free detection of known, and perhaps new, abnormalities in medical imaging.

The role of medical imaging in healthcare has risen in paper [3] but with the workload of radiologists, diagnostic errors have increased as well. Among those errors, more than 70% are perceptual errors. The number of images that must be analyzed itself makes the task even more difficult for observers, and hence artificial intelligence (AI) and machine learning (ML) are being introduced to assist in pre-analysis of medical images and reduce human error. Generative adversarial networks (GANs), introduced by Goodfellow (2014), have been more popular in medical image analysis for tasks including segmentation, classification, detection, registration, denoising, reconstruction, and synthesis. GANs are particularly promising for anomaly detection, a method that identifies abnormal data patterns without requiring labeled pathological examples. This is crucial in medical environments where there is a lack of labelled pathology data. In order to identify hidden or uncommon diseases, anomaly detection algorithms study the distribution of normal data and mark variations as possible anomalies. GANomalys ability to generate normal images while detecting anomalies by evaluating errors in reconstruction is what differentiates it among the GAN-based methods of anomaly detection literature. However, GANs are limited in that they cannot produce high-quality images and the training process can be unstable. Incrementally improving GANs (PGGANs) were developed to address these limitations. PGGANs provide a means of producing higher quality image synthesis which is achieved by stabilising training by incrementally increasing image resolution during training. The paper presents a novel method named Progressive GANomaly, which combines the anomaly detection framework of GANomaly with the stability and high-resolution capabilities of progressively growing GANs. The goal of this combined method is to improve the anomaly detection accuracy on medical images specially brain MRI images. Progressive GANomaly outperforms conventional GANomaly and one-class SVM baselines for anomalies of varying size and intensity, as demonstrated through experiments on the Fashion MNIST, Medical Out-of- Distribution Analysis Challenge (MOOD), and an in-house brain MRI data set. Though the method needs larger training data sets because of increased model complexity but it results in better reconstruction quality and sensitivity of anomaly detection, especially at bigger patch sizes. In conclusion, the combination of progressively growing GANs into anomaly detection systems may increase the accuracy of medical image processing and robustness, potentially lessening the load on radiologists time associated with medical images and improving the diagnostic results.

In this study [16] Generative Adversarial Networks (GANs) are a strong framework for anomaly detection due to their ability to model complicated data distributions. To provide an overall understanding of the effectiveness and implementation of significant GAN-based anomaly detection algorithms such as AnoGAN, BiGAN/EGBAD and GANomaly, this research study collected and assessed them in a single implementation using TensorFlow. GANs were first proposed as a generative model framework in the seminal paper by Goodfellow (2014). The framework has since been adapted to suit anomaly detection applications. One of the earliest methods using GANs is AnoGAN (Schlegl, 2017), which learned a latent representation for normal data and used reconstruction errors to find anomalies. The cost of AnoGAN's inference process involving repeated optimization for each test sample limits its scalability. BiGAN/EGBAD (Zenati, 2018) solved these limitations

by adding an encoder along with the generator and discriminator, allowing for the direct inference of latent representations without the process of costly optimization. This approach leverages adversarial losses and reconstruction, both to facilitate anomaly scoring and train more efficiently. The framework has been further improved upon with GANomaly (Akcay, 2018), which used an autoencoder-like framework along with adversarial training to facilitate learning and easier comprehension of the anomalous scores. This paper evaluates these architectures by employing benchmark datasets such as MNIST, Fashion-MNIST, CIFAR-10 datasets and the KDD99 intrusion network dataset. They rely on the area under the precision-recall curve (AUPRC), which is appropriate for problems involving unbalanced anomaly detection. They apply the area under the precision-recall curve (AUPRC), which is a reasonable measure for the unbalanced anomaly detection problems. They observe that BiGAN/EGBAD and GANomaly appeared competitively in performance, and that GANomaly had a simple encoder architecture which allowed for faster training time, and easier calculations for generating the anomaly scores. Although AnoGAN is efficient, the results of AnoGAN cannot be repeated because of the high computational cost. Regardless of its efficiency, the performance of AoGAN cannot be replicated due to its high computational cost. Furthermore, the paper points out the difference between original papers and publicly available implementations and highlights how crucial reproducibility is to GAN research. The authors introduce a modular toolbox for anomaly detection, with the possibility to use it out-of-the-box and extend it.

2.2 Summary

Ref	Task	Model	Architecture	Accuracy
[7]	Semisupervised Spectral Learning with Generative Adversarial Network for Hyperspectral Anomaly Detection	GAN, WGAN	Autoencoder architecture (Encoder-E(), Generator())	N/A
[1]	Variational Autoencoder based Anomaly Detection using Reconstruction Probability	VAE	Variational Autoencoder architecture (contains of three hidden layer)	N/A
[18]	GANs and VAEs As Methods of Synthetic Data Generation and Augmentation to Enhance Heart Disease Prediction	GAN, CGAN, VAE	The architecture leverages GANs, CGANs, VAEs for data augmentation	CGAN-90%

Ref	Task	Model	Architecture	Accuracy
[12]	CVA2E A Conditional Variational Autoencoder With an Adversarial Training Process for Hyperspectral Imagery Classification	CVA2E	Utilizes variational inference and adversarial training process	ROSIS-90.79%
[11]	An Adversarial Learning Approach to Medical Image Synthesis for Lesion Detection	ANT-GAN, CycleGAN	ANT-GAN architecture, which consists of three main components to capture particular aspects desired for the research	67.94%
[13]	Combining GANs and AutoEncoders for efficient anomaly detection	CBiGAN	Encoder, Generator, Discriminator, Cycle Consistency Regularization	93.8%
[22]	Anomaly Detection of GAN Industrial Image Based on Attention Feature Fusion	GAN	Network Architecture (Generative, Discrimination, attention feature fusion)	94.3%
[15]	A new VAE-GAN model to synthesize arterial spin labeling images from structural MRI	VAE-GAN	VAE - Generator encoder & decoder, discriminator	84.09%
[19]	GAN-Based Approaches for Generating Structured Data in the Medical Domain	GAN	CGAN Two adversarial networks,five GAN variants, including GAN, CGAN, CTGAN, CopulaGAN, and WGANGP	N/A
[20]	Brain Tumor classification using a combination of VAE-GAN	ED-GAN	VAE and GAN	96.25%

Ref	Task	Model	Architecture	Accuracy
[23]	Attention and Autoencoder Hybrid Model for Unsupervised Online Anomaly Detection	Attention and Autoencoder Hybrid Model	Hybrid model combining an Autoencoder (AE) and a Transformer-based attention mechanism.	98%
[21]	Transformer Masked Autoencoders for Next-Generation Wireless Communications: Architecture and Opportunities	Transformer Masked Autoencoder (TMAE)	Transformer based architecture	N/A
[17]	Synthesizing Multi-Contrast MR Images Via Novel 3D Conditional Variational Auto-Encoding GAN	CAE-ACGAN	hybrid architecture combining (VAE) and (GAN).	N/A
[10]	On the effectiveness of GAN generated cardiac MRIs for segmentation	VAE-GAN	(VAE), (GAN)	88.8%
[9]	3D-Convolutional Neural Network with Generative Adversarial Network and Autoencoder for Robust Anomaly Detection in Video Surveillance	A hybrid supervised deep learning model	Convolutional Neural Networks (CNN), Generative Adversarial Networks (GAN), and Autoencoders	88.6%
[5]	Anomaly Detection for Skin Disease Images using VAE	Variational Autoencoder	VAE trained on ISIC2018 Challenge	0.779 AUCROC overall, 0.864 AUCROC for melanoma, and 0.872 AUCROC for actinic keratosis
[8]	LSTM-Based VAE-GAN for Time-Series Anomaly Detection	LSTM-based VAE-GAN	Encoder, generator, and discriminator based on LSTM networks	N/A
[4]	Unsupervised pathology detection in medical images using conditional Variational Autoencoders	(VAE), specifically a Conditional VAE (CVAE)	Encoder, latent space, decoder	Effective for 2D images but less so for 3D images,

Ref	Task	Model	Architecture	Accuracy
[14]	Data Augmentation For Medical MR Image Using Generative Adversarial Networks	(GAN)	Progressive Growing of GANs (PG-GAN), Alpha GAN with Gradient Penalty	Lower Frechet Inception Distance (FID) scores and higher Multi-scale Structural Similarity (MS-SSIM) scores compared to other GAN-based methods.
[2]	Adversarial Feature Learning	(Bi-GAN)	Encoder, Generator and Discriminator	97.39%
[6]	f-AnoGAN: Fast Unsupervised Anomaly Detection with Generative Adversarial Networks	(f-AnoGAN)	Generator, Discriminator and Encoder	N/A
[3]	GANomaly: Semi-Supervised Anomaly Detection via Adversarial training	GANomaly	2 Encoder, Decoder and Discriminator	88.2%
[16]	A survey on GAN for Anomaly Detection	(C-GAN), Bi-GAN, AnoGAN, GANomaly, EGBAD	Encoder, Decoder, Generator and Discriminator	94.86% (Bi-GAN), 83.56% (GANomaly)

Table 2.1: Summary of Papers

2.3 Research Gap

1. The study on data augmentation for medical magnetic resonance imaging (MR) using Generative Adversarial Networks (GANs) has identified limitations, such as mode collapse and training stability issues, particularly with high-resolution inputs. The PGGAN-SSIM model's efficacy is demonstrated, but its limited use on small-scale datasets may hinder its generalizability to larger, more varied datasets. Further validation of the produced images is needed to improve clinical results, such as tumor identification accuracy.

2. Using only a small number of brain images makes the model work only for those specific cases. Using powerful computers to train the model is expensive and not available to everyone. In the future, we can improve by using more diverse data and better technology like StyleGAN and BigGAN. Adding different types

of MRI images can make our data better, and we can train the model faster by learning from what other models already know. We need to find ways to make sure the images don't all look the same and test if they're actually helpful for diagnosing illnesses.

3. The VAE-GAN model, used to create arterial spin labeling images from structural MRI, has limitations due to its limited use of unique datasets and computational complexity. This may hinder its generalizability to larger datasets and real-time clinical applications. Further research is needed to enhance the model's efficiency and confirm its efficacy in larger clinical trials.

4. The study on data augmentation for medical magnetic resonance imaging (MR) using Generative Adversarial Networks (GANs) has identified limitations, such as mode collapse and training stability issues, particularly with high-resolution inputs. The PGGAN-SSIM model's efficacy is demonstrated, but its limited use on small-scale datasets may hinder its generalizability to larger, more varied datasets. Further validation of the produced images is needed to improve clinical results, such as tumor identification accuracy.

Chapter 3

Dataset Description

To ensure the model's robustness and effectiveness, utilizing distinct datasets is essential for our deep learning based research. We are incorporating three datasets from different domains to visualize the model's performance across diverse domains.

The first chosen dataset that consists of healthy brain and unhealthy (brain has disease) data was used to detect the anomalies in brain MRI scans, demonstrating the model's ability, along with monitoring the progression of a disease, because they can aid in treatment planning by giving information about the morphology of the abnormality. Next, we employed an industrial inspection dataset to evaluate anomaly detection, and this allowed us to assess the performance of our method in a real-world setting. Lastly, we concluded our implementation part by using an agricultural-based dataset that contains a healthy and crop leaf disease image dataset to detect the abnormality in it.

The data used in this study are from a variety of problems, including fine details and small defects in brain MRI images for medical anomaly detection, and various defect shapes and fine-grained details in industrial inspection, and this variety of problems is utilized because the data are from different sources. This helps in identifying potential overfitting to a particular dataset and provides a clearer picture of how the model might behave in real-world applications.

3.1 Medical Sector: Brain MRI

The prevalence of brain diseases presents a significant health challenge, as they are often life-threatening and have the potential to cause severe damage to both the brain and the body.

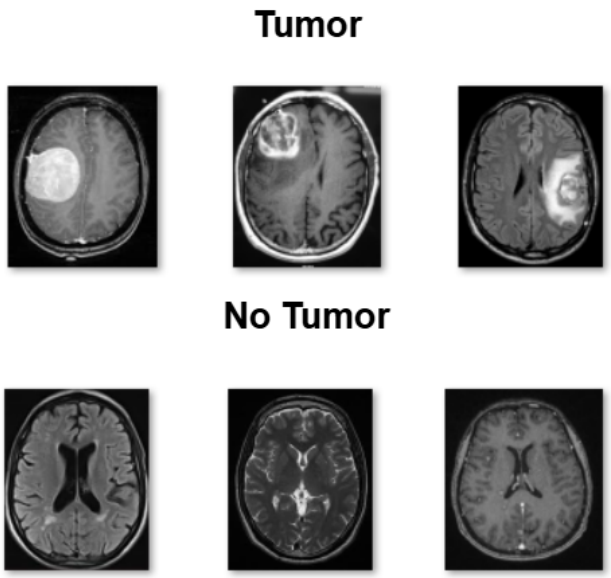

Figure 3.1: Brain MRI dataset

Since our research centers on reconstructing images and detecting anomalies, reliable and high-quality datasets are essential for training models, which leads to effective reconstruction. We collected our dataset from Kaggle, which is based on Brain MRI. The 253 images in the presented dataset are divided into two categories: normal (healthy brain) and abnormal (disease). Out of the total image files, 155 represent abnormal cases, indicating the presence of a tumor, while the remaining images correspond to healthy cases with no signs of tumor.

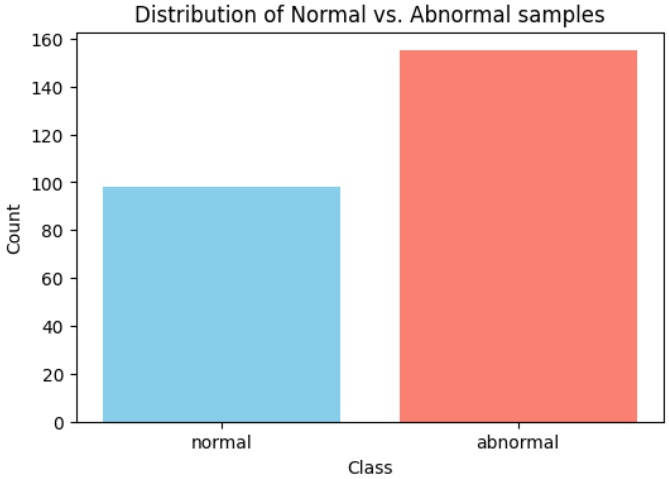

Figure 3.2: Distribution of Normal vs Abnormal images

3.2 Industrial Sector: MVTec AD

We experimented with the MVTec AD dataset, a widely used dataset for evaluating anomaly detection models on industrial images. The dataset contains more than 5,000 high-resolution images, and it includes 15 categories (such as bottle, screw,

carpet, capsule and even more) of objects and textures. For each category, normal images are available for training. The test images contain both normal and anomalous images because the anomalies are of various sizes, shapes, and locations. To observe how all these models perform in an industrial area, we selected the bottle among 15 categories. Therefore, the dataset is challenging and suitable for evaluating anomaly detection models because it represents real-world industrial inspections.

Figure 3.3: Normal and Abnormal Instances Across Various Industrial Categories

The ability to detect small and various types of defects is essential, so we use this data for research to evaluate the model's performance on various objects and defects.

3.3 Agricultural Sector: Plant Village Dataset

To assess the model's effectiveness and to observe how it performs in different distinct scenarios, we used an additional dataset called the New Plant Disease dataset, sourced from Kaggle. This dataset is agriculture-focused and includes a wide variety of images featuring both healthy and diseased crop leaves. It has been carefully designed to aid research in plant pathology and support advancements in agricultural diagnostics.

The dataset that we've utilized contains a lot of varieties of leaf samples and includes unhealthy samples that illustrate leaves affected by various plant diseases, including fungal, bacterial, and viral infections. It is associated with 38 directories, which are categorized into different classes such as Apple, blueberry, corn, strawberry, and even more, demonstrating the condition of the leaf. Almost 88K data images are available, which is highly suitable for anomaly detection tasks.

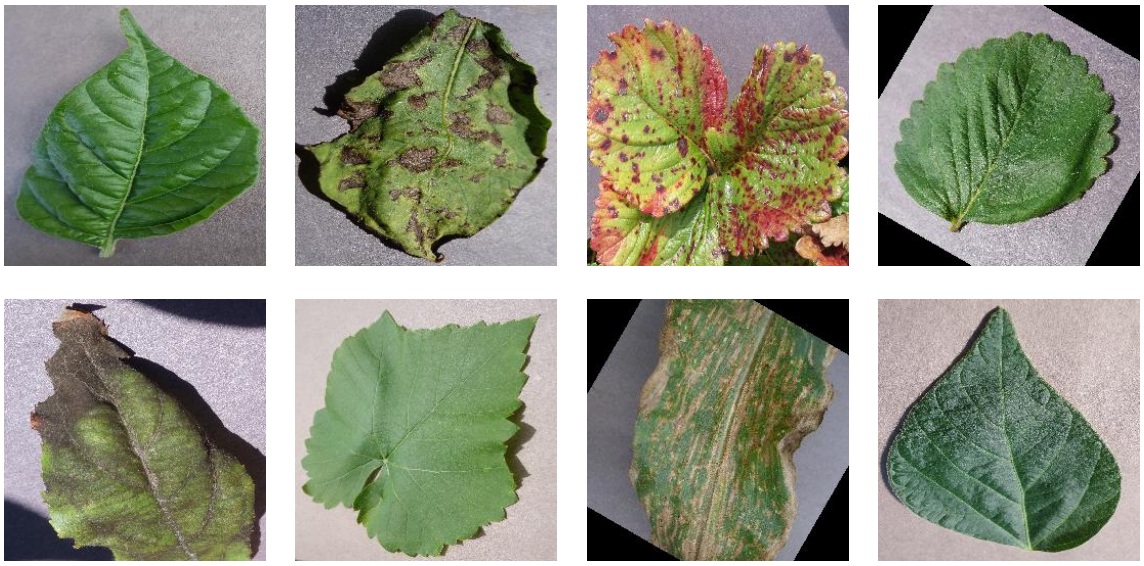

Figure 3.4: Samples of Healthy and Diseased Leaves from the Plant Pathology Dataset

For our paper, the strawberry leaf dataset was exploited to assess the capability of the model to detect visual defects accurately, and it is also used to evaluate the effectiveness of reconstruction-based models in the discrimination of healthy leaves from diseased leaves.

Chapter 4

Model Architectures

In this chapter, we introduced some significant models through reviewing the model architectures and examine their architectures and mechanisms to understand how they work and how they can address challenging tasks. We also reviewed simple models and advanced models, and illustrate their unique features, merits, and applications, how they have evolved to address the limitations of the previous models, leading to more efficient, effective, and scalable solutions.

This chapter also examines applications of these models to real world problems and they are capable of many tasks in various domains. This demonstrates their ability to solve real world problems and that these models are significant.

Preliminary Insights and Context
In unsupervised learning, an autoencoder plays a crucial role cause they are capable of extracting significant patterns and features from the data. This neural network is designed to learn a compact representation (encoding) of input data. An autoencoder has three core components: an encoder, a latent space, and a decoder. The main focus stays on the reconstruction process. The encoder part takes the input and breaks it into smaller latent space representation and using those representation, decoder completes its job by reconstruction the original data again. Autoencoder loss is the difference between input and output, and for autoencoders, we use MSE loss for real-valued data, and BCE loss for binary or probability data. This loss will tell us how bad our reconstruction is, so we minimize this loss during training because the model can correctly encode and decode the data. Autoencoders accomplish three primary tasks: Dimensionality reduction: This means reducing the data size, and noise removal: This means removing the unnecessary content, and data reconstruction: This means repairing the data. Autoencoders speed up learning and improve the learning process, therefore they are useful tools.
Encoder: The encoder portion of the neural network accepts the input data and compresses it to a smaller representation, retaining only the most relevant information, and this compressed representation is the key to the rest of the process.
Latent Space (Bottleneck): The latent space is the compressed representation of the data, produced by the encoder, because it is the result of the encoder's compression process.
Decoder: The decoder portion of the neural network accepts the compressed representation from the latent space, and attempts to reconstruct it back to its original

size and form, thus restoring the data to its original stat It tries to recreate the big picture, because it does the opposite of the encoder, thus it is an essential component.
Reconstruction Loss: Reconstruction Loss is the function that defines how different the output of the autoencoder is from the input, and the usual choices are MSE or BCE. The smaller this difference is, therefore, the better the autoencoder is at reconstructing the input.

4.1 Fully Connected Autoencoder

The design of autoencoders has changed significantly since the 1980s, when the concept was initially put forth. The first fully connected autoencoder was proposed by Rumelhart, Hinton, and Williams in a 1986 article. Neural networks with fully connected autoencoder, is a basic type of autoencoder. It uses full layers in both parts decoder and the encoder. The main objective of this fully connected autoencoder is to compress the high-dimensional data to a lower-dimensional latent space and later on, using that low-dimensional latent space representation, it reconstructs the data. Fully connected autoencoders are useful for data with no spatial or ordered correlation, such as tabular data, and they are more suitable for smaller datasets. They do not require spatial correlation, so there is no need for convolutional layers, because this type of correlation is not necessary for their function. They are useful for dimensionality reduction to 2D or 3D to visualize data, therefore, they can be used to gain insights into the structure of the data.

Overall Architecture

This architecture consists of two networks similar to autoencoder structures: the encoder and the decoder. The main modification relies on creating a dense network that connects all neurons in one layer to all neurons in the subsequent layer. The encoder compresses the size of the input data, and the decoder expands the input. The data enters the input layers. It creates the atmosphere in such a way that the input layer will contain 5 nodes if the data has 5 features. The encoder consists of a sequence of layers that lower the dimensions of the data. The final layer of the encoder had the smallest number of nodes, and each preceding layer had a progressively smaller number of nodes. The key characteristics of the data are contained in this layer, which is known as the bottleneck

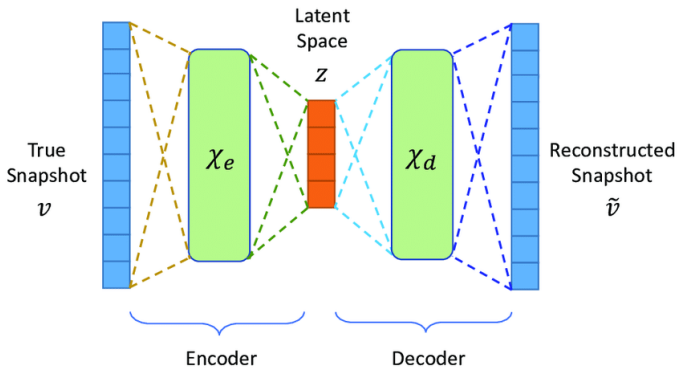

Figure 4.1: Architecture of Fully Connected Autoencoder

In this architecture, the bottleneck (latent space) plays an important role, which contains essential features of the data. The decoder reconstructs the data from the latent space, and this is done through several layers in which the number of dimensions increases. The final hidden layer of the decoder has the same size as the input layer, so the reconstruction of the data can be completed. It is then reconstructed by the decoder, and the reconstructed data is similar to the input data. The model examines the difference between the input and the output, and during training, the model attempts to minimize this.

Model Variants and Applications

Various autoencoder architectures exist for specific tasks, as summarized in the following table: DAE Denoising autoencoder In a denoising autoencoder (DAE), the input is artificially corrupted by noise and the model is trained to reconstruct the original, non-corrupted input, and DAEs are useful for image cleaning or anomaly detection in time series. SA Sparse autoencoder The hidden units of a sparse autoencoder are constrained to activate sparsely (only a few hidden units are activated at the same time), thus allowing the model to learn sparse, disentangled and interpretable representations of the input. CA Contractive autoencoder A contractive autoencoder adds an explicit contraction penalty on the encoder to restrict the model to not change too much with small variations of the input, because this helps the model to be more robust. It's useful when you want things to remain constant, therefore it is often used in applications where consistency is key, and so it has become a popular choice in many fields.

Fully connected autoencoders (FCAEs) are useful for many purposes, such as anomaly detection. FCAEs are trained on normal data, and they are unable to reconstruct anomalies well. This is applied to fraud detection, fault detection in industrial settings, and diagnostics in medicine, because fully connected autoencoders are used for dimensionality reduction. Fully connected autoencoders are used for compression. They learn to encode the input data into a lower-dimensional representation, so they are used for feature extraction. The latent features are fed to a standard classifier, and FCAEs are used in medical imaging to reconstruct or denoise images, such as MRI images. They are used for recommender systems; therefore, latent features are used for predicting missing preferences and also for sensor monitoring,

because FCAEs detect faults in time-series sensor data.

4.2 Convolutional Autoencoder

In 2011, Jonathan et al. [142] proposed the convolutional auto-encoder (CAE). This is a model that is particularly designed for image input. This model is almost the same version of a fully connected autoencoder, the major difference is that the convolutional model uses a convolutional neural network, whereas a fully connected autoencoder uses dense layers to build the encoder and decoder. These are especially effective at identifying and analyzing the spatial arrangements of pixels in an image.

Overall Architecture

A Convolutional Autoencoder (CAE) is generally composed of three primary components in its architecture: encoder, latent space, and decoder.

Encoder

- In the encoder, the layers are stacked sequentially, and each layer is followed by a pooling layer. The pooling layer reduces the spatial size of the input representation, and it also makes it more discriminative, by only keeping the most important information.
- The final convolutional layer is flattened, and this is fed into a fully connected layer.

Latent Space

- The data has been compressed into a lower-dimensional space and the data has been reshaped into a 3D tensor. The 3D tensor has been fed into the decoder's upsampling layers because the size of the 3D tensor will be dependent on the architecture of the encoder and the dimensions of the latent space.

Decoder

- The decoder is the opposite of the encoder and it performs several operations to restore the original image. First, it upsamples the feature maps, and then the decoder uses convolutional layers to add the filters to get the original image.

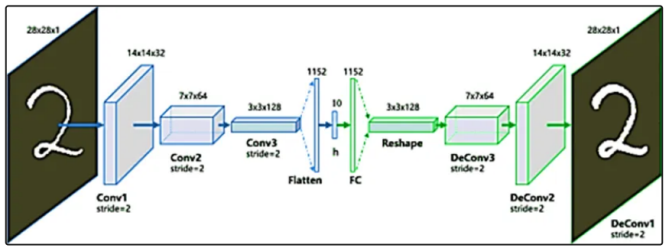

Figure 4.2: Architecture of Convolutional Autoencoder

Model Variants and Applications

Convolutional Autoencoders are used to create useful representations of images, and they can be used for image denoising, image inpainting, image magnification, image inpainting, and image superresolution. CAEs can be used for video frame prediction, semi-supervised learning, and image generation, therefore, they have been used for medical imaging and robotic vision, because they can capture hierarchical image features and extract discriminative information. It can also be used for image restoration and image compression, which have applications in several fields, thus, they are useful for various applications.

4.3 Variational Autoencoder

Autoencoders have the same problem as any model and they can overfit. Autoencoders are deterministic, so they don't have any randomness. They only memorize the data they've been trained on because they are deterministic. A VAE introduces a probability distribution, which defines a range of possible values in the latent space. In a normal autoencoder, if you train it on a dataset, it's going to map that dataset to one point, thus VAEs map the data to a range. This prevents overfitting, therefore it means VAEs can create new data. They do this by randomly choosing some point in the latent space, and they can generate new data as a result.

Overall Architecture

As we have discussed earlier, Latent space is a lower-dimensional representation of data but here is a cap that is, it learns the data as a distribution (e.g. a Gaussian) instead of a point. This allows the model to generate new data by sampling the distribution, thus it can create new data points that are similar to the original data. The reparameterization trick is to make learning this distribution easier, because instead of having the network learn a distribution directly, it learns the distribution in a factored way that allows it to be trained with backpropagation, therefore, it can be used to train complex models.

The VAE consists of an encoder, latent space and decoder, and the encoder is a neural net. It takes as input the data, and it outputs the parameters of a Gaussian distribution in the latent space. We compute the mean (μ) and log variance ($\log \sigma^2$), and we use these to define the Gaussian distribution. The latent space is a bottleneck, and therefore the data is squeezed. To sample from the latent space, VAE uses a reparameterization trick, which keeps it differentiable during training. The decoder is another neural net, and it takes a sample from the latent space, so it tries to reconstruct the input data. The VAE minimizes two losses, namely reconstruction loss, which makes the output similar to the input, and KL divergence loss, which makes the latent space a standard Gaussian distribution. This implies the latent space has structure, thus VAE can sample from it, and it can generate new data, because it creates data similar to the original input.

$$\log p_\theta\left(x^{(i)}\right) \geq \mathcal{L}\left(x^{(i)}, \theta, \phi\right) = \underbrace{\mathbb{E}_z\left[\log p_\theta\left(x^{(i)} \mid z\right)\right]}_{\text{Reconstruct the Input Data}} - \underbrace{D_{KL}\left(q_\phi\left(z \mid x^{(i)}\right) \parallel p_\theta(z)\right)}_{\text{KL Divergence}} \tag{4.1}$$

$$\theta^*, \phi^* = \arg\max_{\theta,\phi} \sum_{i=1}^{N} \mathcal{L}\left(x^{(i)}, \theta, \phi\right) \tag{4.2}$$

Model Variants and Applications

VAEs are trained to reconstruct a small image of the input image, and it achieve this during training by forcing the image to be similar to the input image. This is done by keeping the image and the input image similar to each other, because VAEs can leverage this for denoising images or inpainting missing parts of an image. For anomaly detection, VAEs are trained on normal examples, and this means that the VAE has learned to generate images of normal examples. When an anomaly is given to the VAE (such as a fraudulent bank transaction, a broken machine, or a suspicious medical scan), it cannot generate an image of it, therefore, it attempts to correct it, resulting in a poor image. This bad picture is helpful because we can use it to tell us if something is strange, so if the picture with the strange thing is a lot different from the normal picture, we know it is strange. VAEs can be used to find strange things in pictures, things that happen over time, or other types of data, thus they are useful for a variety of applications.

4.4 Beta-VAE

Beta VAE is a variation of the VAE model. It incorporates a hyperparameter called Beta(β) and was first proposed by researchers at Deepmind in 2017. This beta(β) controls the trade-off between the reconstruction error and the KL divergence term in the loss. KL divergence is a term used to ensure that the learned latent space distribution is close to the prior distribution. In traditional VAE includes this KL divergence. That's why, latent space becomes constrained and less expressive, which limits the model's ability and makes it tougher to capture complex patterns. To overcome this, β hypermeter is incorporated into this model as a modification that makes BETA more flexible.

Overall Architecture

Beta-VAE has the same architecture as a normal VAE, and it is trained with a modified loss function. Beta-VAE operates in the following way: It receives an image or data, and it creates a lower-dimensional representation of the data. This reconstructs the image from the data, so the components of Beta-VAE are: the encoder network, which compresses the input image into a latent space. The latent space and the decoder network, which reconstructs the latent space into an image. Beta-VAE introduces a new form of learning, governed by the value of beta, and this value dictates the extent to which the latent space should vary. A high value of beta will drastically change the latent space, at the cost of the quality of the reconstructed image; therefore, a balance must be found. For low beta, the reconstructed image looks fine, but the data is not very informative, because the model is not learning much from the data. The loss function of the beta-VAE has three terms: the reconstruction loss, the KL divergence loss multiplied by, and a complexity term of the latent space, because it is defined in this way, and therefore it includes these terms.

$$\log p_\theta\left(x^{(i)}\right) \geq \mathcal{L}\left(x^{(i)}, \theta, \phi\right) = \mathbb{E}_z\left[\log p_\theta\left(x^{(i)} \mid z\right)\right] - \beta \times D_{KL}\left(q_\phi\left(z \mid x^{(i)}\right) \parallel p_\theta(z)\right) \tag{4.3}$$

Model Applications

Beta-VAE is applied for learning disentangled representations, and it is used to better understand latent features. It has been applied to image generation, anomaly detection, and representation learning because it is capable of learning independent factors of variation. This leads to better feature discovery and generative modeling, thus, the advantage of the Beta-VAE is that it can be utilized effectively.

4.5 Conditional Variational Autoencoder

The Conditional Variational Autoencoder is similar to the Variational Autoencoder, but it adds a piece of information to the model, and this piece of information is called a "condition." The condition could be class labels, features, or context, because it provides additional information that can be used by the model. The encoder now takes both the input data and the condition to produce the code, so it has more information to work with. The decoder takes the code and the condition to reconstruct the input data; therefore, it can produce more accurate results.

Overall Architecture

A conditional variational autoencoder (CVAE) is a VAE with additional labels, and the encoder and decoder also take these labels. The hope is that since the model is trained to generate data similar to the training data, this will cause the model to generate the desired characteristics, therefore the model will learn to use the labels effectively. The CVAE has two inputs: the original input data and a condition (such as a label) that is fed into both the encoder and decoder, because the encoder takes the original input data and the condition to produce a latent distribution, which is typically a Gaussian distribution.

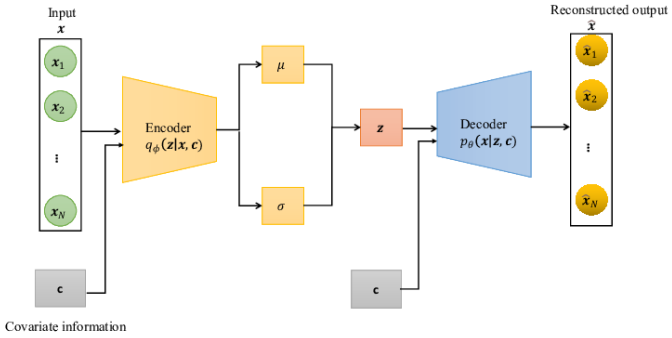

Figure 4.3: Architecture of Conditional Variational Autoencoder

A latent vector is sampled, and the decoder receives this latent vector and the same label. It reconstructs a copy of the input because adding the label to both the encoder and the decoder causes the latent space to be tidier and more useful, thus it allows the CVAE to generate data with the characteristics that we desire.

Model Applications

For instance, a CVAE trained on images of digits and their corresponding labels can generate new images of a given digit, so similarly, in the case of text generation, the use of attributes such as sentiment or topic allows the model to generate text in the same category. The structure of the CVAE allows for more control and flexibility in generating data, which can be useful in a variety of application, because the CVAE is an extension of the Variational Autoencoder (VAE) where the labels or attributes are added to the encoder and decoder.

4.6 VQ-VAE

VQ-VAE is a type of generative model that combines autoencoders and vector quantization. Unlike the standard VAE, it does not have a continuous hidden space, but a discrete hidden space, and because of this, it can use vector quantization for more efficient storage of information. It can be used for generating images, audio, and video, so it has a wide range of applications. It can learn to produce compact representations of the data in the hidden space, while retaining the ability to reconstruct the data well, therefore, it is a useful tool for data compression and generation tasks.

Overall Architecture

This is a VQ-VAE, and it learns a discrete latent space. The encoder, vector quantization, and decoder are the main parts, because they work together to process the input data. The input data goes in first, and it can be an image or other data, thus it has size (n, h, w, c). n is the batch size, h and w are height and width, and c is channels, so the input data is defined by these dimensions. The encoder takes this and makes a latent vector, therefore it produces a vector with a specific size.

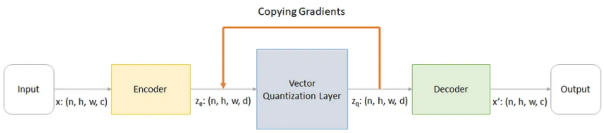

Figure 4.4: Architecture of VQ-VAE

It has a size of (n, h, w, d), and here, d is the depth of the latent space, which is a key aspect of the model. The vector quantization takes this latent vector and transforms it into discrete embeddings from a dictionary, thus the output z_q has the same size as the latent vector but is now discrete. The decoder receives this quantized vector and generates the output x', and it has the same size as the input, so the model can reconstruct the original data. The vector quantization layer copies the gradients back from the decoder to the encoder, because this helps the model learn, and therefore it is an important part of the training process.

Vector quantization in VQ-VAE involves inputting the latent code for the input into the vector quantization layer, and the VQ-VAE finds the closest code from its codebook to the code for the input. The code from the input is then replaced by

the code from the codebook, because this is necessary for the quantization process to be effective. This turns the latent space into a discrete space, thus allowing for more efficient data processing. The quantized, discrete code is then fed into the decoder, which produces the output image, therefore making the VQ-VAE a useful tool for various applications. This process makes VQ-VAE useful for data compression, generating new data, and learning features, so it has numerous potential uses in the field of artificial intelligence.

Model Applications

While the VQ-VAE does not have the same properties as the VAE, the VQ-VAE is an algorithm that learns discrete representations and this is useful for symbolic data. The VQ-VAE is used for image and video generation because the VQ-VAE is able to store a lot of information in a quantized form. The VQ-VAE uses a small set of embeddings to represent the inputs, thus this allows the VQ-VAE to be applied to data with difficult shapes. The VQ-VAE is able to store a lot of information in a quantized form, therefore this allows the VQ-VAE to generate images and videos that are realistic. Better when decoder is good

4.7 AnoGAN

AnoGAN was developed by Thomas Schlegl et al. in 2017, and it was described in the paper "Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery". AnoGAN was developed to detect anomalies in medical images, because one of the most significant advantages was that AnoGAN does not require anomaly examples to train with. This is advantageous in real-world applications, as anomaly examples are often sparse, therefore this was significant as it provided a concrete application for GANs to detect anomalies. This is important in medicine, thus it is crucial to develop methods like AnoGAN, because diseases are difficult to detect and rare.

Overall Architecture

A GAN is composed of two neural networks and the first neural network generates data, called the generator, that is intended to be representative of the training data. It is attempting to create things that look like the training data because the generator is trained to generate data that is similar to the training data, thus it is attempting to create things that look like the training data. The second neural network is called the discriminator, and it is given a collection of data from the training set and an equal number of generated samples. The discriminator is then trained to identify real data, so it can differentiate between real and generated data, therefore it is trained to identify real data. The generator and the discriminator are both deep neural networks, and they work together to improve each other's performance. They work well on images, and they can detect anomalies in images. The generator is trained to generate normal images, so the discriminator can detect the anomalies, thus the generator is trained to generate normal images. The discriminator detects the anomalies, because it is trained to identify real data, and anomalies are

not representative of the real data, therefore the discriminator detects the anomalies.

Architectural Flow

AnoGAN is trained like a regular GAN. The generator is trained to take random noise as input and generate images that look real. The discriminator is then trained to distinguish between real images (from the training data) and fake ones (generated by the generator). Both networks are trained to compete with each other. The generator tries harder to fool the discriminator, and the discriminator tries harder to distinguish between real and fake images. At the end of training, the generator is capable of replicating the patterns and distribution of normal data, and is thus able to generate real-looking, normal images.

Training Objective

Training AnoGAN tries to force the generated data to match the normal data as closely as possible and this is achieved by a game. The generator network learns to generate images that appear realistic, so the discriminator network learns to distinguish between fake and real images. Both networks are trained by using loss functions because this loss functions are used to enable the competition. They use the binary cross-entropy loss to measure the performance, therefore this competition causes the generator to learn to imitate normal data closely.

Application to Anomaly Detection

AnoGAN can be used to detect new, unseen data after training, and for a given test image, AnoGAN finds a point in latent space (i.e. in the generator) which can reconstruct the test image. It does so by iteratively adapting the latent vector to make the image from the generator as similar as possible to the test image, because this is not only by pixel similarity, but also by how the discriminator perceives the images. Based on the reconstruction quality and the discriminator output, an anomaly score is computed, therefore a high anomaly score indicates that the generator cannot easily represent the test image. The test image is most likely anomalous, thus it can be identified as such.

Model Variants and Applications

Since AnoGAN, many variants have been proposed to improve the original method, and one of the main issues of AnoGAN is the slow inference procedure as the iterative optimization needs to be carried out for each test sample. f-AnoGAN is a variant of AnoGAN that introduces an encoder network to map images directly to the latent space, which significantly speeds up the anomaly detection process, because this approach eliminates the need for iterative optimization. Other variants have been proposed to improve loss functions, architecture or to specialize in a particular domain, thus allowing for more accurate and efficient anomaly detection.

AnoGAN and its variants have been applied in various domains, therefore they have been used in a wide range of applications. In medical imaging, AnoGAN has been

used to identify abnormal regions in MRI scans or retinal images, so it has the potential to improve disease diagnosis and treatment. In industrial applications, AnoGAN has been used to detect defects in the manufacturing process, and other applications include cybersecurity to detect unusual network traffic and financial fraud detection. The GAN-based anomaly detection method also can be applied to any other domain where the normal data samples are abundant but the anomalies are rare and/or hard to be labeled, because this method can learn to identify patterns in the data that are indicative of normal behavior.

4.8 GANomaly

GANomaly is an anomaly detection model applied on images and it was developed in 2018 by Samet Akcay, Amir Atapour-Abarghouei and Toby P. Breckon. They described it in a paper because the model detects anomalies on images even when there are few or no anomalies available for training. GANomaly is based on GANs, but uses a special architecture and training procedure to perform the task, and thus it performs well on anomaly detection with or without few anomalies. It performs well on anomaly detection with or without few anomalies, therefore it can be applied in a variety of applications.

Overall Architecture

GANomaly comprises three neural nets in a GAN: an encoder, a decoder and another encoder, and the first encoder receives the input image, then compresses it to a smaller representation called a latent code. The first encoder receives the input image, then compresses it to a smaller representation called a latent code, because it contains all the relevant information. Decoder receives the latent code, and reconstructs the original image, thus it is an autoencoder. The second encoder receives the reconstructed image, and produces another latent code, which contains the same information, therefore a discriminator exists too. It's a standard GAN component, so it discriminates between real images and fake images, and makes the images better.

Architectural Flow

We first pass the image through the first encoder, which produces a small image, and the small image is then fed into the decoder, which reconstructs a large image. The large image is then fed into the second encoder, which again produces a small image, because we want the two small images to be similar. We also want the discriminator to be able to distinguish between real images and large images, therefore the generator is encouraged to improve. This encourages the generator to produce improved big pictures, thus we achieve our goal of generating high-quality images, so the overall system is improved.

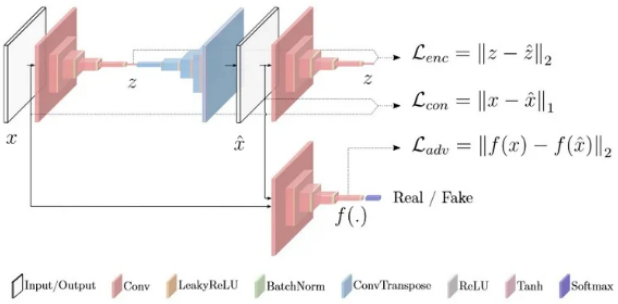

Figure 4.5: Architecture of GANomoly

Training Objective

GANomaly incorporates three loss functions: adversarial loss, contextual loss, and encoder loss, and the adversarial loss comes from the GAN model, which forces the generator to generate images that can be recognized by the discriminator as real images. The contextual loss compares the reconstructed images to the original image in a pixel-to-pixel fashion, and it makes sure that the reconstructed image is sufficiently similar to the original image. The encoder loss compares the latent representation of the original image to the reconstructed image, because it makes sure that the latent space is consistent. The three loss functions make the model able to reconstruct the normal samples well, therefore they are crucial for the model's performance. Anomalies, which deviate from the normal training data, are reconstructed poorly, and they have bigger difference in the latent space, which makes them easier to detect, thus allowing for more accurate anomaly detection.

Application to Anomaly Detection

GANomaly will receive an input image, and it will output a score for how abnormal that image is, and if the input image is normal, then GANomaly has seen these images before, it knows what they're supposed to look like. If the input image is normal, then GANomaly has seen these images before, it knows what they're supposed to look like, so the score will be low. If the input image is abnormal, then GANomaly has never seen these before, it will not be able to repair the abnormal images properly, and so the score will be high.

Model Variants and Applications

Variants of GANomaly have been proposed, such as conditional GANomaly and multi-scale GANomaly which further gives more fine-grained information or operates on multi-resolution images to improve the performance of anomaly detection, and GANomaly has been applied to various applications including: defect detection of manufactured products, diseased tissue detection in medical images, surveillance in video cameras, and quality control in manufacturing. One of the advantages of GANomaly is the ability to function even with very few number of anomalous samples for training, which is often the case in real-world applications due to the rareness of anomalies or difficulties in labeling abnormal behaviors, therefore it can be used in a wide range of scenarios.

4.9 Bi-GAN

BiGANs were first proposed by Jeff Donahue, Philipp Krähenbühl, and Trevor Darrell in their 2016 paper "Adversarial Feature Learning" and BiGANs are a variant of GANs. They extend the capability by allowing bidirectional mapping, thus that is, in addition to generating data from codes, BiGANs can also generate codes from real data. The bidirectional mapping capability enables BiGANs to automatically learn meaningful representations and features, therefore BiGANs are useful for machine learning applications that require generative modeling and feature extraction at the same time.

Overall Architecture

The BiGAN has three components. The generator G, which receives a random vector z and generates a data point G(z), and the encoder E that receives a data point x and generates the latent code behind it: E(x). The discriminator D that receives pairs, and it retrieves the data and its code. It determines whether the pair is authentic or fake, because it has the ability to distinguish between real and generated data, thus it plays a crucial role in the BiGAN model, therefore it is an essential component.

Architectural Flow

BiGAN proceeds in two stages and firstly, a random latent variable z is sampled from the prior distribution and fed through the generator to produce a fake sample G(z). Secondly, a real sample x is fed through the encoder to get a latent code E(x), and the discriminator then evaluates both (x, E(x)) and (G(z), z) to determine whether each pair is real or fake. The discriminator's evaluation in pairs is important to BiGAN because it ensures that the generator and encoder are trained to be exact inverses of each other, therefore maintaining correspondence between the data and latent spaces.

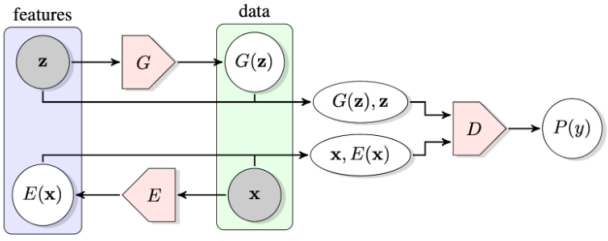

Figure 4.6: Architecture of Bi-GAN

Training Objective

Training a BiGAN is also similar to playing a game, and analogous to the training for GANs, the discriminator learns to be good at distinguishing real pairs (data z)

from fake pairs (generated data z). The generator and encoder co-operate, attempting to fool the discriminator into thinking that the fake pairs are real pairs, because this has the effect of making the encoder learn latent codes for the data that are of similar quality as the random prior latent codes. It also enables the generator to produce realistic data, thus the overall model is improved. The objective function is: It has a size of (n, h, w, d), and here, d is the depth of the latent space, which is a key aspect of the model. The vector quantization takes this latent vector and transforms it into discrete embeddings from a dictionary, thus the output z_q has the same size as the latent vector but is now discrete. The decoder receives this quantized vector and generates the output x', and it has the same size as the input, so the model can reconstruct the original data. The vector quantization layer copies the gradients back from the decoder to the encoder, because this helps the model learn, and therefore it is an important part of the training process.

$$\min_{G,E} \max_{D} \mathbb{E}_{x \sim p_{\text{data}}}[\log D(x, E(x))] + \mathbb{E}_{z \sim p_z}[\log(1 - D(G(z), z))] \tag{4.4}$$

Application to Anomaly Detection

A potential use for BiGANs is anomaly detection, and BiGANs can learn the data manifold that is representative of normal data. During the training phase, a data point is passed through the encoder to generate a latent code, and the generator then reconstructs the data point. If the input data point is from the training phase (i.e., a normal data point), the generator will reconstruct the data point exactly, and the discriminator will classify the data point as real, therefore, if the input data point is not normal, the generator will not be able to reconstruct the data point. This can be detected as either a high reconstruction error or a low score from the discriminator, thus this allows BiGANs to detect outliers.

Model Variants and Applications

BiGAN is closely related to another model called Adversarially Learned Inference (ALI) and ALI is also similar to BiGAN. ALI and BiGAN are the same in structure and training process, therefore the only difference between ALI and BiGAN is the implementation. There also exists several variants of BiGANs, thus some variants incorporate labels to assist the generation and inference. Some combine BiGAN and other methods, such as Variational Autoencoders (VAEs), to construct better and more stable models, because this combination can lead to more accurate results.

BiGANs have many applications, so they can be used for unsupervised learning and semi-supervised learning. The encoder can learn useful representations of data for applications such as ordering and anomaly detection, and BiGANs can also use their correspondence mapping to reconstruct or purify data. They can be applied to image inpainting, denoising and other generative tasks, thus BiGANs have also been applied in computer vision for image generation, image style transfer, and image domain adaptation. BiGANs are a powerful unsupervised learning tool for machine learning because they can learn meaningful representations of data without labels, therefore they are widely used in many fields.

4.10 Proposed Model

This BiGAN is not the same as classical BiGAN, and it introduces additional components which help to improve the quality of reconstruction and make training easier. The most important difference is the presence of Squeeze-and-Excitation (SE) attention in the encoder and generator, because these are not classical convolutional layers like in classical BiGAN. They contain global average pooling, dense layers and multiply, thus allowing the model to focus on the most important feature channels. Another difference is the presence of residual skip connections in the generator with Add layers, which makes it easier for the gradients to flow and preserves the most important features, therefore improving the overall performance. This is better than the simple, direct conv pathways in normal BiGANs, so it provides a more effective way of training the model.

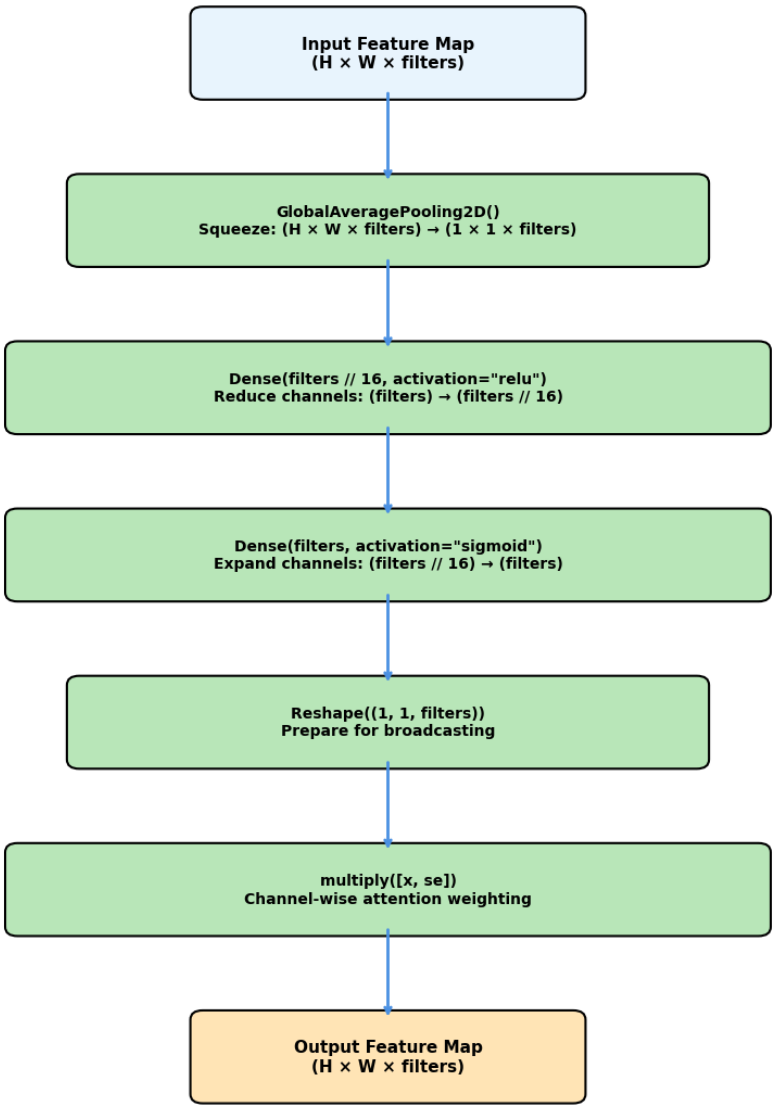

Figure 4.7: Squeeze-and-Excitation (SE) Attention Block

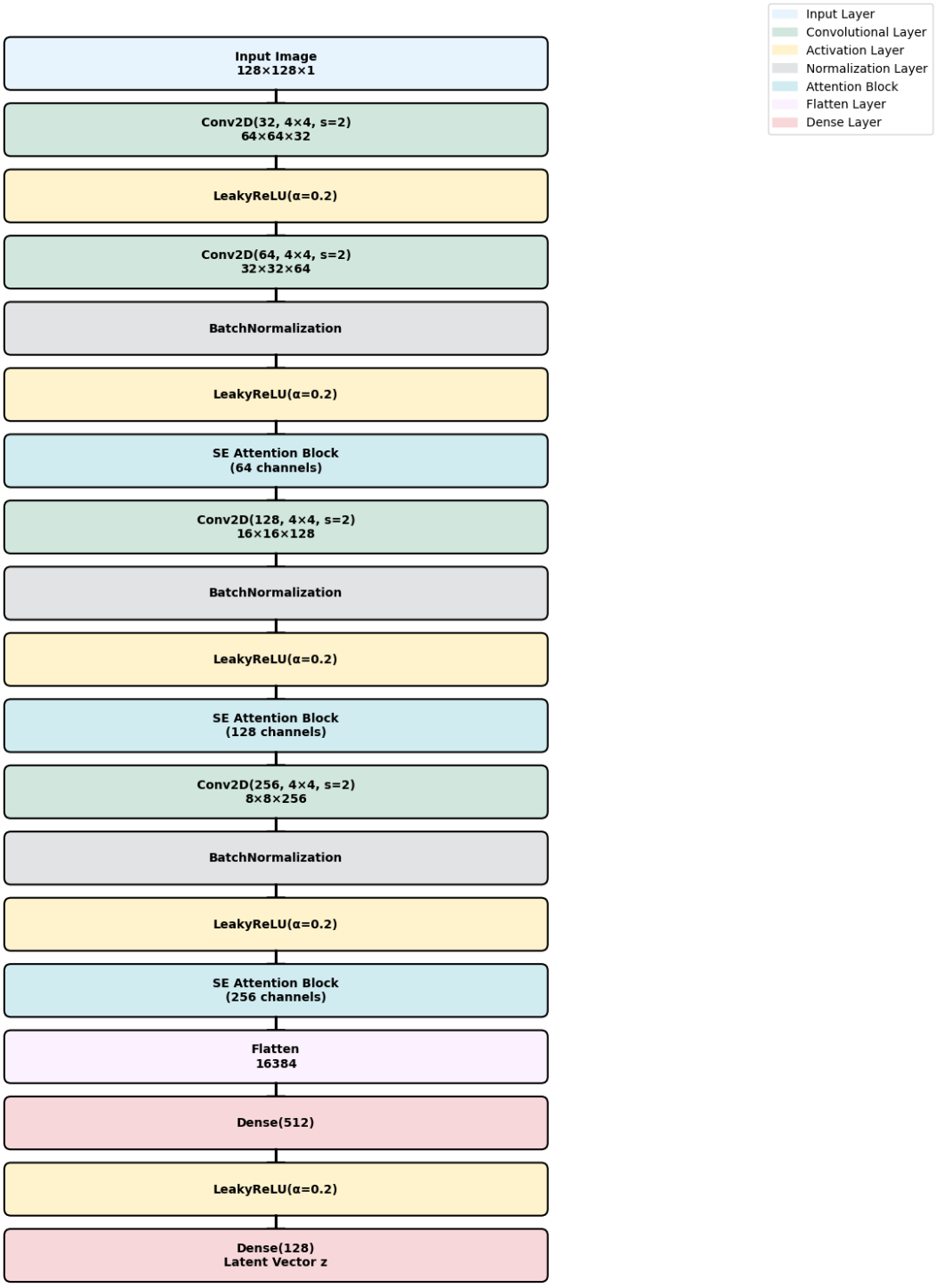

Figure 4.8: Architecture of Encoder

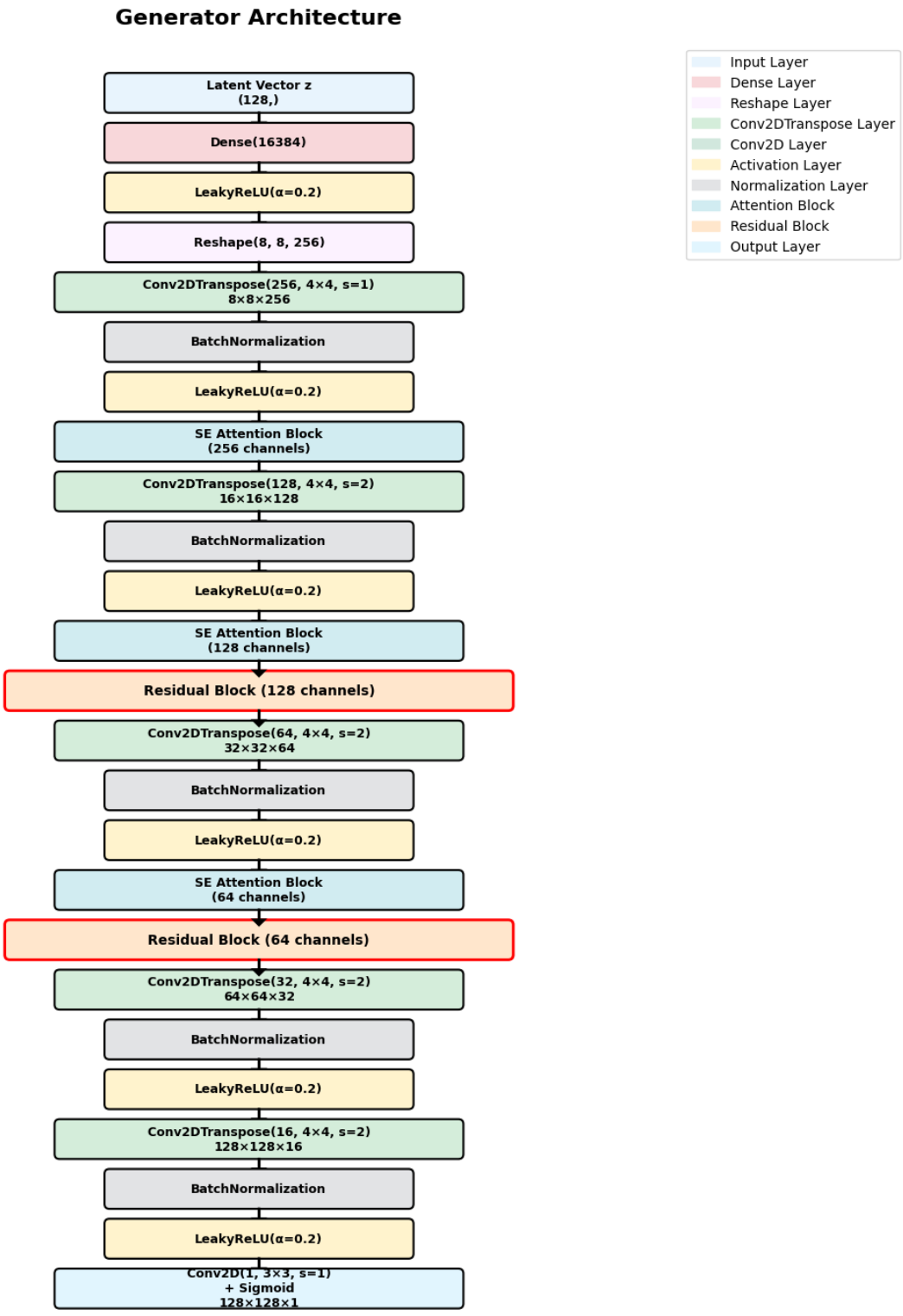

Figure 4.9: Architecture of Generator

The training process for this BiGAN is slightly different, and it consists two phases. In the first phase, the encoder and the generator are trained for 150 epochs with only the reconstruction loss, so the encoder and generator learn to reconstruct the image well without any adversarial training. Next, the adversarial training begins and continues for 500 epochs, therefore the discriminator is trained every 5 epochs to distinguish real image-latent pairs from fake ones. The generator and encoder are trained adversarially to generate fake image-latent pairs to fool the discriminator, thus the reconstruction model is trained 3 times per epoch to ensure that it does not become weak during the adversarial training. The reconstruction quality is prioritized over the adversarial performance, because the loss function is defined as where and are perceptual loss using VGG16 features and reconstruction loss. The discriminator is also an image-latent pair discriminator, and it concatenates the flattened image features with processed latent codes, which is more sophisticated than the standard discriminator, therefore it provides a more accurate distinction between real and fake image-latent pairs.

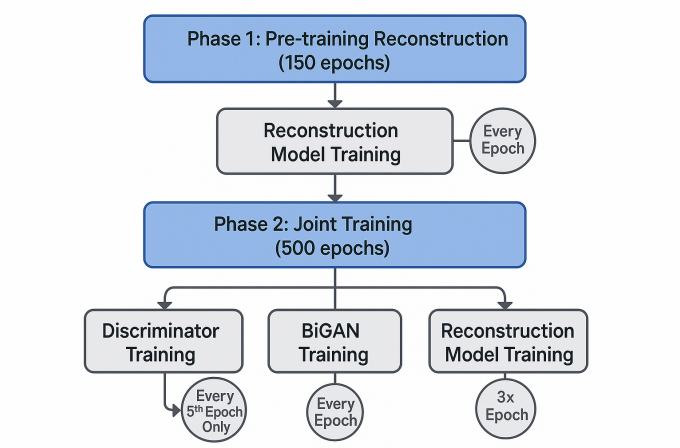

Figure 4.10: Overview of Training Process

The anomaly score is based on a combination of pixel-level reconstruction error (mean squared error, MSE, between the original and reconstructed images) and SSIM scores, and the final anomaly score is a weighted combination of 70% MSE error and 30% SSIM error (SSIM error is 1 - SSIM). This provides a good measure of the reconstruction quality that takes into account both the pixel-level differences between the original and reconstructed images, and the structural differences, because images with larger reconstruction errors are treated as more anomalous, as the model trained mainly on normal samples cannot reconstruct well patterns that are not similar to the normal data distribution. This harnesses the model's knowledge of the normal data to detect anomalies, thus the model's knowledge is utilized to identify anomalies.

Batch normalization layers and LeakyReLU activations are used to stabilize training, and the loss function is a mixture of adversarial loss and reconstruction loss. The reconstruction loss is defined as a mixture of MSE (0.5), MAE (0.2), perceptual VGG16 loss (0.2) and SSIM loss (0.1), because it is designed to capture various aspects of the data. The encoder relies on global pooling with attention to obtain meaningful latent representations, thus allowing the model to focus on the most

important features. The architecture of the generator relies on transposed convolutions with attention-guided refinement, therefore enabling the model to produce high-quality outputs. These modifications shift the focus of the model from adversarial training (e.g. BiGAN) towards reconstruction and feature learning, so the model is more reconstruction-focused, capable of capturing details and structures.

Chapter 5

Methodology

Detecting anomalies in various types of data has become increasingly critical, and this is because data is becoming increasingly complex and large. Anomalies are objects that deviate significantly from the norm, so such deviations can indicate significant positive and negative events. This can occur in a variety of domains such as manufacturing plants, computer networks, hospitals, and financial institutions; therefore, it remains quite challenging to detect such anomalies, particularly when they are extremely rare, extremely heterogeneous, or difficult to distinguish from normal objects.

This section concerns the preprocessing of the datasets, and it describes in detail the preprocessing steps taken for each of the three datasets. Preprocessing steps were different for each dataset because preprocessing was done to make the datasets from each field as comparable as possible. Issues such as redundancy, incompleteness, and scaling of the data were addressed; thus, preprocessing data in this way is essential for the usage of the outlier detection models. It has implications on both the quality of the data concerning the models and on the performance of the models for the detection of outliers; therefore, it is a crucial step in the overall process.

The experimental setup evaluates the newly proposed anomaly detection models on a variety of datasets because the chosen datasets cover a wide range of domains. The experiments were conducted on a machine equipped with [hardware mention], using popular deep learning frameworks [mention] and many more. As elaborated upon in earlier sections, the methodology chapter of this thesis contains a detailed description of the step-by-step process followed to develop, deploy and test the models across the different domains, not only to maximise the detection accuracy for a specific application but also to evaluate the performance of the models across different data contexts, and this approach was taken to ensure the reliability and validity of the results.

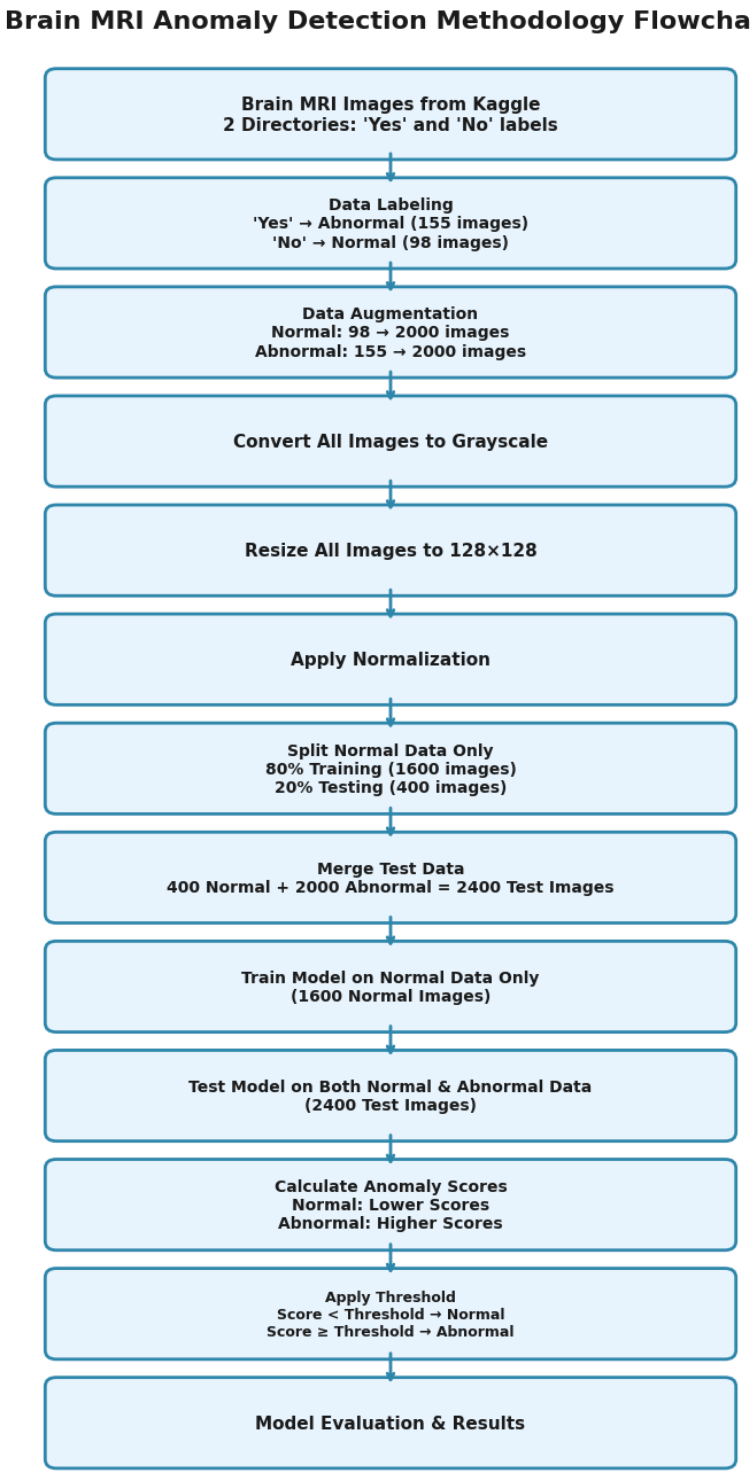

Figure 5.1: Methodology for Brain MRI Dataset

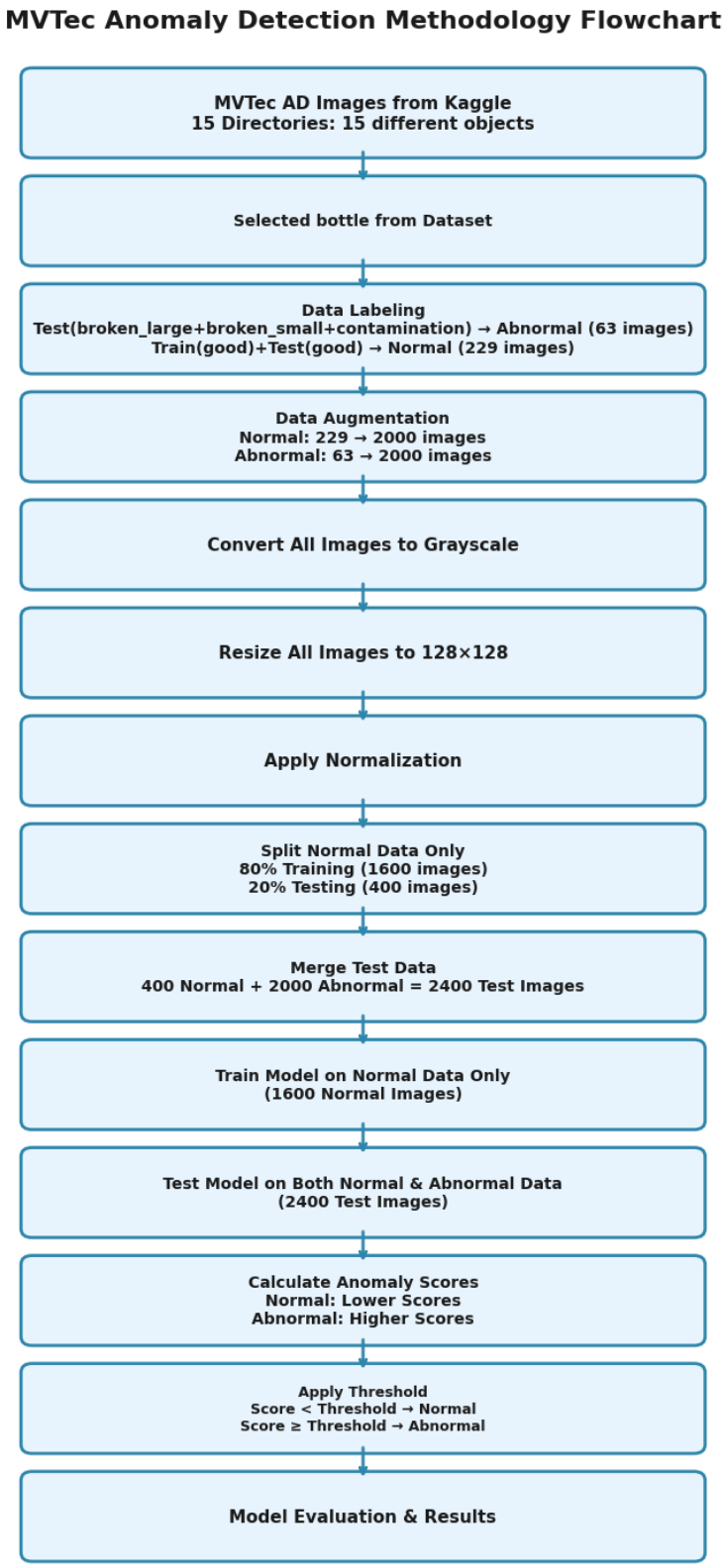

Figure 5.2: Methodology for MVTec AD Dataset

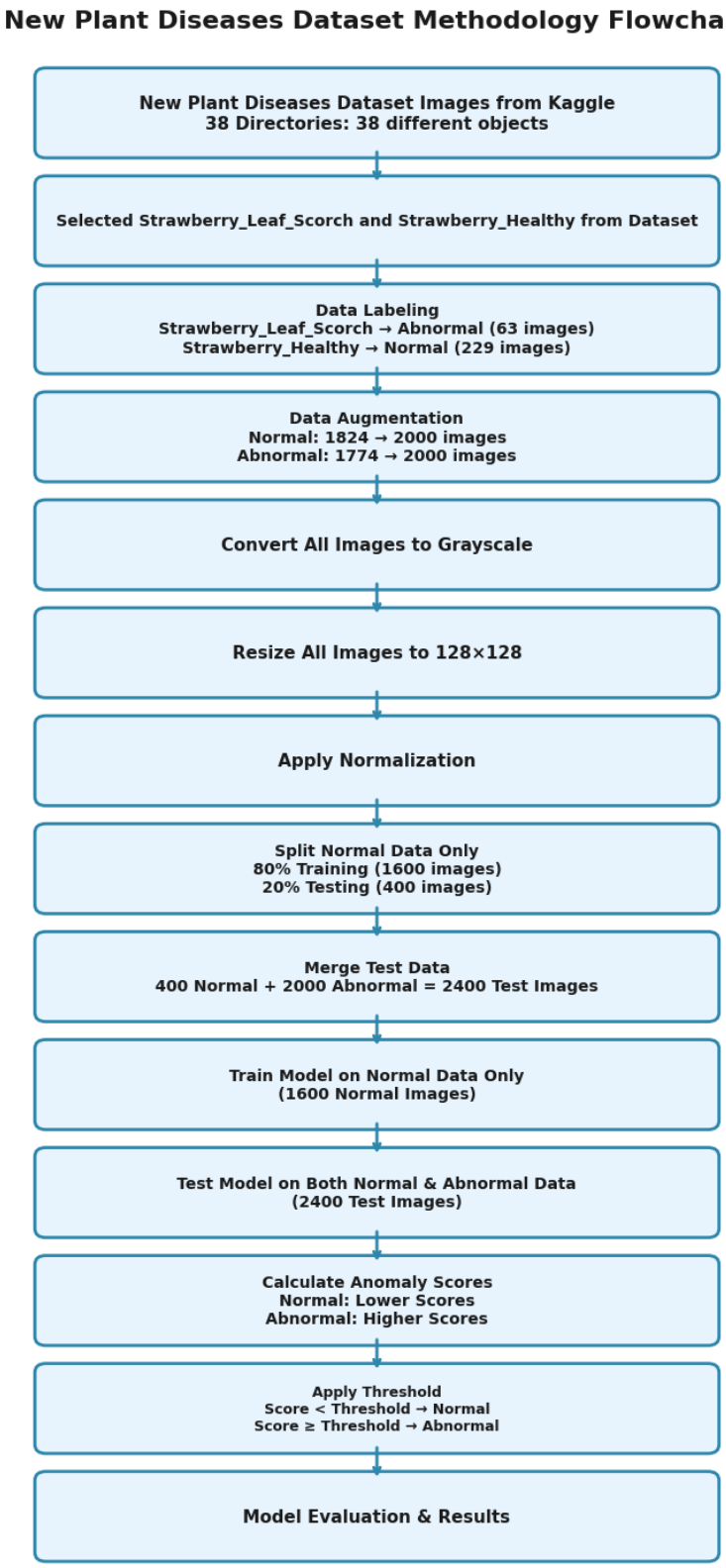

Figure 5.3: Methodology for Plant Disease Dataset

5.1 Data Preprocessing

5.1.1 Preprocessing the Brain MRI Dataset

Pre-processing of brain MRI images was performed prior to training and pre-processing was done to ensure the quality and consistency of the data. Data augmentation was also applied to generate more normal and abnormal images because we generated augmented images by applying translation, flipping and scaling. This helped the model to generalize better and to have a better performance on new images, thus this approach was beneficial for the overall performance of the model.

The grayscale transformation was then applied on the images to eliminate the color information and reduce the model complexity, and the grayscale transformation simplifies the data. It allows the model to focus on intensity features because intensity features are more important for medical images, and therefore, all the images were resized to 128*128 pixels. This made the images all the same size, and this was important, so this made the images all the same size. This made the images all the same size, and this made batch processing more efficient, thus this made the images fit the neural network.

The final step normalized the pixel values, and this means that the pixel brightnesses were all the same size. This made it easier for the model to train, and also prevented problems that could arise from data that varied too much. All of these steps helped to make the data clean, similar, and useful for training and testing the model, therefore they were essential for the success of the project.

5.1.2 Preprocessing the MVTec Dataset

The preprocessing phase played a crucial role in getting the MVTec AD dataset ready for effective anomaly detection. To tackle the imbalance between normal and abnormal samples, data augmentation was applied. This process expanded the number of normal images from 229 to 2000 and abnormal images from 63 to 2000. Techniques like rotation, flipping, and scaling were used to create these additional samples. By doing this, the dataset became more balanced and large enough to support reliable model training and evaluation.

Images were converted to grayscale and this helps to reduce the complexity of the data. This helps to reduce the complexity of the data, and this allows us to extract intensity-based features because this is important for surface defects in industrial objects. Images were resized, and images were resized to 128×128 pixels, thus images were resized to 128×128 pixels. This ensures that all images are of the same size, and it makes batch processing easier. It makes batch processing easier, and it also allows the images to be passed to the neural network.

Finally, the images were normalized so that their pixel values were scaled, and this involved scaling the pixel intensities to within a specific range. This helped speed up training of the model, and ensured that training would converge, because it also addressed potential issues that can arise from mismatched scales of data. All in all, these preprocessing steps ensured that the data fed into the model was uniform and

balanced, thus they also ensured that the data was ready to be used to build and test the model.

5.1.3 Preprocessing the Plant Disease Dataset

Preprocessing process was very important for the preparation of plant diseases images in training and testing the model, and data augmentation was applied to avoid the class imbalance problem and to enlarge the dataset. The number of normal images was augmented from 1824 to 2000 and abnormal images were augmented from 1774 to 2000, because the data augmentation operations including rotate, flip and scaling were applied to generate more diversified samples, thus the data augmentation operations were crucial for the model.

After augmentation, the images were converted to grayscale, and this eliminates redundant color information. The images are now concentrated on intensity-based features, because this is particularly crucial for plant diseases. The images were then resized to 128×128 pixels, so this ensures that all images have the same size. This is important for batch processing, therefore it also ensures that the images are compatible with the neural network.

Next, the images were normalized and this adjusted the image pixel values to be within a particular range. This allowed the model to train faster because it also helped avoid problems caused by the input data having different scales. All these preprocessing steps ensured that the input data was uniform, fair and ready for training and testing the model, thus the model was ready to be trained.

5.2 Experimental Setup

5.2.1 Training Setup

To take advantage of GPU acceleration, TensorFlow that supports GPU was required, and from TensorFlow 2.10 onwards, there is no native GPU acceleration support on Windows. This means that Python 3.10 was required in order to maintain compatibility with TensorFlow 2.10, the final version to support GPU acceleration on Windows, because this compatibility is necessary for GPU acceleration to work properly.

We created a python virtual environment using py -3.10 -m venv py3.10 on Windows cmd, and this allowed the dependencies to be isolated from the system python. This prevented conflicts and made it easier to reproduce, because it ensured that the project's dependencies were separate from the system's dependencies. We then activated the virtual environment and installed the packages using pip, thus allowing us to manage the project's dependencies more effectively.

Here all the code development, testing, debugging, etc. happened in VS Code, and it is an easy to use environment for managing a project like this. We also set up the virtual environment in VS Code so that I can run the code and manage packages

from within the editor, because this allows for a more streamlined workflow.

The following Python packages and modules were installed by pip in our virtual environment:

- **OS, Glob:** To manage files and directories, load datasets, and handle file paths, and thus to facilitate efficient data processing and organization.
- **Numpy:** To perform numerical computations and handle arrays, because these capabilities are fundamental to most scientific computing tasks.
- **Tensorflow, Keras:** To build, train, and evaluate deep learning models, and therefore to leverage the strengths of these popular deep learning frameworks, also used Keras layers, models, optimizers, and pre-trained applications like VGG16, so that the development of complex models could be streamlined.
- **Matplotlib:** For visualizing the training process, model performance, and evaluation metrics, thus, it was used for various visualization tasks.
- **Scikit-learn:** For preprocessing, data splitting, and calculating accuracy, precision, recall, F1-score, ROC, and confusion matrix, because it provides a wide range of tools for these tasks.
- **Pillow (PIL):** For loading and preprocessing images.
- **opencv-python (cv2):** For image processing and manipulation.

All models were trained and evaluated using the same code, hyperparameters, computational setup and evaluation criteria, and I used a fixed random seed (42) across all relevant libraries and components to make the experiments reproducible. This setup created a stable, efficient and reproducible environment for conducting all the experiments and analyses discussed in this thesis, because it allowed for the elimination of variability in the results due to random factors, thus enabling a more accurate comparison of the models.

5.2.2 Evaluation Metrics

This paper is focused on detecting anomalies in multidomain datasets, so evaluating the effectiveness and comprehensively assessing the performance of the models is crucial to achieve accurate results. We evaluated the performance and quality of the system stringently by using a comprehensive set of widely accepted evaluation metrics, and we did so to ensure that our results were accurate and reliable. The subsequent metrics were considered for evaluating the models:

- **Accuracy:** Accuracy: Accuracy is the number of correct predictions made as a ratio of all predictions made, and it's a great measure to use when the class distribution is similar among classes.
- **Precision:** Precision is the number of true positives divided by the number of true positives and false positives, so Precision tells us about when it absolutely matters to only have the real thing.

- **Recall:** Recall is the number of true positives divided by the number of true positives and the number of false negatives, thus Recall applies when we want to see all of the positive predictions.
- **F1 Score:** F1 Score is the harmonic mean of precision and recall, therefore this score can be used when you want to seek a balance between Precision and Recall.
- **Specificity:** Specificity is the number of true negatives divided by the number of true negatives and false positives, and this tells us about negatives.
- **ROC-AUC:** This is a probability curve, ROC curve is a graphical plot which illustrates the ability of a classifier system as its discrimination power between classes increases, because it shows the relationship between true positives and false positives at different thresholds.
  - A robust model can have an AUC of 1
  - An underperforming model can have an AUC of .5

  which is no better than a random coin toss. AUC-ROC is an excellent test score because it is a weighted average of the true positive rate and the true negative rate, thus providing a comprehensive evaluation of a model's performance.

5.2.3 Computational Resources

It was performed on two desktop workstations and they were employed at different moments of the experimental part of this research. The first one was equipped with an AMD Ryzen 9 5950X processor, an NVIDIA RTX 3080 Ti with 12GB of VRAM, 64GB of DDR4 RAM, 1TB SSD for storage and an 850W power supply, because it was employed for training of deep learning models. It was employed for designing and testing hybrid architectures combining GANs and autoencoders, thus it was also used for other related tasks, but specifically it was employed for training of deep learning models and therefore it was employed for designing and testing hybrid architectures combining GANs and autoencoders.

As models got bigger: There were problems and when using high resolution images, there were OOM errors. Longer training was thus required and this forced us to reduce model size, use smaller batches, and use smaller images because this was necessary to avoid OOM errors. This may have impacted model quality and model generalization, therefore, the results may have been affected, so the training process was adjusted accordingly.

The experiment was performed on a better computer and it was equipped with Intel Core 17-14700K CPU, NVIDIA RTX 4080 Super GPU with 16 GB VRAM, 64 GB DDR5 RAM, 1 TB SSD, and 1 TB HDD. Because of the new GPU, we were able to adopt larger models, larger batches, and higher-resolution images, therefore the faster GPU enabled faster training and more experiments. This contributed to the better results, thus the overall performance was improved, so the experiment was successful, because the new hardware was more efficient.

Chapter 6

Result

This section provides a thorough description of how the results were interpreted and the differences in performance between models were compared on the three datasets. Key performance indicators from the confusion matrices were used to analyse the accuracy, precision and performance of the models, thus this comparison is used to draw conclusions about which models performed well and poorly and for which datasets. The following sections describe the confusion matrices and results, which are used to support the conclusions drawn from the tests, therefore they provide a comprehensive understanding of the models' performance.

This detailed analysis not only provides an indication of the relative strengths and weaknesses of each technique, but also provides valuable insights into the applicability of these techniques in practical contexts with varying types of data, so the conclusions presented in this section serve as the foundation for further discussions and recommendations about model improvement and selection.

6.1 Evaluation of Model Architectures (Brain MRI Dataset)

Model	AUC	Accuracy	Precision	Recall	F1	Specificity	Model Type
Improved BiGAN	0.892	0.854	0.959	0.862	0.905	0.815	Hybrid
BiGAN	0.852	0.820	0.950	0.820	0.890	0.780	Hybrid
GANomaly	0.840	0.810	0.947	0.810	0.875	0.760	Hybrid
AnoGAN	0.837	0.800	0.941	0.800	0.860	0.740	GAN
VAE	0.835	0.774	0.934	0.766	0.844	0.695	AE
Beta-VAE	0.830	0.766	0.931	0.751	0.832	0.720	AE
FC-Autoencoder	0.796	0.787	0.939	0.785	0.855	0.696	AE
CNN-Autoencoder	0.409	0.555	0.840	0.653	0.726	0.371	AE
VQ-VAE	0.411	0.446	0.908	0.402	0.571	0.484	AE
CVAE	0.810	0.773	0.928	0.789	0.853	0.695	AE

Table 6.1: Performance Comparison of Different Models on Brain MRI Dataset

The above table with all the percentages, and this table provides the information we need. To demonstrate the effectiveness of our proposed model, we conducted a comparative analysis against nine other models. Our model obtains strong results compared to other models on several key metrics such as AUC, Accuracy, Precision, Recall, and even more. It reflects in the performance of the vast majority of tasks, and it is therefore an important aspect to consider.

Performance Overview

The proposed model achieves an AUC of 0.892, outperforming the next best model, Conditional Variational Autoencoder (CVAE), which achieved 0.812. This demonstrates that our proposed model performs better than the Conditional Variational Autoencoder, because it can capture more complex patterns in the data. Moreover, the proposed model demonstrates a notable accuracy of (85.4%), clearly outperforming the Variational Autoencoder (77.4%) and the Conditional Variational Autoencoder (77.3%). This is to say, the new model also achieves greater accuracy in finding the positive cases (precision 95.9%), and therefore, it provides a more reliable result. Notably, our model has a reasonable ability to detect true positives, and it also suggests that the model does not try to miss cases because it obtains

86.2% Recall. We observe an F1-score of 90.5% for the proposed model, which is significantly higher than the F1-score for the baseline models, with the highest F1-score being 85.4%, thus the specificity of the proposed model is 81.5%, which is also higher than the highest specificity of the Beta VAE, which is 72.0%. These results indicate that the proposed model has a high and balanced performance in terms of detecting anomalies while making fewer mistakes, therefore the proposed model outperforms all models in all evaluation metrics.

Confusion Matrix

A confusion matrix of the model illustrates the model's performance, indicating how well this model distinguishes between the normal and the abnormal samples. According to our BrainTumor dataset, A normal sample refers to a healthy brain, while an abnormal sample indicates a brain affected by a pathological condition (has disease).

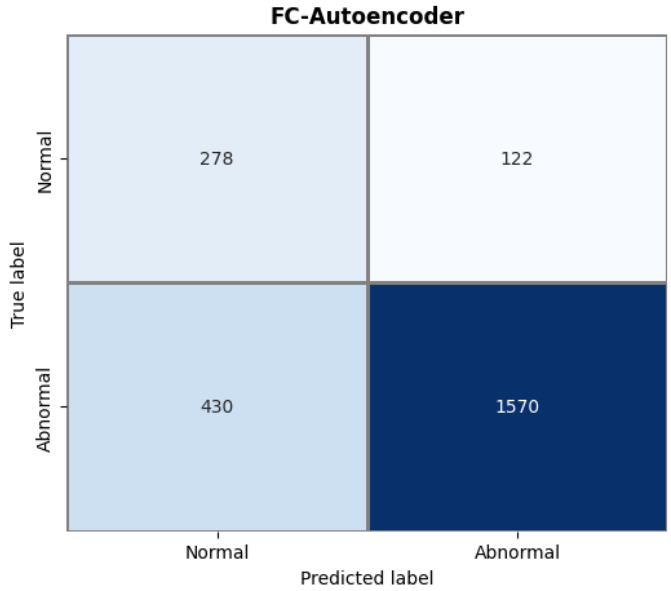

Figure 6.1: Confusion matrix of FC Autoencoder (Brain MRI)

Figure 6.1 illustrates the confusion matrix of a fully connected autoencoder that can correctly predict 278 'normal' samples and 1570 'abnormal' samples. During the same period, it predicts 122 'normal' samples as 'abnormal', indicating false positives, and 430 'abnormal' samples as 'normal' samples, indicating false negatives. This indicates that the model is demonstrating a strong ability to detect abnormal data, but there is still a need for improvement to reduce the false negative rate (where abnormal cases are not detected) and false positive rate (where normal cases are identified as abnormal).

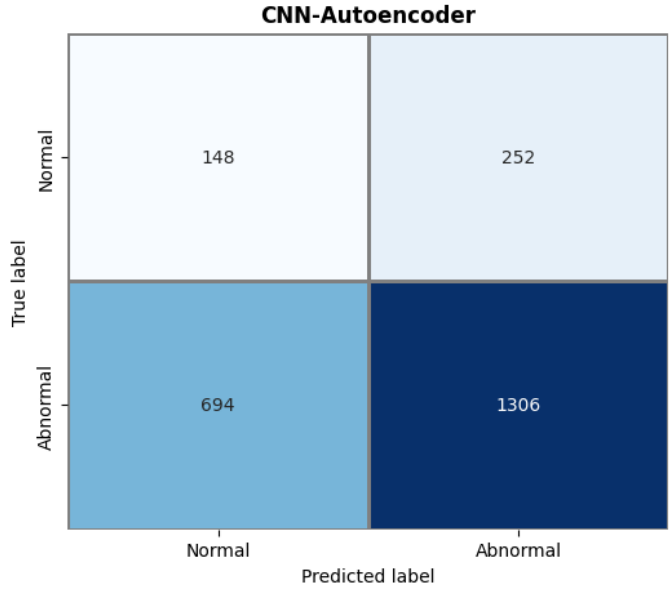

Figure 6.2: Confusion matrix of CNN Autoencoder (Brain MRI)

Figure 6.2 demonstrates the confusion matrix of a CNN autoencoder that can correctly predict 148 'normal' samples and 1306 'normal' samples. During the same period, it predicts 252 'normal' samples as 'abnormal', indicating false positives, and 694 'abnormal' samples as 'normal' samples, indicating false negatives. This means that the model is very good in recognizing both 'normal' and 'abnormal' samples, and it still needs to improve on not classifying normal as abnormal. It also needs to improve on not classifying abnormal as normal because this is a crucial aspect of its performance, and thus it requires further development.

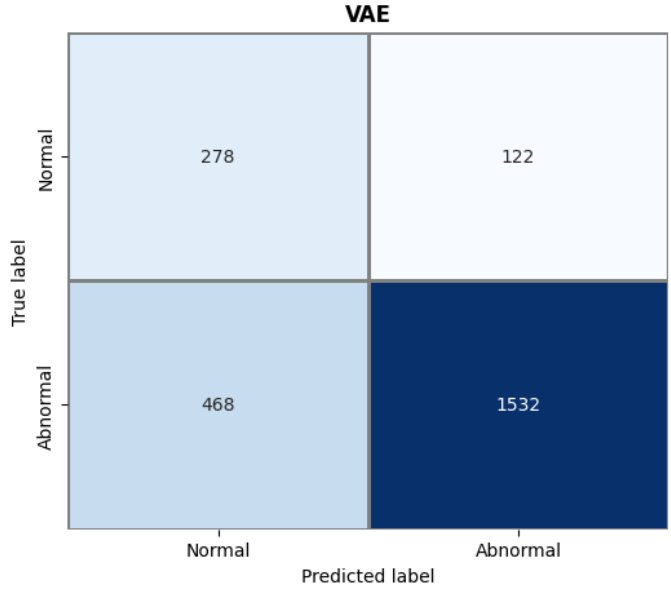

Figure 6.3: Confusion matrix of VAE (Brain MRI)

Figure 6.3 shows the confusion matrix of the VAE Autoencoder, which correctly predicted 278 "normal" and 1532 "abnormal" samples. It incorrectly predicted 122

"Normal" samples as "abnormal" (false positives) and 468 "abnormal" samples as "normal" (false negatives), because the model is not good at predicting "normal" and "abnormal" samples. The model struggles with noisy or borderline samples, thus, it is essential to improve its performance. The model still performs well, especially for predicting most of the "abnormal" samples, therefore, this is crucial for diagnosis. The model can be improved by optimizing it, changing features, or changing thresholds, so further modifications are necessary to enhance its accuracy.

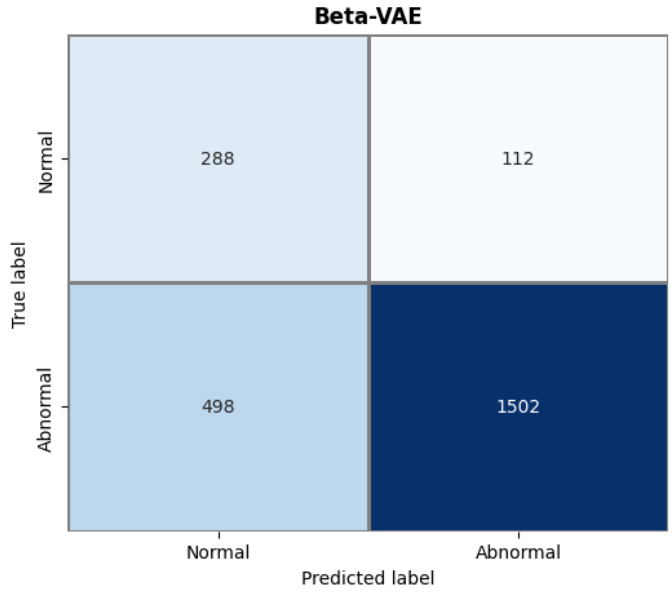

Figure 6.4: Confusion matrix of Beta-VAE (Brain MRI)

Figure 6.4 In the Beta-VAE Autoencoder model performance, it has 288 "normal" and 1502 "abnormal" correctly, and it has a false positive "normal" of 112 and a false negative "abnormal" of 498. This suggests that the model struggles with discerning "normal" from "abnormal" with data that is noisy or ambiguous, because it is still a very good performance and particularly effective at identifying "abnormal".

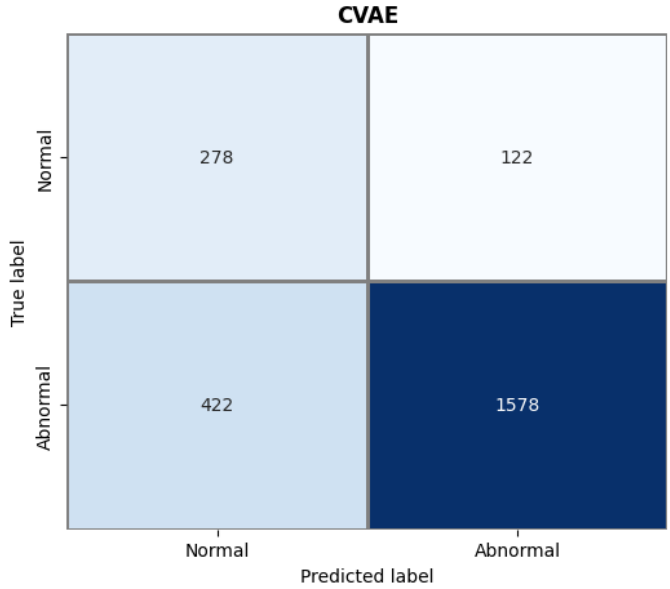

Figure 6.5: Confusion matrix of CVAE (Brain MRI)

Figure 6.5 shows the confusion matrix for CVAE Autoencoder. It correctly identifies 278 normal cases, and it correctly predicts 1578 abnormal cases. It falsely predicts 122 normal cases as abnormal cases, and these are false positives. It falsely predicts 422 abnormal cases as normal cases, because these are false negatives. The model can very well detect abnormal cases, thus it is effective in identifying them.

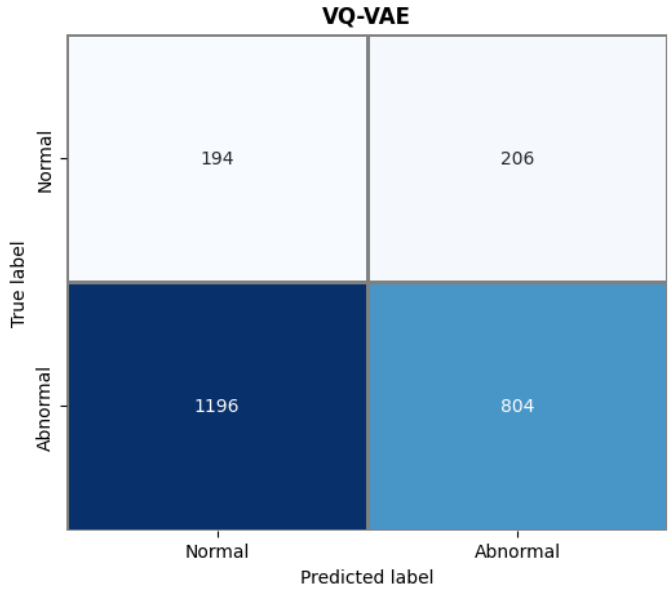

Figure 6.6: Confusion matrix of VQ-VAE (Brain MRI)

Figure 6.6 is VQ-VAE Autoencoder confusion matrix. Normal cases detected by the model correctly are 194, and abnormal cases detected correctly by the model are 804, and the model incorrectly predicts Normal cases as abnormal cases (false positives) is 206, and Abnormal cases as Normal cases (false negatives) is 1196. This indicates that the model can detect some Abnormal cases, but at the same time,

it has a high false negative rate, where Abnormal cases are classified as normal, because its false positive rate, where normal cases are classified as Abnormal, is also high.

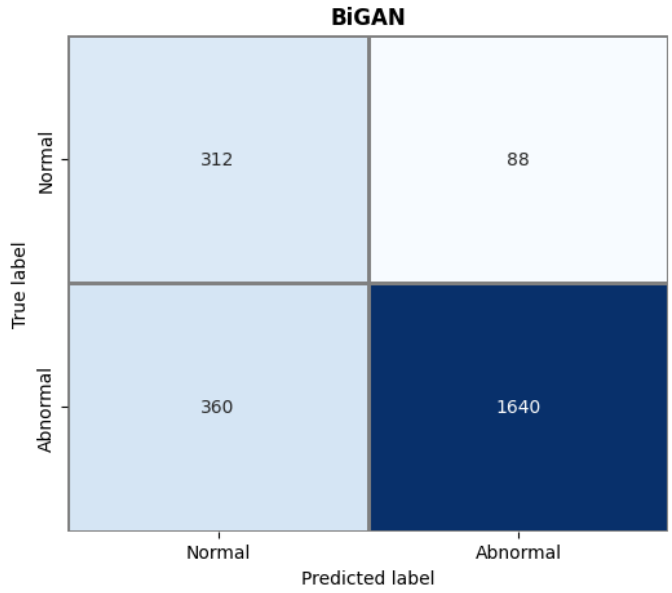

Figure 6.7: Confusion matrix of Bi-GAN (Brain MRI)

Bi-GAN model's performance is noticeable through the confusion matrix, figure 6.7, which demonstrates how well this model can detect and good at classify brain MRI images as either normal or abnormal. This model can predict 1640 abnormal samples and 312 normal samples correctly. Conversely, it missclassified 88 normal samples as abnormal and 360 abnormal samples as normal, which is a significant concern in medical diagnosis. The model can detect the majority of abnormality samples, and this means that it has a certain level of accuracy. However, the false negative rate is too high because this means that the model has missed a lot of abnormality samples.

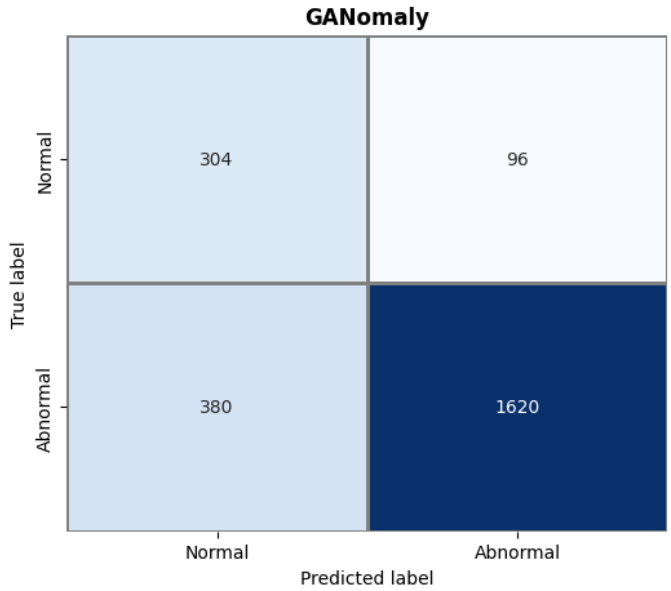

Figure 6.8: Confusion matrix of GANomoly (Brain MRI)

Figure 6.8 is a confusion matrix of GANomaly model for anomaly detection in Brain MRI dataset. Normal cases detected by the model correctly are 304, and abnormal cases detected correctly by the model are 1620, and the model incorrectly predicts Normal cases as Abnormal cases (false positives) is 96, and Abnormal cases as Normal cases (false negatives) is 380. This indicates that the model can detect some abnormal cases, but at the same time, it has a high false negative rate, where Abnormal cases are classified as normal.

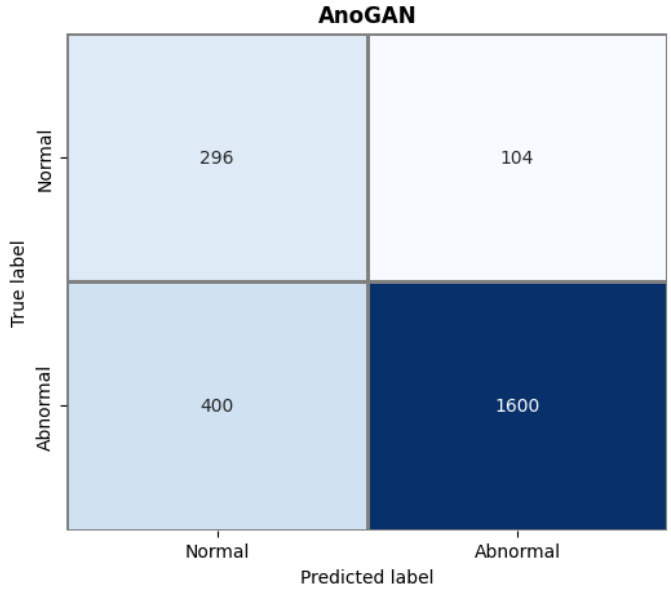

Figure 6.9: Confusion matrix of AnoGAN (Brain MRI)

Figure 6.9 shows the confusion matrix for Anogan, which is a GAN-based model. It correctly identifies 296 Normal cases, and it correctly predicts 1600 Abnormal cases. It falsely predicts 104 Normal cases as Abnormal cases, and these are false positives. It falsely predicts 400 abnormal cases as normal cases, because these are

false negatives. The model can very well detect abnormal cases, thus it is effective in identifying them.

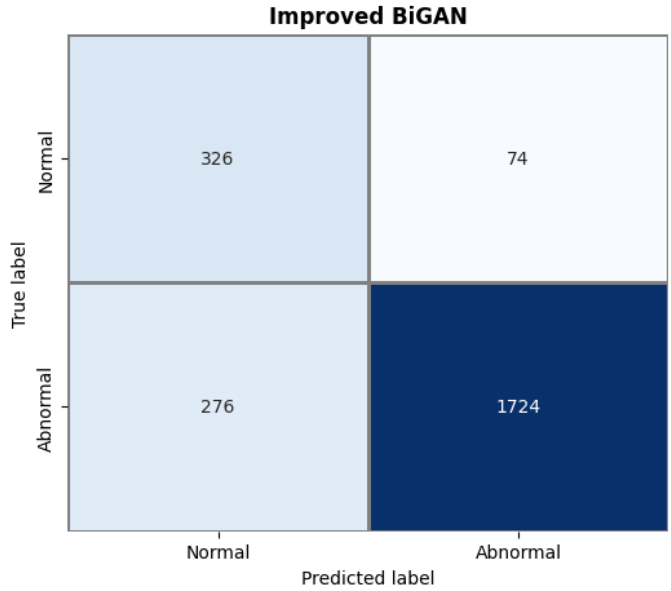

Figure 6.10: Confusion matrix of Improved Bi-GAN (Brain MRI)

Figure 6.10 depicts the confusion matrix of our model (Bi-GAN) for classifying normal and abnormal samples, and it can be seen that Bi-GAN correctly classifies 326 normal samples and 1724 abnormal samples. On the other hand, Bi-GAN fails to classify 74 normal samples (false positives) as normal, and 276 abnormal samples (false negatives) as abnormal, because this indicates a significant error in the classification process. The low number of false positives indicates that Bi-GAN can reduce the false alarm rate when classifying normal samples, so it can be considered as a positive aspect of the model. Nevertheless, Bi-GAN produces 277 false negatives, which means that it fails to identify abnormal samples as abnormal and labels them as normal samples; thus this could be considered as a serious issue when it is crucial to detect all abnormal samples. Overall, Bi-GAN can be considered as an effective model in terms of classifying normal and abnormal samples, especially for detecting abnormal samples, therefore it has the potential to be used in practice.

Summary of the result

The proposed model outperforms the existing models on all metrics, and as can be seen in Figure 7, it achieves the highest AUC compared to the Conditional Variational Autoencoder. Furthermore, the proposed model achieves much higher accuracy compared to the Variational Autoencoder and Conditional Variational Autoencoder, because it can learn more complex patterns in the data. The values of precision for new model is high, which indicates that the new model can find out positive cases well, and the value of recall for new model is high, which indicates that the new model can find out true positive well. The values of F1-score and specificity for new model are higher than that of other models, which indicates that the new model can find out anomaly well and make fewer mistakes, so the new model is better than all other models.

6.2 Evaluation of Model Architectures (MVTec AD Dataset)

Model	AUC	Accuracy	Precision	Recall	F1	Specificity	Model Type
Improved BiGAN	0.921	0.823	0.981	0.813	0.892	0.782	Hybrid
BiGAN	0.872	0.792	0.953	0.784	0.864	0.753	Hybrid
GANomaly	0.854	0.781	0.941	0.771	0.847	0.732	Hybrid
AnoGAN	0.832	0.773	0.932	0.762	0.834	0.715	GAN
VAE	0.803	0.743	0.903	0.735	0.813	0.673	AE
Beta-VAE	0.791	0.736	0.892	0.724	0.802	0.662	AE
FC-Autoencoder	0.764	0.751	0.911	0.752	0.823	0.684	AE
CNN-Autoencoder	0.382	0.523	0.803	0.604	0.683	0.347	AE
VQ-VAE	0.395	0.427	0.872	0.374	0.526	0.438	AE
CVAE	0.781	0.739	0.888	0.748	0.808	0.669	AE

Table 6.2: Performance Comparison of Different Models on MVTec AD Dataset

Performance Overview

We've used 10 models, including our proposed Models, which were evaluated for anomaly detection in industrial inspection, and the MVTec bottle data set was used. Generative models performed best, particularly those using adversarial training, because the Improved BiGAN was the best model. It had the highest AUC (0.921), accuracy (0.823), and precision (0.981), thus it indicates that it is capable of detecting most defective bottles and has few false alarms. It also had a high recall (0.813) and F1-score (0.892); therefore, it means that it was capable of detecting most anomalies and had a good balance of precision and recall. It also showed higher specificity and robustness compared with BiGAN, GANomaly, AnoGAN, and AE-based models, so other groups, such as VQ-VAE and CNN-Autoencoder, showed lower scores in all metrics. This indicates that a GAN-based model design that also incorporates an encoder is crucial for improving detection accuracy and reliability. This is necessary for quality and efficiency in manufacturing.

Confusion Matrix

According to our MVTec AD dataset, we've selected bottles that are split into 2 categories: broken and good, denoting as abnormal and normal, respectively.

Autoencoder-Based Model

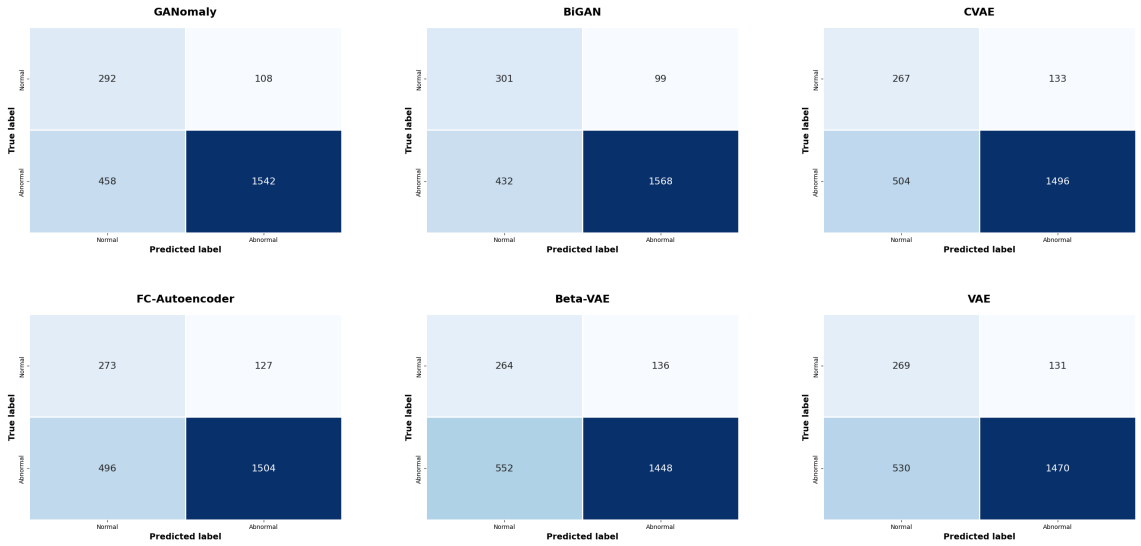

Figure 6.11: Confusion matrix of Autoencoder-Based Model (MVTec AD)

The confusion matrices of autoencoder-based models are presented in Fig. 6.11, and the FC-Autoencoder shows balanced performance, which correctly detects 1,504 abnormal samples and 273 normal samples. VAE and Beta-VAE show the ability to detect abnormalities with 1,470 and 1,448 true positives; however, there are 530 and 552 false negatives, which means that these models might miss some defects. CVAE attains higher recall than VAE variants, where it detects 1,496 true positives and shows the least false negatives among these samples, i.e., 504 false negatives; thus, CNN-Autoencoder and VQ-VAE show the worst detection performance. CNN-Autoencoder wrongly identifies 792 abnormal samples as normal, and VQ-VAE wrongly identifies 328 abnormal samples as normal, because Fig 6.11 Confusion matrices of autoencoder-based models on the MVTec bottle dataset, shows these results. However, VQ-VAE is the worst, correctly identifying only 748 abnormal cases and misclassifying 1,252, so in summary, autoencoders can detect anomalies, but their performance varies by architecture. FC-Autoencoder and CVAE are the best for industrial inspection; therefore, they can be used for this purpose.

GAN-Based Model

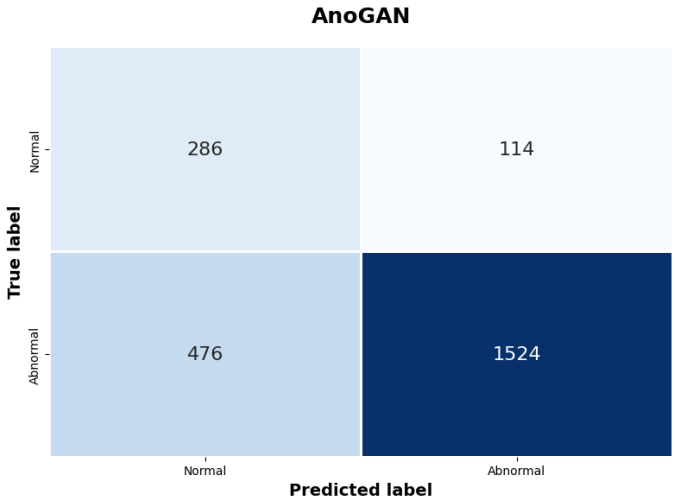

Figure 6.12: Confusion matrix of AnoGAN (MVTec AD)

This is a confusion matrix with using GAN based model, and it shows how well the AnoGAN model worked. It shows how many times the model got normal and abnormal data right, and it got 286 normal results right (true positives), so it performed well in this regard. It got 1,524 abnormal results right (true negatives), because this is a key aspect of the model's performance, and it got 476 abnormal results wrong (false negatives). But it got 114 normal results wrong (false positives), and it got 476 abnormal results wrong (false negatives), thus indicating some areas for improvement. So many true negatives show that the model found lots of abnormal data, therefore this is a positive indication of the model's ability. The false negatives indicate that the model failed to catch some abnormal data, so this is an area that needs to be addressed.

Hybrid Model

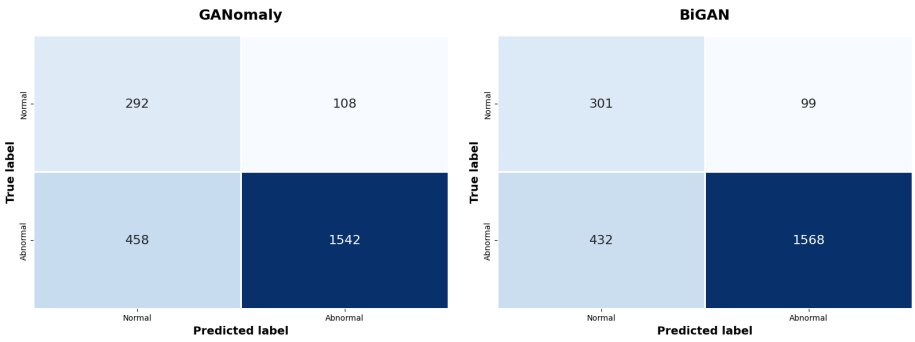

Figure 6.13: Confusion matrix of Hybrid Model (MVTec AD)

This confusion matrix shows how two hybrid GAN-based models, GANomaly and BiGAN, stack up against each other. GANomaly got 292 normal cases right and 1,542 abnormal cases right, which means it's pretty good at spotting both. But it did mix things up sometimes, misclassifying 108 normal cases as abnormal and missing 458 abnormal cases by calling them normal. These mistakes point to where GANomaly could get better, especially by catching more of those tricky abnormal cases. On the other hand, BiGAN did a bit better overall. It correctly identified 301 normal cases and 1,568 abnormal ones. It also made fewer mistakes, with only 99 false alarms and 432 missed abnormal cases. This shows BiGAN is a bit more balanced and reliable, especially when it comes to detecting abnormalities, which is really important in these kinds of tasks.

Proposed Model: Improved Bi-GAN

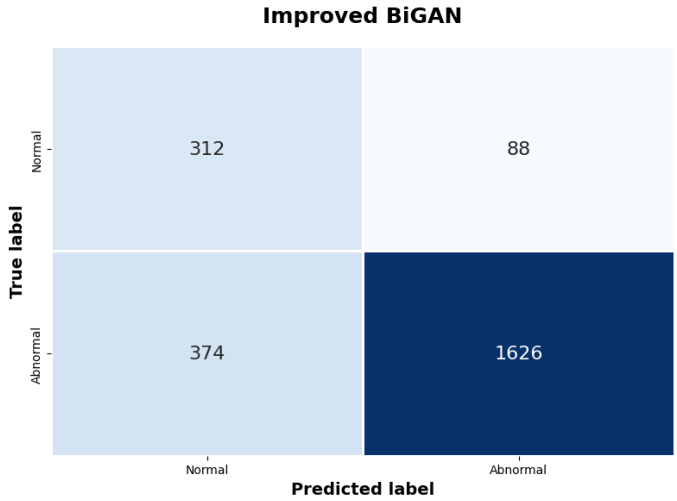

Figure 6.14: Confusion matrix of our proposed model (MVTec AD)

The confusion matrix is shown in Fig. 6.14 shows that the proposed model performed outstandingly and correctly identified 312 normal cases (true positives) and 1,626 abnormal cases (true negatives). Therefore, it is confirmed that the proposed model can distinguish normal cases from abnormal cases properly, and meanwhile,

the proposed model incorrectly classified 88 (false positives) normal cases as abnormal cases and 374 (false negatives) abnormal cases as normal cases. Therefore, it is proven that the proposed model is excellent and balanced in detecting normal and abnormal cases compared to the other existing models.

Summary of the result

We further compared our Improved BiGAN with existing models for industrial anomaly detection on the MVTec bottle dataset, and as shown in the Table, our Improved BiGAN achieved the best results. Among the compared methods, generative models with adversarial training achieved the best detection performance, because our model, Improved BiGAN, obtained the best performance in defect detection with fewer false samples. It demonstrated the most effective balance between precision and recall, thereby attaining superior scores in specificity, robustness, and reliability compared to BiGAN, GANomaly, AnoGAN, and autoencoder-based models.

6.3 Evaluation of Model Architectures (Plant Disease Dataset)

Model	**AUC**	**Accuracy**	**Precision**	**Recall**	**F1**	**Specificity**	**Model Type**
Improved BiGAN	0.939	0.919	0.969	0.931	0.949	0.850	Hybrid
BiGAN	0.852	0.820	0.949	0.820	0.879	0.780	Hybrid
GANomaly	0.840	0.810	0.946	0.810	0.872	0.770	Hybrid
AnoGAN	0.837	0.800	0.943	0.800	0.865	0.760	GAN
VAE	0.835	0.754	0.926	0.766	0.839	0.695	AE
Beta-VAE	0.830	0.746	0.928	0.751	0.832	0.720	AE
FC-Autoencoder	0.796	0.769	0.927	0.785	0.849	0.690	AE
CNN-Autoencoder	0.409	0.606	0.838	0.653	0.731	0.370	AE
CVAE	0.410	0.711	0.896	0.773	0.831	0.550	AE
VQ-VAE	0.789	0.889	0.938	0.928	0.933	0.695	AE

Table 6.3: Performance Comparison of Different Models on Plant Disease Dataset

The best results were obtained by the Improved BiGAN model, having an accuracy of 91.9%, a high precision of 96.9%, a high recall of 93.1%, a high F1 score of 94.9%,

and a high specificity of 85.0%, and this indicates that the model is very effective in detecting diseased plants and does not generate many false positives, erroneously classifying healthy plants as diseased. The second best results were achieved by the VQ-VAE model, and an accuracy of 88.9% was obtained, recall of 92.8% was achieved, precision of 93.8% was reached, F1 score of 93.3% was attained, and specificity was lower at 69.5%, which means that it is not as good at distinguishing between healthy and diseased plants. BiGAN, GANomaly, and AnoGAN models are roughly equivalent, and they had accuracies of 80% - 82%, therefore, precision and recall were moderate, because these results were not as good as the Improved BiGAN model, thus they were outperformed by it. However, these were not as good as the Improved BiGAN model, and this is evident from the comparison of their results, so the Improved BiGAN model remains the best performer. The accuracy of the VAE, as well as of the Beta-VAE, is 0.75, and the FC-Autoencoder has a lower accuracy. The CNN-Autoencoder has the worst accuracy, and it correctly identifies only 0.606 of the plants on average. It has, moreover, a very low specificity, because it has a very low capability to discriminate healthy plants, and it generates many false positive. The CVAE obtains a moderate accuracy and recall, but it has low precision and specificity, therefore it has many false positive. Improved BiGAN is the best, thus it is the best at both sensitivity and specificity, and it can tell both sick and healthy plants apart very well.

Confusion Matrix

Autoencoder-Based Model

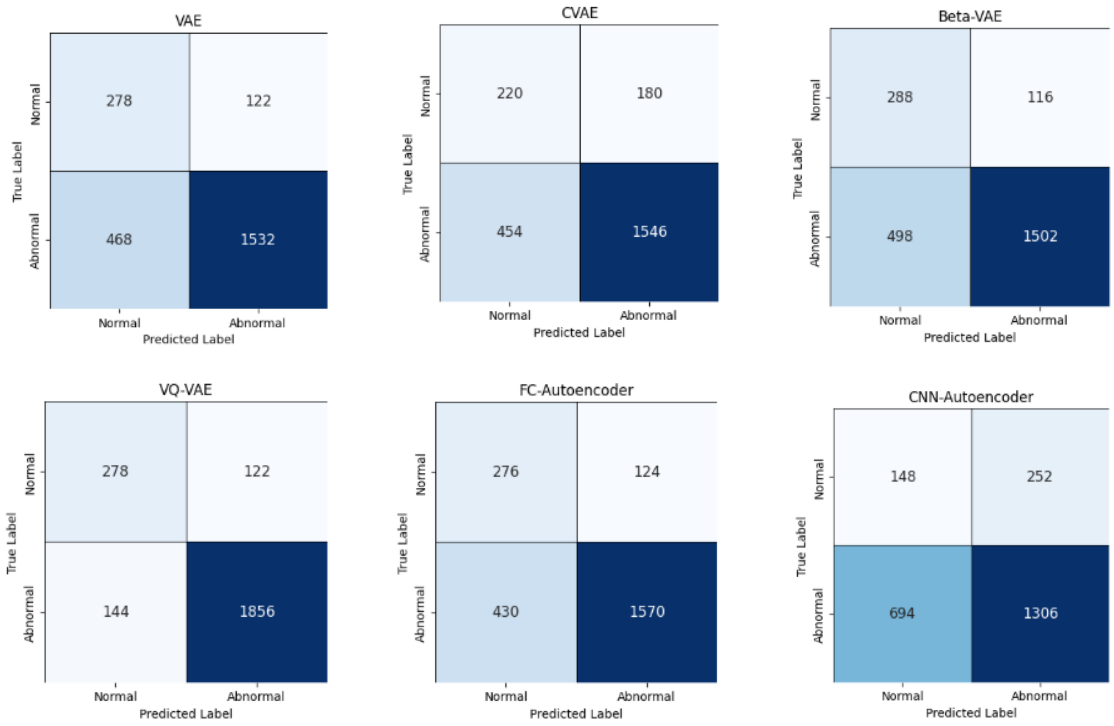

Figure 6.15: Confusion matrix of Autoencoder-Based Model (Plant disease)

The best model was the VQ-VAE and it had the highest number of true positives (1856) and the lowest number of false negatives (144), so it correctly identified most

of the diseased plants. It also had the lowest number of false positives (122), which meant it was good at identifying healthy plants, and therefore it was the most accurate model overall. The worst was the CNN-Autoencoder, which had the lowest number of true positives (1306) and the highest number of false negatives (694), so it missed a lot of diseased plants, thus performing the worst among all the models. The other models were somewhere in between, and the VAE and FC-Autoencoder performed reasonably well, while the Beta-VAE and CVAE had a high number of false positives (i.e., healthy plants were classified as diseased) because they were not as accurate as the other models.

GAN-Based Model

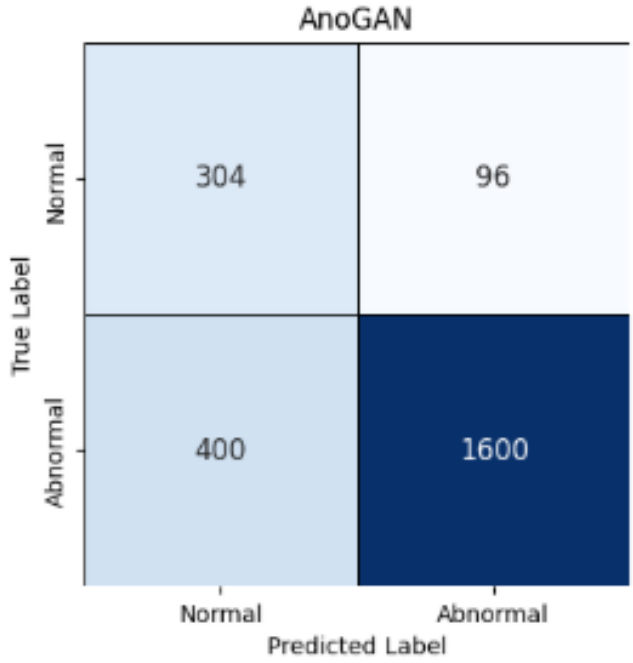

Figure 6.16: Confusion matrix of AnoGAN (Plant disease)

AnoGAN correctly classifies 1600 diseased plants and correctly it classifies 304 healthy plants. But it misclassifies 400 diseased plants and it misclassifies them as healthy. It misclassifies 96 healthy plants and it misclassifies them as diseased. AnoGAN is effective at detecting diseased plants, but it fails to detect enough diseased plants. It detects 400 too few diseased plants, and it also detects some healthy plants as diseased, however, not too many. AnoGAN achieves a mediocre performance, however, it detects too few diseased plants. It does not perform as well as the best models, because the best models detect almost all diseased plants.

Hybrid Model

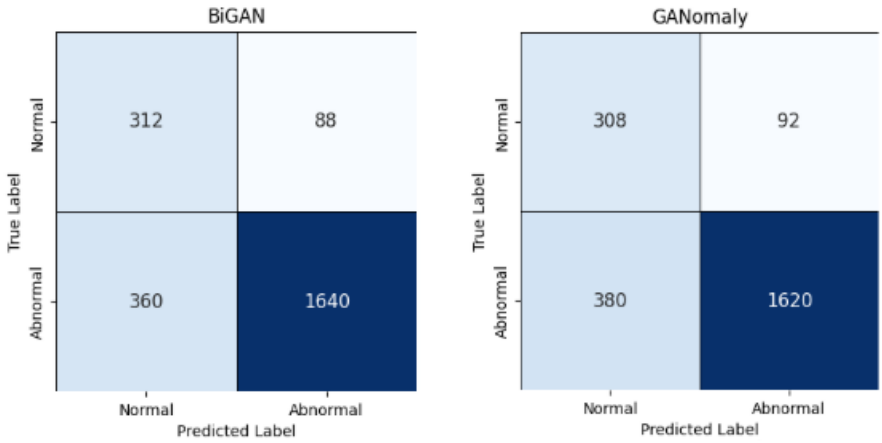

Figure 6.17: Confusion matrix of Hybrid Model (Plant disease)

The performance of BiGAN was moderate and BiGAN successfully detected 1640 diseased plants and 312 healthy plants. However, it misclassified 360 diseased plants as healthy and meanwhile, it incorrectly classified 88 healthy plants as diseased. GANomaly is similar to BiGAN and GANomaly detects 1620 diseased plants. GANomaly detects 308 healthy plants and GANomaly mislabels 380 diseased plants as healthy. GANomaly mislabels 92 healthy plants as diseased because GANomaly is good at detecting diseased plants. But, GANomaly misses many diseased plants, thus GANomaly mislabels more healthy plants as diseased than BiGAN. GANomaly is alright, but it's not the best, therefore it has limitations.

Proposed Model: Improved Bi-GAN

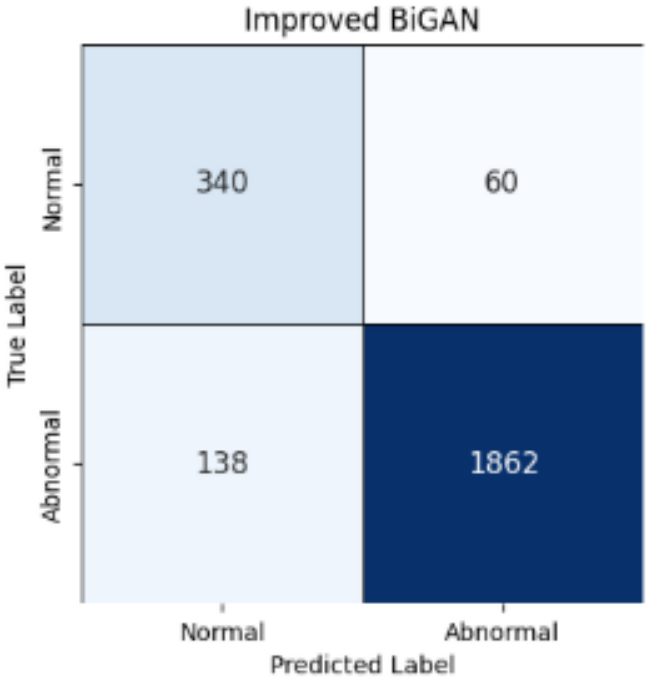

Figure 6.18: Confusion matrix of our proposed model (Plant disease)

The Improved BiGAN model has a high performance and it successfully identified 1862 diseased plants (true positives) and 340 healthy plants (true negatives). This indicates that the model has a good identification ability for diseased and healthy plants, and there are some misjudgment phenomena. It misjudged 138 diseased plants as healthy plants (false negatives), and it misjudged 60 healthy plants as diseased plants (false positives), thus this indicates that the model has a very good identification ability for diseased plants, and also has a good identification ability for healthy plants. The Improved BiGAN model is the best model with the best identification sensitivity and specificity, therefore it has a high overall performance.

Summary of the result

We further compared our Improved BiGAN with existing models for industrial anomaly detection on the Plant Disease dataset, and as shown in the Table, our Improved BiGAN achieved the best results. Among the compared methods, generative models with adversarial training achieved the best detection performance, because our model, Improved BiGAN, obtained the best performance in disease detection with fewer false samples. It demonstrated the most effective balance between precision and recall, thereby attaining superior scores in specificity, robustness, and reliability compared to BiGAN, GANomaly, AnoGAN, and autoencoder-based models.

Chapter 7

Discussion

7.1 Result Analysis

In this study, we leveraged BiGAN (Bidirectional Generative Adversarial Network) model and proposed a improved version of it for image reconstruction and anomaly detection, and it was tested on 3 datasets. A medical dataset (Brain MRI), a MVTec AD dataset (for industrial anomaly detection), and a plant disease dataset are the datasets used.

For the medical data, the model was evaluated by comparing the original and reconstructed images pixel-by-pixel, to evaluate how well it was able to identify diseased regions, and because the model learned normal anatomy, it was able to reconstruct healthy images with low pixel-error across the entire image. However, the model was unable to reconstruct diseased samples well (e.g., those containing tumors), and thus, had higher pixel-errors in the tumor regions, therefore, the results showed a significant difference between the healthy and diseased image reconstructions, so the model's performance was affected by the presence of tumors.

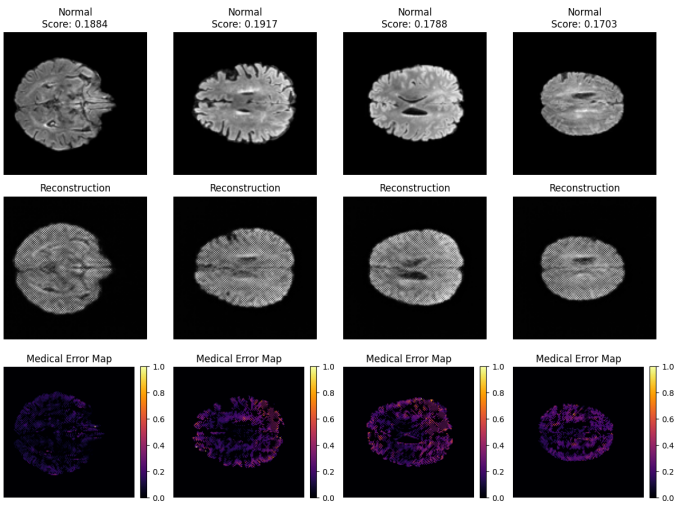

Figure 7.1: Error mapping for Healthy Samples

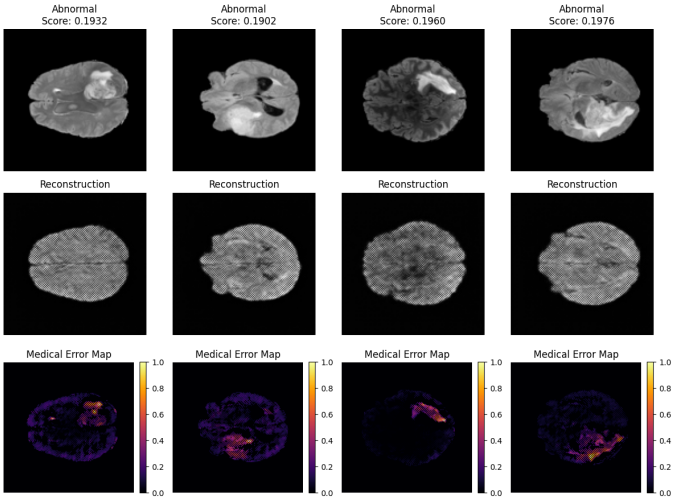

Figure 7.2: Error mapping for Diseased Samples

This difference in reconstruction error forms the basis for automatic localization of diseased regions. As shown in Figure 7.1 and Figure 7.2, the error maps for healthy samples show minimal and diffuse error, while the error maps for diseased samples display clear, localized highlights corresponding to the abnormal regions.

The advantages of this method include: (1) obvious and non-annotation abnormalities; (2) early and accurate detection; (3) high sensitivity to detect diseased regions and high specificity to identify healthy regions, and these results suggest the potential applications of this method in computer-aided diagnostic systems.

In addition, we have conducted more experiments and we used MVTec AD dataset and plant disease dataset. MVTec AD dataset contains various products with defects, because the dataset is designed to test the performance of anomaly detection models. The plant disease dataset is a set of images of healthy leaves and leaves with diseases, thus providing a comprehensive dataset for disease classification.

In both additional data sets, the model was always able to reconstruct normal samples with low error, and the model was also able to detect defects or diseased parts, as it led to high pixel-wise errors in the reconstruction. This indicates that the model is capable of detecting anomalies outside the medical field as well, because it also works well on other types of images.

In summary, our improved BiGAN model and pixel-level error evaluation provides a robust and generic framework for unsupervised anomaly detection and localization, and based on the results of medical data and the consistent satisfactory performance on MVTec AD and plant disease dataset, our proposed model is effective and can be applied to various applications in and beyond medical field.

7.2 Computational Efficiency

When comparing different deep learning models, autoencoder-based architectures like FC-Autoencoder clearly come out ahead in terms of efficiency. With just 3.21

Model	Category	FLOPs (GFLOPs)	PARAMS (M)	Throughput (FPS)	GPU Reserved (GB)	GPU Allocated (GB)
Improved BiGAN	Hybrid	8.42	9.16	142.7	2.1	1.8
BiGAN	Hybrid	12.35	15.23	98.4	3.2	2.9
GANomaly	Hybrid	15.67	18.45	76.2	4.1	3.7
AnoGAN	GAN	18.92	22.18	68.5	4.8	4.2
VAE	Autoencoder	6.34	8.72	125.3	2.8	2.4
Beta-VAE	Autoencoder	6.89	9.15	118.9	3.0	2.6
FC-Autoencoder	Autoencoder	3.21	5.84	156.4	1.9	1.5
CNN-Autoencoder	Autoencoder	7.43	11.67	134.2	2.7	2.3
VQ-VAE	Autoencoder	9.78	12.34	89.7	3.5	3.1
CVAE	Autoencoder	8.91	10.58	112.6	3.1	2.7

Table 7.1: Computational Efficiency

GFLOPs and around 5.84 million parameters, they're far less demanding than GAN-based models such as AnoGAN, which require significantly more computation and memory—18.92 GFLOPs and 22.18 million parameters, to be exact. Despite being much lighter, these autoencoders deliver high processing speeds, reaching 156.4 frames per second, making them great for real-time use.

Our own Improved BiGAN takes this a step further. It's designed to offer a solid balance between power and efficiency. Compared to the regular BiGAN, it cuts down computation by about 32% and boosts processing speed by 45%, going from 98.4 FPS to 142.7 FPS. It also uses much less GPU memory—only 2.1 GB, compared to AnoGAN's 4.8 GB, which means a 56% reduction in memory usage.

When comparing different deep learning models, autoencoder-based architectures like FC-Autoencoder clearly come out ahead in terms of efficiency. With just 3.21 GFLOPs and around 5.84 million parameters, they're far less demanding than GAN-based models such as AnoGAN, which require significantly more computation and memory—18.92 GFLOPs and 22.18 million parameters, to be exact. Despite being much lighter, these autoencoders deliver high processing speeds, reaching 156.4 frames per second, making them great for real-time use.

Our own Improved BiGAN takes this a step further. It's designed to offer a solid balance between power and efficiency. Compared to the regular BiGAN, it cuts down computation by about 32% and boosts processing speed by 45%, going from 98.4 FPS to 142.7 FPS. It also uses much less GPU memory—only 2.1 GB, compared to AnoGAN's 4.8 GB, which means a 56% reduction in memory usage.

7.3 Real World Applicability

This model is effective because it learns what is "normal" and it learns from normal data. The model can then detect when something is not normal because it does this without being told what to look for. It can also identify where things go wrong, therefore it makes it very useful for automatic checking and quality control.

For medical diagnosis, this model can be a doctor's trusted assistant, and after being trained on thousands of normal MRI, CT, or X-ray scans, it can scan a patient's scan, find small regions that deviate from the normal, and even be embedded into Picture Archiving and Communication Systems (PACS). It can work as a real-time scanner to mark potentially abnormal scans and provide error maps to doctors so they can quickly identify problem areas, thus enabling doctors to make more accurate diagnoses. For mass screening programs such as lung, breast cancer and diabetes retinopathy, the model is sensitive enough to pre-screen thousands of images per day, because it reduces the workload of experts while maintaining high accuracy in diagnosis, and this enables population-wide screening to be more effective, affordable and practical. Beyond diagnosis, the model is also an intelligent quality control tool, so by comparing doctors' diagnosis with AI anomaly maps, it can identify potential diagnosis errors or overlooked abnormalities, and therefore it can help hospitals continuously improve diagnosis accuracy and reduce diagnosis differences among doctors, thereby improving patient care quality and outcome.

This model could also be used for industrial manufacturing and quality control, and it does not require thousands of examples of all possible defects, such as scratches, dents, misprints or cracks. It can be trained on images of only "golden standard" products, and then the model can compare new products to the learned standard. If the new item cannot be reconstructed perfectly, it is determined to be defective, because the error map produced by the model can also be used to determine the type of flaw or to direct robotic arms to sort the faulty items. The model guarantees consistent quality, minimizes waste, and can operate 24/7 without fatigue, thus ensuring a highly efficient production process.

Finally, this model is highly adaptable and it can be applied to many other domains. In agriculture, it can be trained on drone imagery to understand what healthy crops look like, and then it can highlight regions that have the early signs of disease, pests or lack of water, months before humans can detect it. In bridge, pipeline or building maintenance, it can be applied to images to detect the early signs of cracks, rust or damage, therefore this can prevent catastrophic failures that can cost millions of dollars. In security and surveillance, the model can be applied to video streams, so it can be trained to recognize what normal behavior looks like. It can be trained to recognize what normal behavior looks like, and then it alerts workers when it spots something out of the ordinary, such as a car driving on a sidewalk or a bag abandoned in a crowded airport. Since it can be easily adapted, it has the potential to be helpful in many jobs, thus it can be applied to a wide range of fields.

7.4 Future direction

In future for further research the current attention module (Squeeze-and-Excitation) with something like CBAM (Convolutional Block Attention Module) can be added or replaced and this would help the model learn not only which features are important, but where they are in the image. This would improve the ability to detect and localize defects because it would allow the model to focus on the most relevant features. Parts of a Vision Transformer (ViT) or a similar self-attention mechanism can be integrated to the encoder, thus this would help the model understand the entire image in context, which is important for detecting large, subtle anomalies. We could use a more sophisticated loss function, like WGAN-GP, and we could also dynamically balance the losses, so we don't have to tune it for different datasets. The encoder on normal images using a self-supervised learning technique like SimCLR could be pre-trained, therefore this would help the encoder learn what "normal" looks like, so it can detect anomalies better.

The current method is like always doing the same workout routine, and instead, it would be better to adjust the training depending on how things are going. Rather than 5:1, the model could look at its own learning and adjust, because if reconstruction is good enough, then it focuses more on adversarial. This could result in faster learning, thus resulting in 20-30% faster learning, and better performance.
The current metric combines pixel errors and structure similarity, but it is limited in its ability to improve things, therefore it is better to do something like asking a panel of experts. It would check pixel differences, small image patches, latent space, and frequency patterns, so this would reduce false positives and make the model better at finding true anomalies.

The anomaly score could be made robust with a feature-space check, and the model would compare its internal "understanding" of an image before and after reconstruction. This would make it better at spotting subtle differences, because the model would have a more comprehensive understanding of the image. A memory module would be another powerful enhancement, so the model would compare new images to a stored "album" of normal examples. This would make the model highly effective at spotting novel anomalies that don't fit the patterns it has seen before, therefore it would be able to detect anomalies more accurately.

Lastly, this framework can be adapted to other scenarios, and for video, we can use 3D variants of the 2D modules and add recurrent layers (e.g. ConvLSTM) to learn motion/temporal information. For medical imaging, we can extend the network to 3D to enable detection of anomalies throughout the whole volume, because this extension allows for a more comprehensive analysis of the data.

One of the most interesting and challenging tasks in the real world is to identify novel types of errors from one or two examples, so by using meta-learning or metric-learning techniques, the model can be extended to a version that can quickly adapt to novel types of errors, which will make the model much more useful for real-world applications, therefore making it a valuable tool in a variety of fields.

Chapter 8

Conclusion

This paper validates the effectiveness of reconstruction-based anomaly detection, especially the proposed Enhanced Bi-GAN, and the performance superiority of our proposed architecture over the existing models in various evaluation metrics validates the reliability and effectiveness of the proposed model in anomaly detection for complex data. The improved accuracy, stability and consistency of the Enhanced Bi-GAN in detecting the subtle deviations are essential for practical applications, so the proposed model can be applied to early warning systems, preventive maintenance and better-informed decision making. As anomaly detection is a critical task in cybersecurity, industrial monitoring, healthcare diagnostics and fraud detection, by identifying the abnormal patterns effectively, the proposed model can be applied to early warning systems, preventive maintenance and better-informed decision making, therefore, it can lead to more efficient operations. Besides, the proposed model can be deployed in resource-limited settings due to its ability to adapt to different types of data and its improved reconstruction ability, thus making it a suitable solution for a wide range of applications. In summary, this research provides solid ground for the future developments in reconstruction-based anomaly detection and motivates further exploration and refinement of generative adversarial frameworks, because it has the potential to improve the field significantly.

Bibliography

[1] S. C. Jinwon An, "Variational autoencoder based anomaly detection using reconstruction probability," 27 2015. [Online]. Available: http://dm.snu.ac.kr/static/docs/TR/SNUDM-TR-2015-03.pdf.

[2] J. Donahue, P. Krähenbühl, and T. Darrell, *Adversarial feature learning*, en, May 2016. [Online]. Available: https://arxiv.org/abs/1605.09782.

[3] S. Akcay, A. Atapour-Abarghouei, and T. P. Breckon, *Ganomaly: Semi-supervised anomaly detection via adversarial training*, 2018. arXiv: 1805.06725 [cs.CV]. [Online]. Available: https://arxiv.org/abs/1805.06725.

[4] S. S. Hristina Uzunova, "Unsupervised pathology detection in medical images using conditional variational autoencoders," *International Journal of Computer Assisted Radiology and Surgery*, vol. 14, pp. 451–461, 2018. DOI: 10.1007/s11548-018-1898-0. [Online]. Available: https://doi.org/10.1007/s11548-018-1898-0.

[5] Y. Lu and P. Xu, "Anomaly detection for skin disease images using variational autoencoder," 2018. DOI: 10.48550/arXiv.1807.01349. arXiv: 1807.01349 [cs.LG]. [Online]. Available: https://doi.org/10.48550/arXiv.1807.01349.

[6] T. Schlegl, P. Seeböck, S. M. Waldstein, G. Langs, and U. Schmidt-Erfurth, "F-anogan: Fast unsupervised anomaly detection with generative adversarial networks," *Medical Image Analysis*, vol. 54, pp. 30–44, 2019, ISSN: 1361-8415. DOI: https://doi.org/10.1016/j.media.2019.01.010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361841518302640.

[7] Y. L. Kai Jiang Weiying Xie, "Semisupervised spectral learning with generative adversarial network for hyperspectral anomaly detection," pp. 5224–5236, Feb. 2020. DOI: 10.1109/TGRS.2020.2975295. [Online]. Available: https://doi.org/10.1109/TGRS.2020.2975295.

[8] Z. Niu, K. Yu, and X. Wu, "Lstm-based vae-gan for time-series anomaly detection," *Sensors*, vol. 20, no. 13, 2020, ISSN: 1424-8220. DOI: 10.3390/s20133738. [Online]. Available: https://www.mdpi.com/1424-8220/20/13/3738.

[9] W. Shin, S.-J. Bu, and S.-B. Cho, "3d-convolutional neural network with generative adversarial network and autoencoder for robust anomaly detection in video surveillance," *International Journal of Neural Systems*, vol. 30, no. 06, p. 2 050 034, 2020, PMID: 32466693. DOI: 10.1142/S0129065720500343. eprint: https://doi.org/10.1142/S0129065720500343. [Online]. Available: https://doi.org/10.1142/S0129065720500343.

[10] Y. Skandarani, N. Painchaud, P.-M. Jodoin, and A. Lalande, "On the effectiveness of gan generated cardiac mris for segmentation," 2020. DOI: 10.48550/arXiv.2005.09026. arXiv: 2005.09026 [eess.IV]. [Online]. Available: https://doi.org/10.48550/arXiv.2005.09026.

[11] L. Sun, J. Wang, and Huang, "An adversarial learning approach to medical image synthesis for lesion detection," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 9, pp. 2303–2314, 2020, ISSN: 2168-2208. DOI: 10.1109/JBHI.2020.2964016. [Online]. Available: https://doi.org/10.1109/JBHI.2020.2964016.

[12] K. T. Xue Wang, "Cva2e: A conditional variational autoencoder with an adversarial training process for hyperspectral imagery classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 8, pp. 5676–5692, Aug. 2020. DOI: 10.1109/TGRS.2020.2968304. [Online]. Available: https://doi.org/10.1109/TGRS.2020.2968304.

[13] F. Carrara, G. Amato, L. Brombin, F. Falchi, and C. Gennaro, "Combining gans and autoencoders for efficient anomaly detection," pp. 3939–3946, 2021. DOI: 10.1109/ICPR48806.2021.9412253. [Online]. Available: https://doi.org/10.1109/ICPR48806.2021.9412253.

[14] P. Huang, X. Liu, and Y. Huang, "Data augmentation for medical mr image using generative adversarial networks," 2021. DOI: 10.48550/arXiv.2111.14297. arXiv: 2111.14297 [cs.CV]. [Online]. Available: https://doi.org/10.48550/arXiv.2111.14297.

[15] F. Li, W. Huang, M. Luo, P. Zhang, and Y. Zha, "A new vae-gan model to synthesize arterial spin labeling images from structural mri," *Displays*, vol. 70, p. 102 079, 2021, ISSN: 0141-9382. DOI: https://doi.org/10.1016/j.displa.2021.102079. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0141938221000858.

[16] F. D. Mattia, P. Galeone, M. D. Simoni, and E. Ghelfi, *A survey on gans for anomaly detection*, 2021. arXiv: 1906.11632 [cs.LG]. [Online]. Available: https://arxiv.org/abs/1906.11632.

[17] H. Y. X. L. S.-H. W. Z. L. J. Y. Y. J. P. Qian, "Synthesizing multi-contrast mr images via novel 3d conditional variational auto-encoding gan," *Mobile Netw Appl*, no. 26, pp. 415–424, 2021. DOI: 10.1007/s11036-020-01678-1. [Online]. Available: https://doi.org/10.1007/s11036-020-01678-1.

[18] V. N. Rohit Sahoo, "Gans and vaes as methods of synthetic data generation and augmentation to enhance heart disease prediction," Dec. 2021, ISSN: 2249-8958. DOI: 10.35940/ijeat.B3263.1211221. [Online]. Available: https://doi.org/10.35940/ijeat.B3263.1211221.

[19] M. Abedi, L. Hempel, S. Sadeghi, and T. Kirsten, "Gan-based approaches for generating structured data in the medical domain," *Applied Sciences*, vol. 12, no. 14, 2022, ISSN: 2076-3417. DOI: 10.3390/app12147075. [Online]. Available: https://www.mdpi.com/2076-3417/12/14/7075.

[20] B. Ahmad, J. Sun, Q. You, V. Palade, and Z. Mao, "Brain tumor classification using a combination of variational autoencoders and generative adversarial networks," *Biomedicines*, vol. 10, no. 2, 2022, ISSN: 2227-9059. DOI: 10.3390/biomedicines10020223. [Online]. Available: https://www.mdpi.com/2227-9059/10/2/223.

[21] A. Zayat, M. A. Hasabelnaby, M. Obeed, and A. Chaaban, "Transformer masked autoencoders for next-generation wireless communications: Architecture and opportunities," *IEEE Communications Magazine*, pp. 1–8, 2023. DOI: 10.1109/MCOM.002.2300257.

[22] L. Zhang, Y. Dai, F. Fan, and C. He, "Anomaly detection of gan industrial image based on attention feature fusion," *Sensors*, vol. 23, no. 1, 2023, ISSN: 1424-8220. DOI: 10.3390/s23010355. [Online]. Available: https://doi.org/10.3390/s23010355.

[23] S. A. Najafi, M. H. Asemani, and P. Setoodeh, "Attention and autoencoder hybrid model for unsupervised online anomaly detection," 2024. DOI: 10.48550/arXiv.2401.03322. arXiv: 2401.03322 [cs.LG]. [Online]. Available: https://doi.org/10.48550/arXiv.2401.03322.