SRH University Heidelberg

Master's Thesis

An Approach to Medical Image Processing Based on Generative Adversarial Networks

Author	: Abhishek Chandru
Matriculation Number	: 11038053
Course	: M.Eng.-Information Technology
Department	: School of Engineering and Architecture
Address	: Am Steingarten 12, 68169 Mannheim
Email Address	: abhishek.c1528@gmail.com
1st Supervisor	: Prof. Dr. Achim Gottscheber
2nd Supervisor	: Prof. Dr. Milan Gnjatović
Begin	: 14 November 2024
End	: 7 April 2025

Acknowledgement

The process of developing a master's thesis titled " An Approach to Medical Image Processing based on Generative Adversarial Networks" was made possible and aided with the help of numerous individuals whose influence cannot be overstated. In the first place I extend my sincere gratitude to my advisor, whose competence, counsel and tireless support enabled me to navigate my way through researching and writing. Their informed criticism and encouragement were the key drivers in my paper revision. I would like to express my gratitude to the department's faculty and staff for enabling me to reach this point in my academic and research career.

I express my gratitude to my friends and colleagues for fostering a vibrant and collaborative research environment. Their companionship and intellectual interchange made the arduous task more pleasurable and fruitful for me. I am most appreciative to those who were part of the experiments, their data were theones that shaped this study.

In the course of writing this thesis, I have not only experienced my own personal growth but also witnessed the coming together of many to form a collective achievement. Lastly but not least, I thank you all for sharing this journey with me.

Declaration of Authorship

I hereby declare,

that my thesis is my own original work. I have written it independently without outside help and have not used any sources other than those indicated - in particular, no sources not named in the references.

that I have appropriately indicated any direct quotations or passages taken from literature, as well as the use of intellectual property from other authors, by providing the necessary citations within the work. This applies equally to the sources used for text generation by Artificial Intelligence (AI). In the case of utilizing AI-generated quotes, I have independently verified and ensured the source's reliability, which includes both the verbatim question/task and its corresponding answer (by means of screenshot).

that the **thesis** was not previously presented to another examination board.

I am aware that in case of suspicion, this thesis may be checked for plagiarism and for appropriate use and reference of others' work.

I am aware of legal consequences in cases of incorrect declaration.

Place, Date Signature

Abstract

Medical imaging has significantly advanced healthcare diagnostics, yet the detection of subtle anomalies, such as early-stage brain tumors in MRI scans, remains challenging due to the complexities of image analysis and the scarcity of annotated data. This thesis explores the application of Generative Adversarial Networks (GANs), specifically Vector Quantized GAN (VQGAN+D), to enhance anomaly detection and image reconstruction accuracy in brain MRI images. By integrating vector quantization into the GAN architecture, the study addresses issues of capturing detailed anatomical structures, reducing reliance on extensively annotated datasets, and improving model interpretability. The proposed VQGAN+D framework demonstrates superior performance in identifying subtle yet clinically significant anomalies, thus contributing substantially towards bridging the gap between automated diagnostic systems and clinical practice.

Kurzfassung

Die medizinische Bildgebung hat die Diagnostik im Gesundheitswesen erheblich vorangebracht, dennoch stellt die Erkennung subtiler Anomalien, wie etwa frühzeitiger Hirntumore in MRT-Aufnahmen, aufgrund der Komplexität der Bildanalyse und des Mangels an annotierten Daten weiterhin eine Herausforderung dar. Diese Masterarbeit untersucht den Einsatz Generativer Adversarialer Netzwerke (GANs), insbesondere des Vector Quantized GAN (VQGAN+D), um die Genauigkeit bei der Anomaliedetektion und Bildrekonstruktion in Hirn-MRT-Bildern zu verbessern. Durch die Integration der Vektorquantisierung in die GAN-Architektur adressiert die Studie die Herausforderungen bezüglich der Erfassung detaillierter anatomischer Strukturen, reduziert den Bedarf an umfangreich annotierten Datensätzen und verbessert die Interpretierbarkeit des Modells. Das entwickelte VQGAN+D-Framework zeigt überlegene Leistung bei der Identifikation subtiler, jedoch klinisch relevanter Anomalien und trägt somit wesentlich dazu bei, die Lücke zwischen automatisierten Diagnosesystemen und der klinischen Praxis zu schließen.

Table of contents

Acknowledgement II

Abstract IV

Kurzfassung V

List of Tables X

List of Figures XI

List of Abbreviations XII

1 Introduction 1

1.1 Motivation 1

1.2 Background 1

1.3 Problem Statement 3

1.4 The Challenge of Limited Annotated Data 3

1.4.1 Complexities and Subtleties of Brain MRI Scans 3

1.4.2 Issues Concerning Stability and Interpretability in Generative Adversarial Networks 4

1.4.3 Limited Access to High-Quality Training Data 4

1.5 Research Questions 5

1.6 Thesis Structure 5

2 Literature Review 7

2.1 GANs in Medical Image Analysis 7

2.2 Brain MRI Imaging: Challenges, Advances, and Future Directions 9

2.3 Key Challenges in Brain MRI Analysis 9

2.3.1 Subtle Variations in Tissue Contrast 9

2.3.2 Technical Artifacts and Imaging Flaws 9

2.3.3 Recent Developments in Brain MRI Examination 10

2.4 Advancements with VQGAN in Brain MRI 12

2.5 GAN-based Synthetic Data Generation in MRI 14

3 Theoretical Background **17**

3.1 Fundamentals of Generative Adversarial Networks (GANs) **17**

3.1.1 GAN Components and Training 17

3.2 Variants of GANs: **18**

3.2.1 Deep Convolutional GANs (DCGAN) 18

3.2.2 Wasserstein GAN (WGAN) 19

3.2.3 Conditional GAN (CGAN) 19

3.2.4 Progressive Growing GAN (PGGAN) 19

3.2.5 Practical Implications in Medical Imaging 19

3.3 Vector Quantization and VQGAN+D **20**

3.3.1 Understanding VQGAN: Architecture and Functionality 20

3.3.2 Enhancing Anomaly Detection: The Role of the Discriminator in VQGAN+D 20

3.3.3 Advantages of Discrete Latent Representations in Medical Image Reconstruction 21

3.3.4 Applications and Implications in Clinical Practice 21

4 Proposed methodology **22**

4.1 Dataset **22**

4.2 Implementation Details and Tools **24**

4.2.1 PyTorch 24

4.2.2 Torchvision 24

4.2.3 NumPy 25

4.2.4 OpenCV (cv2) 25

4.2.5 Matplotlib and Seaborn 26

4.2.6 TensorBoard 26

4.3 Preprocessing Techniques **27**

4.3.1 Resizing for Standardized Input Dimensions 28

4.3.2 Conversion to Tensor Format 28

4.3.3 Dataset Structure and Loading 28

4.4 VQGAN+D Model Architecture **29**

4.4.1 MRI Image Acquisition and Preprocessing 29
4.4.2 Feature Encoding Using VQGAN+D 30
4.4.3 Vector Quantization: Enhancing Feature Representation 30
4.4.4 The Role of the Codebook in Learning Discrete Features 31
4.4.5 Image Reconstruction via VQGAN Decoder 31
4.4.6 Role of the Discriminator in Evaluating Reconstruction Quality 32
4.4.7 Anomaly Detection via Reconstruction Error Computation 32
4.4.8 Final Decision: Classifying MRI Scans as Normal or Anomalous 33

4.5 Training Procedure 33
4.5.1 Generator (VQGAN) Loss Function 33
4.5.2 Discriminator Loss Function 34
4.5.3 Optimization Strategy and Hyperparameters 34
4.5.4 Adam Optimizer: Overview and Functionality 35
4.5.5 Training Process and Checkpoints 37

4.6 Testing and Evaluation 39
4.6.1 Anomaly Detection Process 39
4.6.2 Adaptive Thresholding for Anomaly Classification 40
4.6.3 Evaluation Metrics 40
4.6.4 Summary of Testing and Evaluation 42

5 Result 42

5.1 Comparative Analysis of Model Training First 10 vs Last 10 Epochs . 43

5.2 Images Generated across Epochs 45

5.3 Detailed Analysis of Confusion Matrix 47

5.4 Comparison of Generated Images (DCGAN vs. VQGAN) 51
5.4.1 Similarity of Generated images: 51
5.4.2 Texture and Noise Patterns: 53

5.5 Comparative Analysis of Confusion Matrices: VQGAN+D vs DCGAN 54

6 Conclusion and Future Research 56

Bibliography **57**

List of Tables

Table 1. Confusion Matrix Structure 41
Table 2. Comparison for first 10 and last 10 epochs 44

List of Figures

Figure 1. Healthy Brain Scans 23
Figure 2. Anomaly Brain Scans 23
Figure 3. Preprocessing of input images 27
Figure 4. VQGAN+D Model Architecture 29
Figure 5. Training process of VQGAN+D 37
Figure 6. The losses for first 10 epochs during the training process of the model .. 42
Figure 7. Images generated at epoch 0 45
Figure 8. Images generated at epoch 200 45
Figure 9. Images generated at epoch 400 45
Figure 10. images generated at epoch 600 46
Figure 11. images generated at epoch 800 46
Figure 12. images generated at epoch 1000 46
Figure 13. confusion matrix 47
Figure 14. ROC curve of the anomaly detection 49
Figure 15 Precision- Recall curve 50
Figure 16. Healthy images Original and Reconstructed using VQGAN 52
Figure 17. Anomaly images Original and Reconstructed using VQGAN+D 52
Figure 18. Anomaly images Original and Reconstructed using DCGAN 53
Figure 19. Anomaly images Original and Reconstructed using DCGAN 53
Figure 20. Confusion matrix of VQGAN+D 54
Figure 21. Confusion matrix of DCGAN 55

List of Abbreviations

GAN	Generative Adversarial Networks
VQGAN	Vector Quantized GAN
CNN	Convolutional Neural Networks
VQVAE	Vector Quantized Variational Autoencoder
CSF	Cerebrospinal Fluid
MHE	Multiplex Histogram Equalization
MBOBHE	Multi-Purpose Beta Optimised Bi-Histogram Equalization
ESGAN	Enhancement and Segmentation GAN
WGANs	Wasserstein GANs
PGGAN	Progressive Growing GAN
EMA	Exponential Moving Average
BCE	Binary Cross-Entropy
TPR	True Positive Rate
FPR	False Positive Rate
AUC	Area Under the Curve
DCGAN	Deep Convolutional Generative Adversarial Networks
MRI	Magnetic Resonance Imaging
GPU	Graphics processing unit
CUDA	Compute Unified Device Architecture

1 Introduction

1.1 Motivation

Medical imaging has revolutionized how doctors diagnose and treat diseases, especially in critical areas like brain tumor detection. Most current systems, however, depend on significant volumes of labeled data, which in the medical profession can be challenging and costly to gather. Conventional machine learning algorithms find it difficult to precisely identify minor or rare irregularities in the absence of annotated images.

In medical imaging, deep learning—especially Generative Adversarial Networks (GANs)—has created hitherto unrealized opportunities. First, to apply GANs for unsupervised anomaly detection Schlegl et al. (2017) and Akcay et al. (2018) let models learn the patterns of "normal" images and highlight anything that deviates. This method greatly lessens reliance on labeled data. However, conventional GAN models can miss anomalies or false detections since they cannot adequately capture the minute features of medical images. Here Vector Quantized GAN (VQGAN) finds use. VQGAN+D can retain minor structural elements by including discrete latent representations, hence enhancing the clarity and accuracy of anomaly identification.

Using VQGAN+D on brain MRI data, our work expands on previous developments in an effort to increase early-stage tumor identification. Our methodology facilitates the bridging of the gap between automated diagnosis and real-world clinical accuracy by capturing finer data, therefore enabling radiologists to make faster and more accurate judgments.

1.2 Background

Medical imaging has emerged as a fundamental element of contemporary healthcare, providing clinicians with exceptional insight into the human body via modalities like X-rays, computed tomography (CT), and magnetic resonance imaging (MRI). The images presented are crucial for the accurate diagnosis of diseases, the formulation of treatment plans, and the ongoing assessment of patient progress. Nonetheless, the examination of these images presents considerable complexity.

Historically, the discipline depended on meticulously crafted features—deliberately designed attributes that encapsulate particular elements of an image. Specialists would carefully develop these attributes and subsequently employ supervised machine learning models to categorize and examine images. While this method has played a crucial role in the initial applications of medical imaging, it is accompanied by notable limitations. Manual feature engineering is characterized by its labor-intensive and costly nature, and it is significantly reliant on the presence of expertly labeled data. In numerous clinical settings, the collection of such annotations presents significant challenges, rendering these conventional approaches less scalable and at times impractical.

The emergence of deep learning has initiated a transformative period in the field of image analysis. Convolutional Neural Networks (CNNs) have significantly transformed the domain by enabling the automatic extraction of hierarchical features from raw image data. This automation minimizes the necessity for manual involvement and has resulted in significant performance improvements across various visual recognition tasks. Nonetheless, convolutional neural networks generally necessitate extensive, labeled datasets to function optimally, presenting a significant challenge within the medical field. Furthermore, although convolutional neural networks demonstrate strong performance in situations where abnormalities are clearly visible, they frequently encounter difficulties in identifying the subtle anomalies that are essential for early diagnosis in medical imaging. This challenge has generated significant interest in exploring alternative methodologies that can derive insights from data while minimizing dependence on manual labeling. GANs represent a significant advancement in the field. A GAN comprises two opposing networks: one is a generator responsible for producing images, while the other is a discriminator tasked with assessing those images. By engaging in a repetitive adversarial process, the generator acquires the ability to create images that closely resemble authentic ones. This capability renders GANs especially attractive for unsupervised anomaly detection, wherein the objective is to comprehend the characteristics of "normal" and subsequently recognize deviations from that standard. Nonetheless, conventional GANs occasionally struggle to accurately represent the intricate structural details that are crucial for precise medical diagnosis, as they may neglect subtle yet important image characteristics.

In response to these limitations, recent studies have developed more sophisticated models that enhance the representation of images. A significant advancement was introduced by van den Oord et al. through VQ-VAE . The VQ-VAE which discretizes the latent space into finite tokens, effectively capturing intricate and semantically relevant features that continuous representations might overlook [1]. Expanding upon this notion, Esser et al. further developed the concept with VQGAN, incorporating vector quantization into the GAN framework to achieve highly detailed reconstructions. In the field of medical imaging, where even the slightest alterations can hold clinical importance, such comprehensive representations are of immense value.[2]

Furthermore, extensive evaluations such as the one conducted by Kazeminia et al. offer a thorough examination of the potential benefits and obstacles associated with the application of GANs in the field of medical image analysis. The findings emphasize the significant potential of integrating unsupervised learning methods with sophisticated deep learning frameworks, while also pointing out persistent challenges, including issues related to training stability, interpretability, and the necessity for enhanced anomaly detection approaches [3].

1.3 Problem Statement

While Generative Adversarial Networks (GANs) have introduced innovative opportunities for medical image processing, several challenges remain that may hinder their broad implementation in clinical settings. The recognized limitations impede the ability to accurately detect subtle irregularities, generate high-quality synthetic images, and ensure reliable diagnostic outcomes. This study aims to address the key challenges outlined below

1.4 The Challenge of Limited Annotated Data

Acquiring expert annotations for medical imaging datasets often involves substantial expenses, considerable time investment, and challenges in practicality, particularly in large-scale supervised learning contexts. While unsupervised learning methods reduce dependence on labeled data, they face difficulties in effectively capturing complex image details, potentially leading to missed or misclassified anomalies.

1.4.1 Complexities and Subtleties of Brain MRI Scans

This study centres on brain MRI scans, which pose unique challenges for reconstruction and anomaly detection due to their intricate anatomical structures, varied tissue contrasts, and subtle pathological characteristics. Subtle but clinically significant changes—like early-stage tumors, lesions, or minor structural anomalies—can complicate the process of distinguishing them from normal variations. Many conventional models that utilize GANs encounter difficulties in accurately reconstructing complex details, leading to missed anomalies or erroneous detections, which ultimately threatens the accuracy of diagnostics.

1.4.2 Issues Concerning Stability and Interpretability in Generative Adversarial Networks

A notable issue linked to traditional GAN architectures is their instability during the training phase, often resulting in mode collapse, where the model fails to generate a diverse range of significant outputs. The observed instability can result in outcomes that are both unpredictable and inconsistent, making it difficult to depend on GAN-based methods within a clinical setting. Moreover, the unclear nature of these models limits their interpretability, leading to difficulties for medical professionals, including doctors and radiologists, who struggle to understand or validate the decision-making processes at play. The lack of clear and understandable outcomes remains a considerable obstacle to the incorporation into clinical practice.

1.4.3 Limited Access to High-Quality Training Data

Medical imaging datasets, particularly regarding anomaly detection, face constraints stemming from privacy concerns, variations in imaging protocols, and the high costs linked to data collection. The ability to generate high-quality synthetic data could potentially mitigate this limitation by augmenting real datasets with clinically realistic images. Nonetheless, existing approaches that employ GANs often face challenges in preserving complex details, ensuring a diverse range of images, and producing data that aligns with the practical demands of real-world clinical applications.

Proposed Methodology:

A Framework for Anomaly Detection Using VQGAN. This document outlines a framework for detecting anomalies utilizing a Vector Quantized Generative Adversarial Network (VQGAN), specifically designed for brain MRI scans to tackle the recognized challenges. This approach goes further than simple anomaly detection; it aims to generate high-quality synthetic medical images, offering crucial resources for the training, validation, and development of algorithms. The main objectives of this framework include:

Detailed Representation:

Utilizing vector quantization to create discrete latent representations, allowing the model to detect and highlight subtle anomalies that might otherwise go unrecognized.

Robust Reconstruction:

This involves enhancing the accuracy and reliability of anomaly detection by employing a hybrid loss function that integrates reconstruction loss, adversarial loss, and perceptual loss. This approach ensures that the model maintains structural integrity and realism in its outcomes.

Improved Stability and Clarity:

Employing VQGAN's discrete latent space to stabilize the training process, thus reducing mode collapse and inconsistencies that often impact traditional GANs. Providing distinct outcomes that improve clinical decision-making, thus assisting radiologists in understanding and validating the generated results.

This study addresses challenges associated with limited data, model reliability, and clarity of interpretation, contributing to the advancement of GAN-based anomaly detection and achieving significant strides toward real-world applications in the healthcare sector. The combination of anomaly detection and synthetic data generation offers considerable potential for improving early disease identification, AI-assisted diagnosis, and advancements in medical imaging research.

1.5 Research Questions

The work presented in this thesis revolves around several key questions:

RQ1: How does integrating vector quantization into a GAN framework (i.e., using VQGAN) enhance anomaly detection in medical images compared to traditional GAN methods?

RQ2: What benefits do discrete latent representations provide when trying to capture the detailed features necessary for detecting anomalies?

RQ3: In what ways does the combination of reconstruction, adversarial, and perceptual losses improve the performance and robustness of the model?

RQ4: What are the practical challenges of applying this method across various imaging modalities and clinical scenarios?

RQ5: How can the interpretability of the model's outputs be improved to support clinical decision-making more effectively?

These questions guide our investigation into the strengths, weaknesses, and real-world applicability of the proposed method.

1.6 Thesis Structure

The thesis is organized into distinct chapters, systematically covering the research aspects of employing Generative Adversarial Networks (GANs), particularly Vector Quantized GANs (VQGAN), in medical image processing for anomaly detection in brain MRI scans.

Section 1: Introduction This Section provides a foundational understanding of the research topic, detailing the motivation, background of medical imaging and GAN technology, problem statement, specific challenges, research questions, and an overview of the thesis structure.

Section 2: Literature Review critically reviews existing research on GAN applications in medical imaging, particularly in anomaly detection for brain MRI scans. It discusses current methodologies, identifies gaps, and outlines how the thesis aims to address these gaps through advancements in GAN technology, specifically VQGAN.

Section 3: Theoretical Background This section explains the fundamental principles behind GANs, including detailed insights into various GAN architectures such as DCGAN, WGAN, CGAN, and PGGAN. It particularly emphasizes vector quantization and its integration within GAN architectures, establishing a comprehensive theoretical understanding essential for subsequent chapters.

Section 4: Proposed Methodology Detailed descriptions of the dataset utilized, tools and software frameworks (PyTorch, Torchvision, NumPy, OpenCV, Matplotlib, Seaborn, and TensorBoard), preprocessing techniques, and the specific architecture of the proposed VQGAN+D model are provided. This chapter thoroughly explains feature encoding, vector quantization processes, the role of the discriminator, and the methodology behind anomaly detection via reconstruction error.

Section 5: Experimental Setup and Results this section outlines the experimental procedures, including model training with specific optimization strategies, hyperparameter settings, and performance evaluation methods. Results from various epochs, comparative analyses of generated images, and evaluation metrics such as confusion matrices and ROC curves are presented and analyzed.

Section 6: Discussion section six provides an in-depth interpretation of the experimental outcomes, critically discussing the strengths, limitations, and practical applicability of the VQGAN+D framework. It addresses the initial research questions, evaluates the clinical implications of the findings, and discusses challenges encountered during the research.

Chapter 7: Conclusion and Future Research the final section summarizes the key contributions of the thesis, highlights implications for medical diagnostics, and suggests future research directions. Potential improvements and broader applications of the VQGAN+D methodology in clinical contexts are also discussed.

References and Appendices

This section comprises an extensive bibliography of all referenced materials used throughout the thesis.

2 Literature Review

2.1 GANs in Medical Image Analysis

The application of GAN has significantly enhanced the field of medical image analysis. Goodfellow et al. presented GANs, which employ an innovative training approach that integrates two neural networks: a discriminator tasked with distinguishing between generated images and real ones, and a generator that produces synthetic images. This adversarial learning drives the generator to produce outputs that are exceptionally realistic and closely resemble genuine medical images. Initially, GANs underwent thorough examination for the creation of general visual content, demonstrating their potential across a range of image-generation tasks due to their strong ability to capture complex data distributions [4].

A significant early application of GANs within the domain of medical imaging was AnoGAN, introduced by Schlegl et al. AnoGAN utilized deep convolutional generative adversarial networks specifically designed for unsupervised anomaly detection in medical imaging, demonstrating its effectiveness in identifying pathological features within retinal images. The key innovation of AnoGAN is its unsupervised approach to comprehending normal anatomical variability based on healthy image data. The methodology involves directing a generator to create standard anatomical images, while the discriminator is responsible for distinguishing these synthetic images from genuine ones. This dual learning process enables the detection of anomalies by assessing deviations from the defined "normal" manifold. While it introduces an innovative approach, AnoGAN faces several limitations, including unstable training dynamics, challenges in accurately reconstructing complex anatomical details, and increased sensitivity to hyperparameter tuning [13].

In light of the constraints identified in AnoGAN, GANomaly Akcay et al. proposed an advanced adversarial framework that integrates adversarial and autoencoder losses. GANomaly addresses the complexities involved in identifying anomalies within medical images by employing an adversarial autoencoder mechanism, resulting in improved stability and robustness in the detection process. This model effectively detects subtle anomalies by obtaining latent representations of images in conjunction with their corresponding reconstruction errors. Although GANomaly demonstrated enhanced performance compared to AnoGAN, particularly in identifying subtle anomalies, it encountered difficulties, especially in the precise reconstruction of complex medical images like brain MRIs [14].

Hamghalam and Simpson presented a novel method that leverages GANs in the field of medical imaging, specifically using conditional GANs (cGANs) to achieve the segmentation of brain tumors. The approach involved generating high-contrast synthetic MRI images, resulting in significant improvements in both voxel-wise and region-wise segmentation. The suggested framework, ESGAN, combines classifier loss with adversarial loss, which effectively improves the contrast between different brain tissues and subsequently enhances segmentation accuracy. This study highlights the potential of GANs in tackling the difficulties associated with overlapping intensities in MR images, a common challenge within the realm of medical image analysis [9]. Recent developments in GAN architectures, particularly through the integration of vector quantization, were presented by van den Oord et al. with the Vector Quantized Variational Autoencoder (VQ-VAE) and further advanced by Esser et al. (2021) in VQGAN. The integration of discrete latent representations within VQGAN markedly enhances the model's ability to preserve complex visual details. Esser et al. introduced an innovative method that incorporates transformers within a VQGAN framework, facilitating the proficient modeling of global image context and high-resolution details, which are crucial for medical applications such as brain MRI analysis. The use of a discrete representation in conjunction with an adversarial training approach significantly improved the model's capacity to produce high-resolution, intricately detailed images, thereby enabling the precise detection of subtle anomalies that traditional GANs may overlook [16][1].
Alalwan et al. enhanced the utilization of GANs by integrating synthetic data generation with federated learning to address the data privacy challenges frequently faced in medical research. The methodology involved the training of federated convolutional neural networks (CNNs) using synthetic data generated by GANs, which notably improved diagnostic accuracy while preserving the confidentiality of patient information. This study emphasizes the significance of data produced by GANs in clinical environments, especially in delicate fields like medical imaging, where constraints stemming from data scarcity and privacy concerns present considerable obstacles to the progress of research [30].
In summary, methods employing GANs, including AnoGAN, GANomaly, ESGAN, and VQGAN, have collectively represented important advancements in the domain of medical imaging. The developments in these approaches have methodically tackled the limitations of traditional medical imaging methods, particularly in the areas of anomaly identification and synthetic data creation. However, issues like training instability, interpretability, and the creation of

realistic synthetic data continue to exist, requiring sustained research and development initiatives. This research addresses these shortcomings by presenting a detailed VQGAN-based framework for detecting anomalies, specifically tailored to manage the complexities of brain MRI scans, thus improving accuracy in anomaly detection, realism in synthetic data, and interpretability [12][30][10].

2.2 Brain MRI Imaging: Challenges, Advances, and Future Directions

Brain MRI imaging represents a highly intricate and challenging domain within medical imaging, largely owing to the complex structures and nuanced variations present in the brain. In contrast to other imaging modalities, MRI scans provide exceptionally detailed anatomical characteristics, necessitating careful analysis to distinguish among different tissue types, including gray matter, white matter, cerebrospinal fluid (CSF), and pathological conditions such as tumors, lesions, or edema. The tissues display nuanced variations in intensity and texture, which complicates the process of automated analysis significantly.

2.3 Key Challenges in Brain MRI Analysis

2.3.1 Subtle Variations in Tissue Contrast

A significant challenge in the analysis of brain MRI lies in the overlapping intensity distributions among various tissues. In contrast to the more clearly defined structures observed in various medical imaging modalities, brain tissues often exhibit comparable intensity levels. This similarity poses challenges for automated systems tasked with differentiating between healthy and abnormal areas. This matter is especially significant in activities such as anomaly detection and segmentation, where accurate recognition of irregularities is crucial [9]. For example, identifying minor lesions or tumors in their initial stages necessitates a model capable of discerning slight intensity variations while remaining unaffected by typical anatomical differences.

2.3.2 Technical Artifacts and Imaging Flaws

MRI scans are affected by variations in intensity and artifacts resulting from elements like uneven magnetic fields, patient motion, and inconsistencies in hardware. Such imperfections may occasionally imitate or conceal genuine anomalies, resulting in erroneous positives or overlooked diagnoses. Conventional computational techniques, along with certain contemporary machine learning methods, face challenges in addressing these inconsistencies. This necessitates the development of more advanced algorithms that can effectively differentiate between authentic anomalies and imaging artifacts.

2.3.3 Recent Developments in Brain MRI Examination

In response to these challenges, investigators have examined various sophisticated methods that enhance the interpretation of MRI images and the detection of anomalies.

- **Conditional GANs for Improved Segmentation**
  A conditional GAN framework aimed at improving image contrast to achieve more accurate tumor segmentation. ESGAN employs a combination of classifier and adversarial losses to enhance image quality and more effectively delineate tumor boundaries. Although the model has shown impressive outcomes, it continues to exhibit considerable sensitivity to hyperparameter adjustments and necessitates substantial computational resources, potentially obstructing practical application [9].
- **Integration of Features and Hybrid U-Net Structures**
  Another significant development is presented by Nizamani et al., who created hybrid U-Net architectures that integrate sophisticated preprocessing techniques such as:
  CLAHE is a technique that enhances the contrast of images by applying adaptive histogram equalization while limiting the amplification of noise. Multipeak Histogram Equalization (MHE) and Multi-Purpose Beta Optimized Bi-Histogram Equalization (MBOBHE)
  The application of these techniques serves to normalize intensity variations and enhance the clarity of subtle tissue boundaries, thereby significantly improving segmentation accuracy. Nonetheless, the computational requirements of these models, along with the necessity for precise tuning of preprocessing parameters, remain obstacles to widespread implementation [31].
- **Unsupervised Anomaly Detection Utilizing Generative Adversarial Networks**
  The introduction of AnoGAN by Schlegl et al. in 2017 marked one of the initial uses of GANs within the realm of medical imaging. AnoGAN introduced a novel approach to anomaly detection by utilizing unsupervised learning techniques, focusing exclusively on healthy data. This methodology enables the model to identify abnormalities by recognizing deviations from the established "normal" anatomical patterns. Although this represents a notable advancement, AnoGAN is accompanied by various limitations, such

as issues with training stability and challenges in precisely reconstructing intricate anatomical structures, especially in brain MRI scans [13].

- **Vector Quantization for Enhanced Image Representation**
  In response to the constraints of conventional GANs, Esser et al. (2021) presented VQGAN, a model that combines vector quantization with transformer networks. The discrete latent representation offered by VQGAN significantly improves the model's capacity to capture complex image structures. The incorporation of transformer components enhances the modeling of long-range relationships, resulting in superior image quality and more effective anomaly detection. Nonetheless, similar to numerous deep learning frameworks, VQGAN encounters difficulties regarding training stability and scalability, especially when utilized for high-resolution brain MRI scans [16].
- **The Role of Synthetic Data in Medical Imaging**
  In order to improve model robustness and supplement training datasets, researchers have looked at creating synthetic data due to the lack of high-quality medical imaging datasets. Alalwan et al. (2024) effectively utilized GAN-generated synthetic MRI images within federated learning frameworks, enabling various institutions to collaboratively train models while maintaining the confidentiality of patient data. This methodology effectively tackles issues related to limited data availability and privacy considerations, leading to notable enhancements in accuracy. Nonetheless, guaranteeing the clinical authenticity and dependability of synthetic images continues to be a persistent challenge, necessitating thorough validation by healthcare experts prior to the broad application of such data in clinical settings [30].
- **Future Directions and Research Priorities**
  The analysis of brain MRI is a highly complex area within medical imaging, necessitating ongoing improvements to boost model stability, interpretability, and diagnostic precision. Subsequent investigations ought to concentrate on Enhancing the stability of training in GAN models to mitigate challenges such as mode collapse and to guarantee consistent and dependable outputs. Improving the clarity of anomaly detection models, allowing clinicians to comprehend and have confidence in AI-generated outcomes. Thoroughly validating synthetic medical images to connect AI-generated data with clinically applicable imaging resources.

2.4 Advancements with VQGAN in Brain MRI

In recent years, significant advancements in Generative Adversarial Networks (GANs) have considerably improved medical image processing capabilities, notably through architectures like Vector Quantized Generative Adversarial Networks (VQGAN). The introduction of VQGAN marked a substantial leap forward, especially in addressing the challenges inherent in brain MRI analysis. Brain MRI images are characterized by complex anatomical structures, subtle pathological features, and varying intensity distributions. Thus, traditional GAN-based methods faced limitations in accurately detecting subtle yet clinically significant anomalies. The introduction of VQGAN, with its discrete latent representations and powerful transformer-based mechanisms, offers a promising pathway to address these complex challenges effectively.

VQGAN, introduced by Esser et al. represents a transformative improvement in the GAN paradigm by combining vector quantization (VQ) with transformers. Unlike conventional GANs, which rely on continuous latent spaces that might fail to capture detailed structural features crucial for medical image analysis, VQGAN uses vector quantization to discretize the latent space into distinct visual tokens. These tokens represent essential visual elements that facilitate capturing complex, detailed structures within images. The primary strength of this approach lies in its ability to learn and represent rich, contextually meaningful image features at multiple scales. This enhanced representation capability is particularly valuable for medical imaging tasks, including anomaly detection in brain MRIs, where preserving detailed anatomical structures is crucial [16].

The architecture of VQGAN combines convolutional neural networks (CNNs) with transformers to balance local and global contextual information. CNNs provide robust local feature extraction, while transformers excel at modeling long-range dependencies within images. This hybrid model ensures that the fine anatomical details are accurately captured, enabling precise detection of subtle anomalies that traditional GAN architectures might overlook. Additionally, VQGAN's patch-based discriminator significantly improves the visual quality of generated images, making them more realistic and clinically relevant.

Recent studies have validated the effectiveness of VQGAN for medical imaging tasks, particularly in anomaly detection and synthetic image generation. For example, conditional GAN-based methods similar to VQGAN frameworks for brain tumor segmentation. Their work emphasized generating synthetic MRI images with improved contrast for better tumor sub-region identification. They successfully demonstrated that conditional GAN approaches significantly enhance the accuracy of segmentation methods by providing synthetic data with reduced intensity overlaps. This synthetic data generation capability is particularly beneficial for addressing data scarcity, a common challenge in medical imaging [9].

Similarly, UNet models integrated with CNN and advanced preprocessing techniques like CLAHE, MHE, and MBOBHE. Although not directly based on VQGAN, their emphasis on feature extraction and image enhancement parallels the philosophy of VQGAN, emphasizing the importance of detailed feature representations. Such preprocessing methods can also be effectively combined with VQGAN to further boost the performance of anomaly detection and segmentation tasks, enhancing model reliability and accuracy [32].

One pivotal contribution in the development of VQGAN-based methods is their potential to significantly improve anomaly detection. AnoGAN, was groundbreaking for unsupervised anomaly detection in medical imaging but suffered from instability and limited reconstruction detail. VQGAN addresses these drawbacks by using discrete latent representations combined with powerful transformer mechanisms [12]. Unlike AnoGAN, VQGAN provides stable training and generates more detailed and interpretable outputs. These improvements are crucial for detecting subtle anomalies, such as early-stage tumors, small lesions, or tissue alterations in brain MRIs.

However, despite these promising developments, several practical challenges persist in the application of VQGAN to brain MRI analysis. Training VQGAN models remains computationally intensive, particularly for high-resolution images required in medical applications. Furthermore, the interpretability of generated results, although improved compared to traditional GANs, still requires refinement to ensure clinical acceptance. Enhancing the explainability of model outputs remains essential for translating research into practical clinical tools.

Another essential aspect that recent research has begun to address is the integration of synthetic data generated by VQGAN with federated learning frameworks. A novel approach that combined GAN-based synthetic data generation with federated learning. This method demonstrated significant improvements in diagnostic accuracy without compromising patient data privacy. Their work highlighted the potential of VQGAN to generate clinically useful synthetic data for federated learning applications, enhancing data availability and diversity across different clinical institutions [30].

In the broader context, VQGAN also enables conditional synthesis, where specific spatial or semantic information guides the generation process. For instance, the synthesis of high-resolution images guided by semantic segmentation maps, depth maps, or class labels, showcases the versatility and effectiveness of VQGAN in various controlled generation tasks. This capability to incorporate specific conditions into image generation opens exciting possibilities for generating tailored synthetic datasets for specialized clinical scenarios or training more robust segmentation models [16].

Additionally, ongoing research aims to enhance VQGAN models by refining their training stability and reducing the computational requirements. Techniques such as spectral normalization, self-attention mechanisms, and adaptive loss functions have been investigated to further stabilize GAN training and enhance anomaly detection performance. Spectral normalization, in particular, has proven effective in stabilizing training and improving convergence properties, enabling better and more reliable results in medical imaging applications.

Future developments of VQGAN in medical imaging are likely to explore hybrid architectures further, combining the strengths of transformers, CNNs, and advanced normalization techniques. These hybrid approaches could significantly improve anomaly detection accuracy, reduce computational costs, and enhance interpretability, making GAN-based methods more practical and clinically viable.

2.5 GAN-based Synthetic Data Generation in MRI

The emergence of Generative Adversarial Networks (GANs) has profoundly impacted medical imaging by addressing one of its most persistent challenges: the scarcity of high-quality annotated datasets. Synthetic data generation using GANs has become a pivotal strategy for augmenting datasets, thereby enhancing the training and generalization capability of machine learning models. This is especially crucial for MRI analysis, where acquiring large, annotated datasets is both time-consuming and costly. The generation of synthetic MRI data through GANs can significantly mitigate data limitations, promoting robust model training and improving diagnostic accuracy.

GANs consist of two primary neural networks—a generator that synthesizes new data and a discriminator that distinguishes real from generated images. Through an adversarial training process, the generator progressively learns to produce synthetic images indistinguishable from real data, making GANs exceptionally suitable for medical image augmentation. The realism and variability of synthetic images generated by GANs have proved particularly effective in improving the generalizability of deep learning models across different imaging conditions and patient populations.

One notable advancement is the development of conditional GANs (CGANs), which allow for controlled image synthesis by conditioning the generation process on additional information, such as class labels or segmentation maps. Hamghalam and Simpson presented a significant application of CGANs, termed Enhancement and Segmentation GAN (ESGAN), specifically designed for brain MRI data. ESGAN synthesizes high-contrast MRI images that minimize intensity overlaps between various tumor tissues, greatly enhancing segmentation performance. This model leverages an adaptive calibration mechanism and employs a classifier loss combined with adversarial training to enhance the contrast and distinctness of tumor tissues, directly addressing the common problem of overlapping intensity distributions in MRI scans [9].

Similarly, the powerful impact of GAN-based synthetic data generation on medical image classification tasks. Their work involved generating synthetic images of liver lesions using GANs, which significantly improved the performance of CNN classifiers trained on limited datasets. Their study conclusively showed that synthetic data could effectively augment small datasets, enhancing model accuracy and reducing overfitting. Such outcomes underscore the potential of GANs to address data scarcity in MRI-based diagnostic tasks effectively [33].

Further advancing this concept, is the integration of GAN-generated synthetic data within federated learning frameworks. Federated learning enables collaborative training across multiple institutions without compromising patient data privacy. Alalwan et al. employed GANs to generate synthetic MRI data that closely mimicked real data distributions, subsequently using this data to train federated CNNs. Their method significantly improved diagnostic accuracy while maintaining data privacy, highlighting the practical benefits of synthetic data generation not only in addressing data scarcity but also in enhancing collaborative research opportunities [34].

In another impactful study, Han et al. (2018) utilized GAN-based augmentation to enrich brain MRI datasets specifically for glioma segmentation tasks. By generating synthetic images that capture complex pathological features, their approach improved segmentation model performance significantly. This method addressed issues related to limited patient samples and enhanced model generalization capabilities by presenting diverse synthetic cases representing various tumor characteristics and imaging scenarios [35].

The literature also underscores the challenges inherent in GAN-based synthetic data generation for MRI. Ensuring clinical realism is critical—synthetic images must faithfully replicate anatomical structures and pathological features without introducing misleading artifacts. Achieving such high fidelity requires meticulous fine-tuning of GAN architectures and rigorous validation. Techniques like Wasserstein GANs (WGANs) and progressive growing GANs have been introduced to improve stability and image quality, addressing typical GAN training issues like mode collapse and instability.

Moreover, GAN-based approaches have been expanded by integrating advanced architectural modifications to improve the quality and realism of generated data further. Karras et al. (2018) introduced Progressive Growing GANs (PGGANs), which progressively increase the resolution of generated images during training. This methodology significantly improved the synthesis of high-resolution medical images, maintaining stability and enabling the generation of detailed anatomical structures critical for precise medical analyses.

In addition, Esser et al. (2021) presented VQGAN, which leverages vector quantization and transformers to enhance the fidelity and contextual coherence of synthetic images. Their method significantly improves GAN-generated image quality, making synthetic MRI data highly detailed and clinically relevant. Such advancements in architectural design demonstrate the importance of combining sophisticated neural network techniques to achieve superior synthetic data quality.

Challenges in generating clinically acceptable synthetic MRI data remain significant despite these advancements. Validating the clinical utility of synthetic images requires rigorous assessment to ensure they accurately represent real-world pathology. This validation typically involves expert clinical evaluations and extensive comparative analyses with real clinical data, demanding substantial interdisciplinary collaboration between computational scientists and clinical experts.

Recent efforts have emphasized evaluating the clinical realism of GAN-generated data. Wang et al. (2020) systematically evaluated synthetic MRI data for clinical applicability using radiological expert assessments. Their study revealed that although GAN-generated synthetic images closely resembled real MRI data visually, subtle artifacts and unrealistic textures sometimes persisted, highlighting the critical need for rigorous validation in clinical settings.

3 Theoretical Background

3.1 Fundamentals of Generative Adversarial Networks (GANs)

In 2014, Ian Goodfellow and his team introduced a groundbreaking approach known as Generative Adversarial Networks (GANs), which enables the generation of highly realistic synthetic data, particularly in the form of images, utilizing an adversarial training framework. GANs represent a pivotal advancement in the field of generative modeling, transitioning from conventional explicit density estimation to the implicit learning of intricate data distributions. This is achieved through the competitive interaction between two neural networks: the generator and the discriminator [4].

The fundamental concept underlying GANs is both simple and impactful. The generator network endeavours to produce synthetic images from random noise, with the objective of closely resembling actual data. Conversely, the function of the discriminator is to differentiate between authentic images and those generated synthetically by the generator. The adversarial training process compels both networks to consistently improve their functions, ideally resulting in the generator creating images that are indistinguishable from real data [4].

$$\min_G \max_D V(D,G) = E_{x\sim p_{data}(x)}[\log D(x)] + E_{z\sim p_z(z)}[\log(1 - D(G(z)))]$$

where:

• G (Generator) learns to map a latent variable z (sampled from a prior distribution pz (z)) to a data space, generating synthetic samples G(z).

• D (Discriminator) is a classifier that distinguishes between real data x ~pdata(x) and generated data G(z).

• $E_{\{(x\sim pdata(x))\}}[logD(x)]$ encourages the discriminator D to correctly classify real samples as real.

• $E_{z\sim pz\,(z)}\left[\log\left(1 - D\big(G(z)\big)\right)\right]$ encourages the generator G to generate samples that fool the discriminator into classifying them as real.

3.1.1 GAN Components and Training

The fundamental structure of GANs includes two primary neural network components:

Generator (G): A neural network that synthesizes data. The input to the generator is typically a random noise vector drawn from a Gaussian or uniform distribution, transformed through multiple layers into a realistic image that resembles the true data distribution.

$$J_G = -\frac{1}{m}\sum_{i=1}^{m} logD(Gz_i))$$

Where:

J_g measures how well the generator is fooling the discriminator

z_i is the latent noise sampled from a prior distribution $p_z(z)$

$G(z_i)$ is the probability that the discriminator assigns to the generated sample being real.

The generator aims to minimize this loss, which encourages the production of samples that the discriminator classifies as real (i.e., $D\big(G(z_i)\big)$approaches 1).

Discriminator (D): A neural network trained to accurately distinguish between real data samples and synthetic samples produced by the generator. The discriminator outputs a probability that the input is real.

$$J_d = -\frac{1}{m}\sum_{i=1}^{m}\Big[logD(x_i) + \log\Big(1 - D\big(G(z_i)\big)\Big)\Big]$$

Where:

J_D measures how well the discriminator can classify real vs. generated samples.

x_i is a real image from the dataset.

$D(x_i)$ represents the probability that the discriminator correctly classifies a real image.

$1 - D\big(G(z_i)\big)$ represents the probability of correctly identifying a generated image as fake.

The discriminator aims to maximize this function, improving its ability to differentiate between real and fake images.

3.2 Variants of GANs:

3.2.1 Deep Convolutional GANs (DCGAN)

DCGAN, proposed by Radford et al. was among the first significant architectural improvements to standard GANs. DCGAN incorporated convolutional neural networks (CNNs) into both the generator and discriminator, achieving superior performance in generating high-quality images. Key innovations included the use of batch normalization, ReLU activations in the generator, and LeakyReLU activations in the discriminator. These advancements notably stabilized training and enhanced image realism [5].

3.2.2 Wasserstein GAN (WGAN)

Wasserstein GAN, introduced by Arjovsky et al. employed the Wasserstein distance (also known as Earth Mover's distance) to improve GAN training stability. WGAN provided more informative gradient feedback to the generator, significantly reducing mode collapse and vanishing gradient issues.[6]. Additionally, incorporating gradient penalty methods Gulrajani et al. further stabilized training, making WGAN a widely adopted variant, especially suitable for high-quality medical image generation tasks [7].

3.2.3 Conditional GAN (CGAN)

Conditional GANs, initially introduced by Mirza and Osindero, expanded the versatility of GANs by enabling image generation conditioned on external inputs such as class labels or segmentation masks. This conditional approach guides the generator to focus on specific subsets of data distributions, making it particularly beneficial for medical imaging tasks, such as tumor segmentation and anomaly detection. By generating targeted synthetic datasets, CGAN significantly enhances the effectiveness of diagnostic models in clinical contexts [9][10].

3.2.4 Progressive Growing GAN (PGGAN)

The Progressive Growing GAN approach, proposed by Karras et al., marked another critical enhancement in GAN technology. This method involves training the neural networks progressively, starting from low-resolution images and incrementally increasing the complexity and resolution of the generated images. Such progressive training stabilizes the learning process, significantly enhancing the realism of high-resolution synthetic images. This capability is particularly valuable in medical imaging, where detailed, high-resolution synthetic datasets are essential for accurate diagnostic evaluations and modeling [10][11].

3.2.5 Practical Implications in Medical Imaging

Variants of GANs have demonstrated substantial potential within medical imaging, particularly in data augmentation, anomaly detection, and assisting diagnostics. GAN-generated synthetic images address critical challenges, including data scarcity and patient privacy concerns, both of which significantly impact the development of robust medical machine learning models. For instance, synthetic MRI scans created using GAN techniques have notably improved diagnostic accuracy and model generalization across varied patient demographics, ultimately contributing to enhanced clinical outcomes [9][12].

3.3 Vector Quantization and VQGAN+D

In the realm of medical imaging, the accurate reconstruction of images is paramount for effective diagnosis and treatment planning. Traditional GANs have made significant strides in this area; however, they often grapple with challenges related to capturing intricate details and ensuring high-fidelity reconstructions[13]. The advent of Vector Quantized GANs (VQGANs) has introduced a novel approach that leverages discrete latent representations to enhance image quality[16]. Building upon this foundation, the integration of a discriminator—referred to as VQGAN+D—has further refined the capabilities of these models, particularly in the context of anomaly detection.

3.3.1 Understanding VQGAN: Architecture and Functionality

VQGAN represents a sophisticated fusion of vector quantization techniques with the adversarial training paradigm of GANs[3]. At its core, VQGAN comprises three primary components: an encoder, a vector quantizer, and a decoder. The encoder processes input images to produce continuous latent representations, which are subsequently mapped to discrete codes by the vector quantizer[10]. These discrete codes are then transformed back into the image space by the decoder, aiming to reconstruct the original input with high fidelity.

The vector quantizer plays a pivotal role by enforcing a discrete latent space, effectively capturing essential image features while mitigating issues such as blurriness often observed in traditional autoencoders [3]. This discrete representation facilitates the model's ability to learn a compact and informative codebook, enhancing the quality of generated images. The adversarial component of VQGAN introduces a discriminator that distinguishes between real and reconstructed images, compelling the generator to produce outputs that are not only accurate but also perceptually convincing [1]. This adversarial training encourages the generator to capture fine-grained details, leading to more realistic reconstructions.

3.3.2 Enhancing Anomaly Detection: The Role of the Discriminator in VQGAN+D

The integration of a discriminator in VQGAN, resulting in VQGAN+D, significantly bolsters the model's efficacy in anomaly detection within medical imaging [1]. The discriminator serves a dual purpose: it not only differentiates between real and generated images but also provides valuable feedback to the generator, guiding it towards producing more accurate and realistic reconstructions [3].

In the context of anomaly detection, VQGAN+D operates under the premise that the model is trained predominantly on normal (i.e., non-anomalous) medical images [2]. During inference, when the model encounters an image containing anomalies, the reconstruction process tends to falter, as the model is unfamiliar with these aberrations. Consequently, the reconstructed image deviates from the input, particularly in regions corresponding to anomalies. By analyzing the discrepancies between the input and its reconstruction, clinicians can pinpoint areas of interest that may warrant further investigation [1].

The discriminator's feedback is crucial in this process. It ensures that the generator maintains high-quality reconstructions of normal anatomical structures, thereby amplifying the contrast between normal and anomalous regions. This heightened sensitivity to anomalies enhances the model's diagnostic utility, enabling the detection of subtle pathological changes that might elude traditional imaging analyses.

3.3.3 Advantages of Discrete Latent Representations in Medical Image Reconstruction

The adoption of discrete latent representations within VQGAN+D offers several compelling advantages for medical image reconstruction:

- **Enhanced Image Fidelity:** The discrete latent space encourages the model to learn a finite set of prototypical image features, which promotes sharper and more accurate reconstructions. This is particularly beneficial in medical imaging, where the clarity of anatomical details is critical for diagnosis [2].
- **Improved Anomaly Detection:** By focusing on normal image patterns during training, the model becomes adept at reconstructing typical anatomical structures. Anomalous regions, being unfamiliar, are poorly reconstructed, resulting in noticeable discrepancies that can be leveraged for anomaly detection [3].
- **Efficient Representation:** Discrete codes facilitate a more compact representation of images, reducing computational complexity and storage requirements. This efficiency is advantageous in clinical settings where large volumes of imaging data are routinely processed.
- **Robustness to Noise:** The quantization process inherently filters out minor variations and noise, leading to more stable and reliable reconstructions. This robustness ensures that clinically relevant features are preserved while irrelevant artifacts are suppressed.

3.3.4 Applications and Implications in Clinical Practice

The integration of VQGAN+D into clinical workflows holds significant promise:

Early Detection of Pathologies: The model's sensitivity to anomalies enables the early identification of subtle pathological changes, facilitating prompt intervention and improved patient outcomes.

Reduction of False Positives: By accurately reconstructing normal anatomical structures, VQGAN+D minimizes the likelihood of false-positive detections, thereby reducing unnecessary follow-up procedures and associated patient anxiety [1].

Augmentation of Radiologist Expertise: As a decision-support tool, VQGAN+D can assist radiologists by highlighting regions of interest, streamlining the diagnostic process, and potentially reducing workload [3].

Conclusion:

The evolution from traditional GANs to VQGAN and subsequently VQGAN+D represents a significant advancement in medical image reconstruction and anomaly detection. The incorporation of discrete latent representations, coupled with the guiding influence of a discriminator, enhances the model's ability to produce high-fidelity reconstructions and detect anomalies with greater precision. These developments hold substantial potential for improving diagnostic accuracy and efficiency in medical imaging, ultimately contributing to better patient care.

4 Proposed methodology

4.1 Dataset

The dataset utilized in this thesis consists of a carefully selected combination from three publicly available MRI brain scan collections: Figshare, Sartaj, and BR35H[]. Specifically, this research focuses exclusively on images classified as either healthy or glioma-affected, deliberately excluding other tumor types. 2000 healthy images and 1621 glioma images are used. This selective approach streamlines the anomaly detection process, enabling generative models to better differentiate subtle pathological features characteristic of gliomas from healthy brain tissue.

Gliomas, a common type of malignant brain tumor, vary significantly in size, shape, and intensity across MRI scans, thereby posing challenges for effective automated detection and diagnosis. Combining multiple datasets helps achieve a broader spectrum of variation in anatomical and pathological features, enhancing the robustness of trained models by reducing susceptibility to overfitting to specific imaging protocols or individual datasets.

The selected datasets collectively contribute diverse images crucial for training robust generative models. This diversity improves the models' generalization capability, enabling more accurate identification of anomalies across varying imaging conditions and scanners. Figure [1] presents two examples of healthy brain MRI images, showcasing typical characteristics without pathological features. Figure [2] illustrates two examples of glioma-affected MRI scans, clearly highlighting tumor regions and emphasizing the visual complexity involved in accurate automated detection.

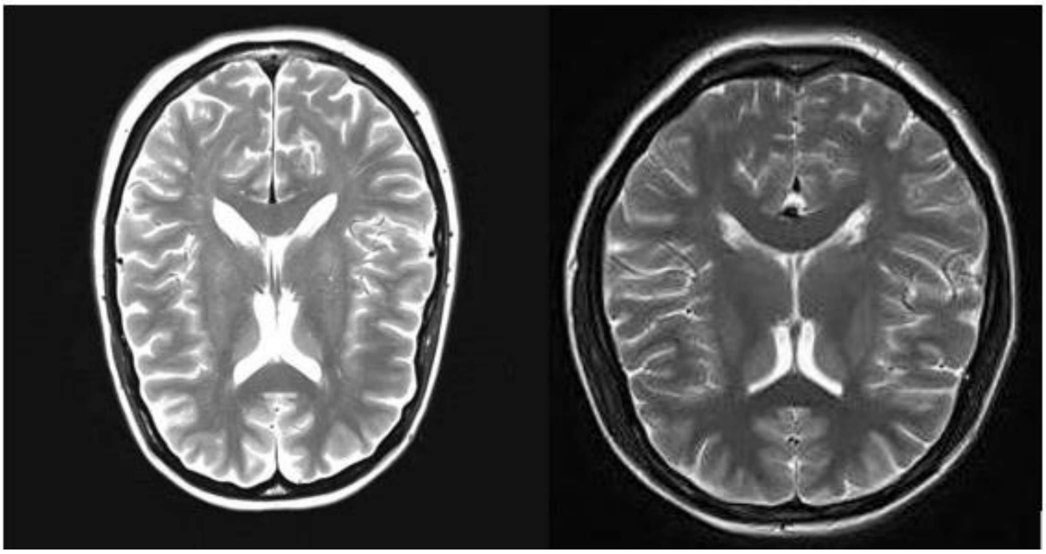

Figure 1. Healthy Brain Scans

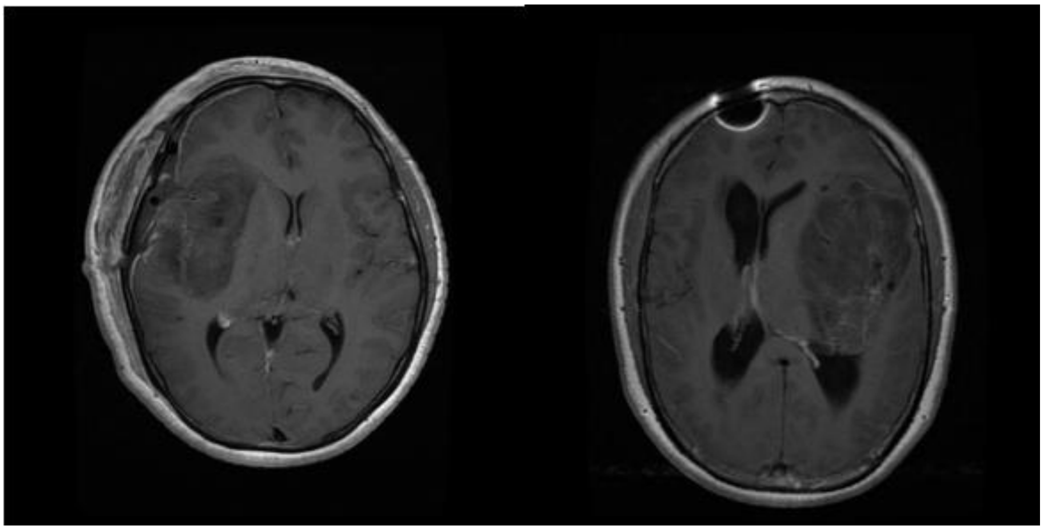

Figure 2. Anomaly Brain Scans

These images are extracted from real-world clinical MRI scans, making them highly relevant for practical anomaly detection applications. The dataset's availability in standardized grayscale format simplifies preprocessing and ensures compatibility with deep learning models.

4.2 Implementation Details and Tools

The VQGAN+D-based anomaly detection system calls for a strong and ordered programming environment including image processing tools, numerical computation libraries, and deep learning frameworks. Every element in this work is very important for designing the architecture of VQGAN+D, managing picture transformations, running tensor-based computations, and visualizing model performance. Choosing the suitable software tools guarantees computational efficiency, steady training, and consistent anomaly identification considering the complexity of MRI-based anomaly detection.

4.2.1 PyTorch

PyTorch, one of the most popular deep learning frameworks, is used to build and train the VQGAN+D model. Designed by Facebook AI Research (FAIR), PyTorch is well-known for its dynamic computation graph, which permits flexible model creation and effective debugging. PyTorch offers an easy approach to create neural networks unlike conventional static graph-based architectures, which is perfect for VQGAN+D as several connected components—such as the generator, discriminator, and vector quantizer—must be simultaneously optimized [17].

PyTorch's automatic gradient computation and backpropagation simplification method (Autograd) is a main benefit. Training VQGAN+D depends critically on this ability since it guarantees effective parameter updates for the generator and discriminator networks. PyTorch also makes use of CUDA (Compute Unified Device Architecture), therefore enabling computations on NVIDIA GPUs, greatly lowering training time. Using PyTorch in tandem with GPU acceleration guarantees scalability and efficiency since VQGAN+D consists in computationally intensive adversarial learning and high-density image synthesis.

PyTorch defines the VQGAN+D architecture in this work containing the encoder, vector quantizer, and decoder components. Crucially for training an anomaly detection system, it also helps to compute adversarial loss, perceptual loss, and reconstruction loss. The tensor-based operations in PyTorch further facilitate smooth model convergence by providing faster gradient computations. This study's use of PyTorch guarantees a scalable and effective training pipeline, enabling the model to produce excellent MRI reconstructions while maintaining important structural details.

4.2.2 Torchvision

Mostly used for MRI image preprocessing and feature extraction, Torchvision is a necessary expansion of PyTorch. Preprocessing is a crucial step in guaranteeing consistent model training since MRI datasets sometimes include images of different resolutions, aspect ratios, and intensity distributions. Before feeding MRI scans into VQGAN+D[18], Torchvision offers a set of image processing tools that assist to normalize and standardize them.

Torchvision is in charge of scaling MRI pictures to a set resolution of 128×128 pixels in this work, therefore guaranteeing that all input data keeps a consistent format. Moreover, pixel intensity normalisation is used to scale values between [-1,1] hence minimising excessive brightness and contrast fluctuations that can compromise training stability. Furthermore used are methods of data augmentation include random rotations, flips, and contrast changes to increase generalization and stop overfitting.

Feature extraction also depends critically on Torchvision. Well-known convolutional neural network VGG16 is used in perceptual loss computations to compare produced MRI pictures with actual ones. VQGAN+D can create high-quality reconstructions quite similar to real anatomical structures by extracting deep feature representations from MRI data. By means of these changes, Torchvision guarantees successful preprocessing of the input data thereby enabling the model to learn meaningful representations and detect abnormalities more precisely.

4.2.3 NumPy

Effective numerical computation is the foundation of deep learning, hence NumPy is the fundamental numerical computing instrument applied in this work. Management of high-dimensional MRI data depends critically on high-performance matrix operations, multidimensional array manipulations, and statistical computations made feasible by NumPy[19].

This work guarantees smooth compatibility with PyTorch tensors by converting MRI images into structured numerical arrays with NumPy. Since MRI scans show pixel intensity values that must be normalized and translated, NumPy provides the necessary mathematical operations to conduct scaling, matrix transformations, and statistical preprocessing. NumPy is also useful for computing model evaluation measures like Mean Squared Error (MSE) and Structural Similarity Index (SSIM), which both check the quality of reconstructed MRI images.

By cleverly mixing with PyTorch and Torchvision, NumPy guarantees fast and accurate tensor operations, hence reducing computational overhead.

4.2.4 OpenCV (cv2)

Preprocessing is essential for medical imaging, and OpenCV (cv2) is utilized in order to enhance MRI image quality before training. The noise, intensity swings, and artifacts regularly found in medical images could compromise the performance of VQGAN+D. Thanks to OpenCV's suite of image processing techniques, which

helps to minimize these issues [20], just the most important anatomical information are maintained.
Among OpenCV's primary tools in this work is Gaussian filtering, which reduces noise in MRI images while preserving fine structural details. Moreover, contrast normalizing techniques help to highlight brain areas by which minor anomalies are more easily found. Another essential preprocessing step done using OpenCV is skull stripping, the removal of non-brain areas from MRI data. Separating the brain tissue allows VQGAN+D to focus exclusively on analyzing relevant anatomical aspects, hence improving its ability to spot anomalies.

4.2.5 Matplotlib and Seaborn

Visualization is an essential component of assessing deep learning models, hence in this work the output of VQGAN+D is investigated with Matplotlib and Seaborn. These instruments give detailed understanding of model performance by enabling the presentation of loss curves, anomaly heatmaps, and reconstructed MRI scans. Showing MRI scan reconstructions using Matplotlib helps to compare generated and real images mostly. It also provides training loss plots, which help researchers track generator network convergence with the discriminator. One use of statistical data visualization with Seaborn, an extension of Matplotlib, is the distribution of anomaly detection scores over several MRI images. This work integrates several visualization tools to guarantee transparent and interpretable model performance.

4.2.6 TensorBoard

To monitor model training in real time, TensorBoard is utilized for tracking key performance metrics and visualizing intermediate results. Since VQGAN+D relies on adversarial training, tracking multiple loss functions simultaneously is crucial to diagnosing potential issues such as mode collapse, gradient vanishing, and unstable training dynamics [22].

In this study, TensorBoard logs loss values (adversarial loss, perceptual loss, and reconstruction loss) at each training step. Additionally, it visualizes model weight updates, allowing for early detection of training instability. TensorBoard’s built-in image visualization features also enable a direct comparison between real and generated MRI images, ensuring that the model is progressing toward higher-quality reconstructions.

By incorporating TensorBoard into the training pipeline, this study provides a transparent and structured approach to model evaluation, ensuring that anomalies are detected with high precision.

4.3 Preprocessing Techniques

Preprocessing is a crucial step in medical image analysis, particularly for deep learning models applied to MRI scans. Raw images often exhibit intensity variations, resolution inconsistencies, and scanner-dependent artifacts, which can negatively impact model performance. Standardizing these images through a structured preprocessing pipeline ensures better generalization and improved anomaly detection.

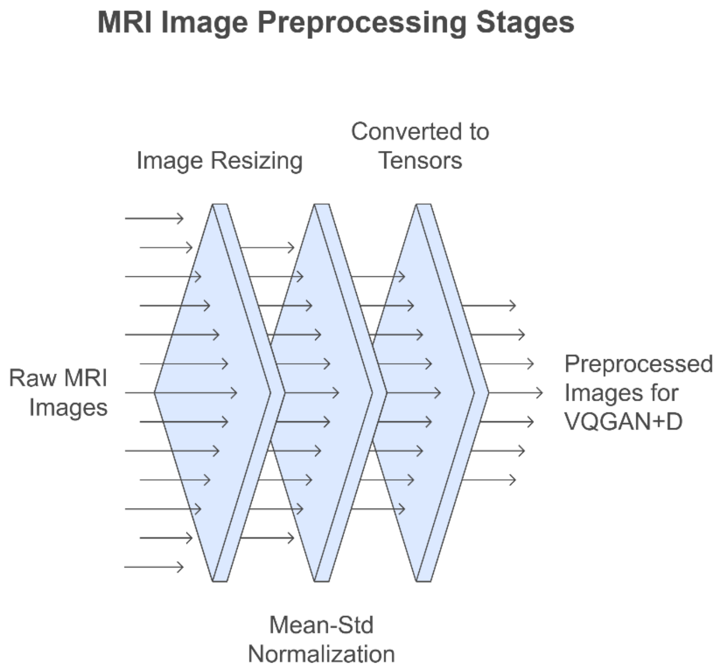

Figure 3. Preprocessing of input images

MRI images are subject to intensity variations caused by differences in imaging protocols, scanner hardware, and patient conditions. These inconsistencies can cause deep learning models to focus on insignificant brightness variations rather than learning meaningful anatomical features. To address this, z-score normalization (also called mean-std normalization) is applied using the following formula [17]:

$$I_{normalized} = \frac{I-\mu}{\sigma}$$

where:

$I_{normalized}$ is the normalized pixel intensity,

I represents the original intensity value,

$\mu = 0.5$ is the mean intensity of the dataset,

$\sigma = 0.5$ is the standard deviation of pixel intensities.

This normalization ensures that pixel values are centered around zero mean with unit variance, improving stability in gradient updates during training [1][2]. The method is widely used in medical imaging applications, including segmentation and anomaly detection, as it reduces scanner-specific biases [2].

4.3.1 Resizing for Standardized Input Dimensions

MRI datasets often contain images of varying resolutions due to different scanning conditions. Since deep learning models require fixed input dimensions, all images are resized to 128 × 128 pixels before being fed into the VQGAN+D model [24].

Resizing provides several benefits:

Ensures consistency across different scans, allowing stable feature extraction.

Reduces computational complexity, enabling efficient training.

Optimizes memory usage, preventing GPU memory overflow.

Bilinear interpolation is applied during resizing to preserve anatomical structures and prevent distortions, a technique commonly used in medical image preprocessing [24].

4.3.2 Conversion to Tensor Format

Deep learning models, especially those implemented in PyTorch, require images to be converted into tensor format before training. After resizing and normalization, the images are transformed into PyTorch tensors, which allow efficient matrix operations on GPUs. The conversion is performed as follows:

$$I_{tensor} = ToTensor(I_{normalized})$$

where:

I_{tensor} is the image in tensor format,

$I_{normalized}$ is the preprocessed image after resizing and normalization,

ToTensor() is a PyTorch function that converts the image from a NumPy array to a tensor representation.

Additionally, images are batched using PyTorch's DataLoader, which enables parallel processing and efficient memory handling [25].

4.3.3 Dataset Structure and Loading

The MRI dataset used in this study consists of two categories:

Normal Brain Scans – Healthy MRI images with no visible abnormalities.

Anomalous Brain Scans – Images containing tumors or structural anomalies.

All images are loaded in batches using PyTorch's DataLoader, which ensures that GPU memory is optimally utilized during training and inference [25]. The dataset is split into training and testing subsets, with the same preprocessing steps applied to both.

This preprocessing pipeline ensures that all images are uniform, standardized, and ready for anomaly detection using the VQGAN+D framework.

4.4 VQGAN+D Model Architecture

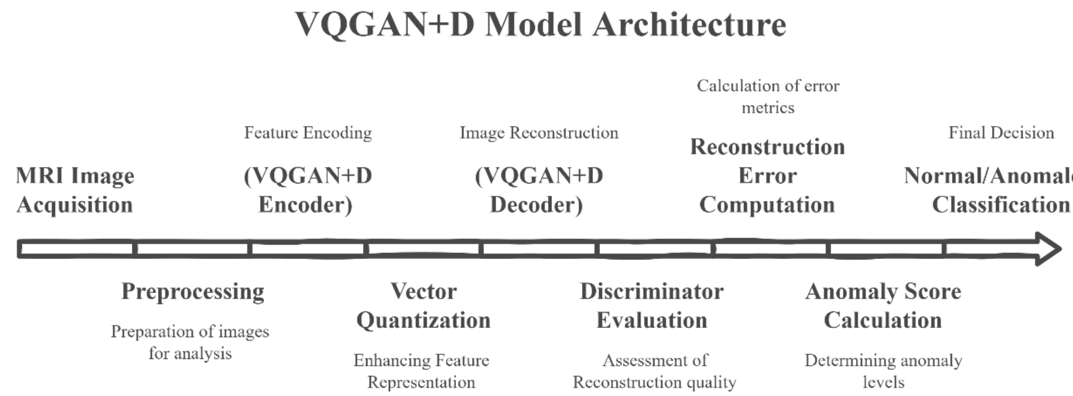

Figure 4. VQGAN+D Model Architecture

The VQGAN+D-based anomaly detection system is structured as a multi-stage pipeline that processes brain MRI images and detects anomalies through vector quantized feature representations and adversarial learning mechanisms. The system follows a logical progression, as illustrated in Figure 4, which begins with image acquisition and preprocessing, followed by feature encoding, vector quantization, image reconstruction, and anomaly scoring based on reconstruction error computation. The final output is a classification indicating whether the given MRI scan represents a normal or abnormal brain structure.

4.4.1 MRI Image Acquisition and Preprocessing

The initial stage of this workflow involves obtaining brain MRI scans from standardized medical datasets, including Figshare, SARTAJ, and Br35H [10]. These datasets provide a balanced collection of healthy and pathological MRI scans, which are essential for learning normal anatomical structures and evaluating the model's ability to detect unseen anomalies.

Once acquired, MRI scans undergo preprocessing to ensure consistency in format and quality. Medical imaging data often contains noise, intensity variations, and artifacts, which can interfere with deep learning models if not addressed appropriately. The preprocessing pipeline includes image resizing, pixel intensity normalization, and noise reduction techniques to enhance structural clarity while preserving essential brain features. Pre-processed images are then prepared for feature extraction through the VQGAN+D encoder. While this section provides an overview of preprocessing, further details are discussed in Section 4.5.

4.4.2 Feature Encoding Using VQGAN+D

Following preprocessing, the MRI images are passed through the VQGAN+D encoder, which converts them into a compact latent representation that captures key anatomical details. Unlike standard GAN architectures that rely on continuous latent spaces, VQGAN+D incorporates a discrete vector quantization mechanism, ensuring that fine-grained structural variations are effectively retained [1].

This encoding process is mathematically expressed as:

$$z = E(x)$$

Where:

$E(x)$ is the encoder function, transforming the input image x into its lower-dimensional latent representation z. This latent space representation is designed to retain medically significant features, allowing for efficient reconstruction while facilitating anomaly detection.

4.4.3 Vector Quantization: Enhancing Feature Representation

A fundamental improvement of VQGAN+D over traditional GANs is its vector quantization step, which ensures that high-resolution MRI features are discretized into a finite codebook of embeddings [3]. The model maps each encoded latent vector to its closest pre-learned prototype vector, allowing for better generalization and structural consistency during reconstruction.

This quantization step is formulated as:

$$z_q = arg_{e_i \in \varepsilon}^{min} ||z - e_i||$$

Where:

e_i are learned codebook embeddings.

ε is the set of all embeddings.

z_q is the closest discrete vector selected from the codebook.

The use of discrete latent representations allows VQGAN+D to maintain fine details in MRI scans while preventing feature loss that could impact anomaly detection performance [24].

4.4.4 The Role of the Codebook in Learning Discrete Features

The codebook serves as a lookup table of feature embeddings that store representations learned from the dataset. Rather than operating in a free-form latent space, the encoder's output is quantized into a discrete set of vectors, ensuring that feature representations remain structured and interpretable.

During training, the codebook embeddings are iteratively updated based on the encoder's output, refining the stored feature prototypes to better capture normal anatomical structures in MRI scans. This process prevents overfitting to noise and redundant details, allowing for more robust feature extraction [27].

To ensure stable learning, the codebook values are updated using an exponential moving average (EMA) rather than conventional gradient descent. The update equation for each embedding e_i is given by:

$$e_i^{(t+1)} = e_i^{(t)} + \lambda\left(z - e_i^{(t)}\right)$$

where:

$e_i^{(t)}$ is the current embedding at timestep t.

λ is the update coefficient (learning rate for the codebook).

z is the encoder output assigned to this codebook entry.

This technique ensures gradual adaptation of embeddings, maintaining stability in the codebook while allowing it to learn optimal feature representations [4].

(subheading inside vector quantization) Impact of the Codebook on Anomaly Detection

The discrete nature of the codebook is particularly useful for detecting anomalies in brain MRI scans. Since the codebook is trained on normal anatomical structures, it lacks representations for anomalous regions such as tumors, lesions, or abnormalities. When an MRI scan containing an unseen pathology is passed through the model, its feature representations do not align with any existing codebook embeddings, leading to high reconstruction errors.

In effect, the codebook acts as a constraint that forces the model to reconstruct images using only known feature embeddings. If an MRI scan contains previously unseen structures, the decoder fails to accurately reconstruct those regions, resulting in higher anomaly scores [24].

4.4.5 Image Reconstruction via VQGAN Decoder

Once the encoded MRI scan is mapped to its quantized form, it is passed through the VQGAN+D decoder, which reconstructs an approximation of the original image. The reconstruction process is designed to retain normal anatomical structures, ensuring that the generated images closely resemble real MRI scans. The decoder function is expressed as:

$$\hat{x} = G(z_q)$$

Where:

G(zq) is the generator function, reconstructing an output image $\hat{x}$ from the quantized latent representation.

Ideally, if the input MRI scan belongs to the normal distribution, the reconstructed image should be nearly identical to the original. However, if an anomalous region such as a tumor, lesion, or irregular tissue growth is present, the model fails to accurately reconstruct those regions, leading to noticeable discrepancies between the input and output images.

4.4.6 Role of the Discriminator in Evaluating Reconstruction Quality

The discriminator in VQGAN+D plays a crucial role in evaluating the quality of the reconstructed images by distinguishing between real MRI scans and their generated counterparts [24]. This adversarial component ensures that the generator produces high-fidelity images that accurately capture structural details of the brain. The discriminator provides feedback signals, helping refine the generator's ability to reconstruct realistic MRI scans while preserving medically significant features.

4.4.7 Anomaly Detection via Reconstruction Error Computation

Anomaly detection in VQGAN+D is based on the principle that healthy MRI **scans** should be reconstructed with minimal error, whereas abnormal scans—containing tumors, lesions, or other pathological structures—will exhibit higher reconstruction errors [26]. The system computes reconstruction loss, measuring the difference between the original and reconstructed images. This difference is used to calculate an anomaly score, indicating whether an MRI scan is normal or abnormal.

If the anomaly score exceeds a defined threshold, the MRI scan is classified as abnormal, suggesting the presence of a potential pathological condition. Conversely, low reconstruction errors indicate that the MRI scan belongs to the learned distribution of healthy brain structures.

4.4.8 Final Decision: Classifying MRI Scans as Normal or Anomalous

The final output of the VQGAN+D anomaly detection pipeline is a classification decision indicating whether the given MRI scan is normal or contains anomalies. The decision is based on the reconstruction error analysis and discriminator evaluation, providing a structured and reliable framework for unsupervised medical anomaly detection.

By leveraging vector quantization, adversarial training, and reconstruction-based anomaly detection, VQGAN+D offers a powerful tool for identifying subtle structural deviations in brain MRI scans. This capability is particularly valuable in early-stage tumor detection and automated radiological assessments, where precision and sensitivity are paramount.

4.5 Training Procedure

The effectiveness of the VQGAN+D anomaly detection model heavily depends on a structured and carefully orchestrated training procedure. The core training strategy revolves around optimizing the generator and discriminator networks through a series of interconnected loss functions, each critical to capturing meaningful features and producing accurate image reconstructions. This section comprehensively details the specific loss functions, network architectures, and optimization strategies aligned with the implemented model.

4.5.1 Generator (VQGAN) Loss Function

The generator network in the VQGAN+D framework aims to reconstruct MRI scans that closely match their original inputs. To achieve this, a composite loss function comprising L1 Loss (Reconstruction Loss), Perceptual Loss (Multi-Scale Feature Loss), and Adversarial Loss is employed.

The Reconstruction Loss, implemented using an L1 Loss, ensures pixel-level accuracy in the reconstructed images and is mathematically defined as:

$$L_{L1(x,\hat{x})} = \frac{1}{N}\sum_{i=1}^{N} |x_i - \hat{x_i}|$$

Where:

x is the original MRI image

$\hat{x}$ is the reconstructed image.

This loss encourages the model to preserve the essential details of the original images and prevents distortion, which is vital for accurate medical anomaly detection [16].

The perceptual loss in this study is uniquely implemented using a custom-designed Multi-Scale Perceptual Loss Network rather than traditional pretrained models like VGG16. This network features two parallel branches with convolutional layers of different kernel sizes, alongside a residual connection that directly maps the input into the feature space. Mathematically, the Perceptual Loss is formulated as:

$$L_{perc}(x,\hat{x}) = \left||\emptyset(x) - \emptyset(\hat{x})|\right|^{\{2\}}$$

Here:

ϕ(·) denotes the feature extraction operation carried out by the custom-defined Multi-Scale Perceptual Loss Network. This innovative network significantly improves the model's sensitivity to structural and multi-scale details, essential for effective anomaly detection in MRI scans.

Additionally, the Adversarial Loss component employs the Binary Cross-Entropy (BCE) loss, aiming to improve the realism of the generated reconstructions. It is defined as:

$$L_{adv} = -\log\left(D(\hat{x})\right)$$

This loss compels the generator to produce MRI reconstructions indistinguishable from genuine scans, enhancing their fidelity.

4.5.2 Discriminator Loss Function

The discriminator is trained to distinguish between real MRI scans and those reconstructed by the VQGAN. Its primary objective is implemented through the Binary Cross-Entropy (BCE) loss:

$$L_{BCE} = -\frac{1}{2}[\log\big(D(x)\big) + \log\big(1 = D(\hat{x})\big)][4]$$

The discriminator's role is adversarial, continuously pushing the generator to improve the quality and authenticity of its reconstructions, thus directly enhancing anomaly detection accuracy.

4.5.3 Optimization Strategy and Hyperparameters

The choice of an optimization strategy and careful selection of hyperparameters significantly influence the performance, stability, and speed of training deep learning models. In this VQGAN+D framework, the Adam optimizer is employed for training both the generator (VQGAN) and the discriminator networks. Adam, introduced by Kingma & Ba (2014), is an adaptive optimization method particularly popular in deep learning due to its ability to handle noisy gradients and its robustness across various types of model architectures and loss landscapes.

4.5.4 Adam Optimizer: Overview and Functionality

Adam, which stands for Adaptive Moment Estimation, combines the advantages of two other extensions of stochastic gradient descent—AdaGrad and RMSProp. Specifically, it calculates adaptive learning rates for each parameter based on first-order (mean) and second-order (variance) moments of gradients. This adaptation ensures smoother and faster convergence during the training of neural networks.

Mathematically, Adam computes adaptive learning rates for each parameter as follows:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \text{ [39]}$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \text{[39]}$$

Where:

m_t and v_t represent the exponential moving averages of gradients and squared gradients, respectively, at timestep t.

β_1 and β_2 are hyperparameters controlling the decay rates of these moving averages.

g_t represents the gradient at time step t.

The bias-corrected moment estimates are then used to update the parameters, θ

$$\widehat{m}_t = \frac{m_t}{1-\beta_1^t} \quad \hat{v}_t = \frac{v_t}{1-\beta_2^t}$$

The model parameters are updated using:

$$\theta_t = \theta_{t-1} - \alpha \frac{\widehat{m}_t}{\sqrt{\hat{v}_t + \epsilon}}$$

Here, α is the learning rate, $\beta_1 and\, \beta_2$ control the exponential decay rates of the moment estimates, and ϵ prevents division by zero.

Hyperparameter Configuration

In this implementation, the Adam optimizer is configured with a learning rate of 2×10^{-4} along with moment parameters β_1=0.5 and β_2=0.999. The learning rate of 2×10^{-4} was chosen to provide stable convergence, facilitating gradual and steady updates to the model parameters. Selecting β_1=0.5 and β_2=0.999 has become common practice in GAN training due to their effectiveness in stabilizing training and preventing oscillations in loss values, enabling smoother convergence toward optimal parameter configurations.

Gradual Increase of Adversarial Loss Weighting

A critical and unique aspect of the training process in this study is the incremental adjustment of the adversarial loss weight throughout training epochs. Initially set at a low value, this weight progressively increases, allowing the model to first focus on accurate and stable image reconstruction before gradually incorporating adversarial pressure to improve realism. This weighting strategy helps alleviate the instability issues commonly associated with GAN-based training, including mode collapse and vanishing gradients.

The adversarial weight increment is mathematically defined as:

$$adv\_weight = \min(0.1 + epoch \times 0.0005, 0.5)$$

In practice, the adversarial weight starts at 0.1 at the beginning of training, then incrementally increases with each epoch until it reaches a predefined maximum of 0.5 Such gradual modulation of the adversarial component ensures a balance between structural fidelity and perceptual realism, critical for robust anomaly detection in MRI scans.

4.5.5 Training Process and Checkpoints

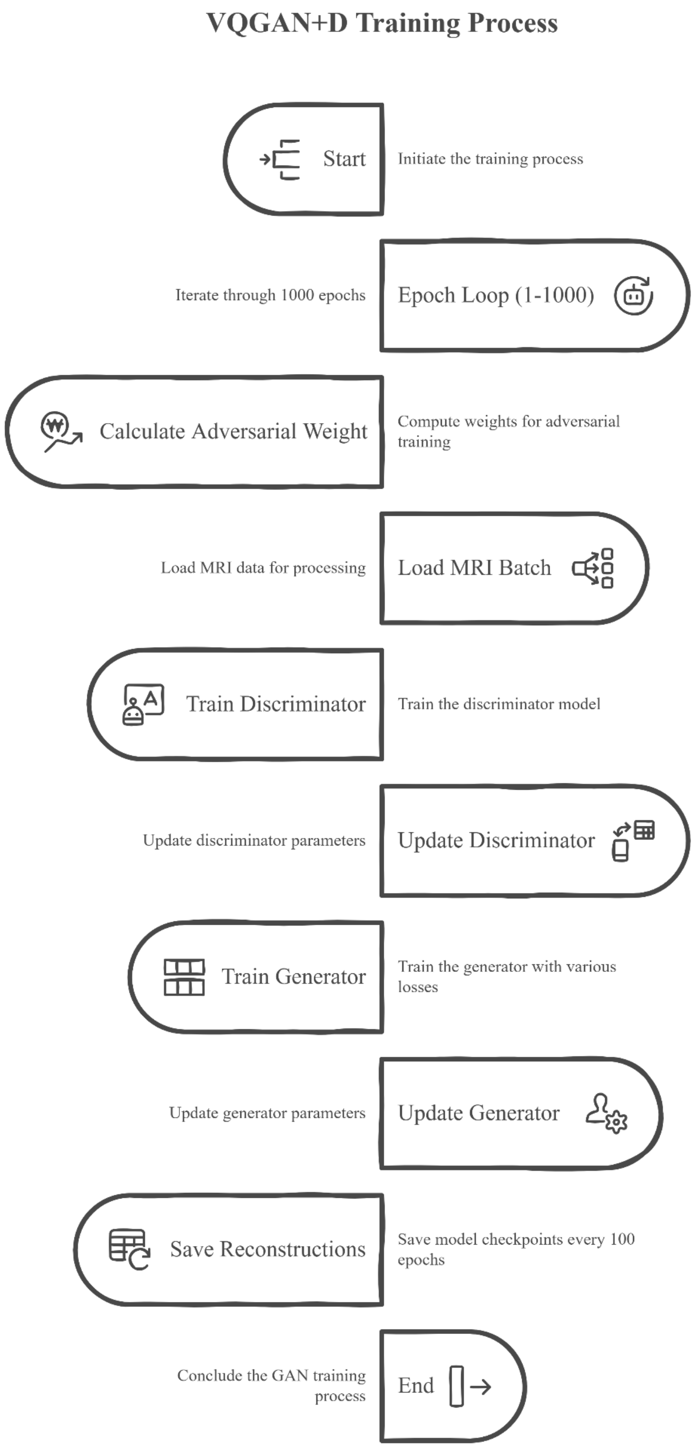

Figure 5. Training process of VQGAN+D

The training procedure for the VQGAN+D model is shown in Figure 5. Which consists of a structured process spanning 1000 epochs. At the start of each epoch, an adversarial weight is dynamically adjusted, increasing gradually from an initial value of 0.1 and capping at 0.5. This gradual increment ensures a balanced emphasis, initially on accurate reconstruction and later on enhancing the realism of the reconstructions, significantly stabilizing the adversarial training process.

Within each epoch, MRI images are loaded batch-wise from the dataset. First, the discriminator is trained by distinguishing between real MRI images and those reconstructed by the VQGAN. The discriminator's objective is quantified using Binary Cross-Entropy (BCE) loss, encouraging it to correctly classify real images as authentic and reconstructed images as artificial.

Following this, the VQGAN (generator) creates reconstructions of the input MRI images, known as fake images. These images are evaluated by the discriminator, and an adversarial loss for the generator is calculated, incentivizing the generator to produce reconstructions indistinguishable from real scans. Alongside this, the reconstruction loss (L1 loss) measures pixel-wise differences, ensuring high structural fidelity in the reconstructed images. Additionally, the custom multi-scale perceptual loss, as previously explained, ensures the preservation of essential perceptual features.

The generator's total loss is computed as a weighted combination of these three components—reconstruction loss, perceptual loss, and adversarial loss—weighted progressively to increase adversarial realism over the training period. This approach allows the generator to initially emphasize accurate reconstruction and gradually integrate more realistic detail generation as training progresses, enhancing the robustness and accuracy of anomaly detection (Kingma.

Periodic checkpoints at intervals of 100 epochs and specifically at epoch 500 provide visual confirmation of model progress by saving sample reconstructions. These checkpoints allow monitoring of incremental improvements in image reconstruction quality and offer insights into the convergence behavior of the training procedure.

Justification and Impact on Model Performance

The carefully designed optimization strategy and specific hyperparameter choices significantly impact the model's overall performance. Utilizing Adam optimizer with adaptive learning rates ensures reliable and rapid convergence, helping avoid common pitfalls in GAN training, such as instability and convergence difficulties. The progressive weighting scheme for adversarial loss further enhances the generator's capacity to initially focus on reconstructive accuracy and progressively emphasizes producing realistic details. This controlled approach directly contributes to the improved accuracy and robustness of the VQGAN+D model in detecting anomalies within MRI images, aligning with the overarching objectives of this research.

4.6 Testing and Evaluation

After training, the VQGAN+D model is evaluated to measure its effectiveness in anomaly detection on brain MRI scans. This section details the testing procedure, the anomaly detection process, and the evaluation metrics computed from the testing code.

4.6.1 Anomaly Detection Process

Once the VQGAN+D model is trained, the saved model is loaded, and it is tested on anomaly scans. The goal is to determine whether a given MRI scan is normal or contains anomalies based on its reconstruction error. Since the model learns to generate images that match the distribution of healthy MRI scans, any deviations in the reconstruction indicate potential anomalies.

- **Passing Input Images through the Model**
  Each test image is fed into the trained VQGAN+D model, which attempts to reconstruct it using the discrete latent representations. The quality of the reconstructed image is then compared to the original MRI scan to detect abnormalities.
- **Computing the Anomaly Score**
  The anomaly score is a measure of how much the reconstructed image deviates from the original scan. A higher anomaly score suggests an abnormal scan. However, in this proposed methodology, the scans having less anomaly score is classified as anomaly because the score is computed using three primary metrics:
  Mean Squared Error (MSE)
  Measures the pixel-wise difference between the input MRI scan and its reconstructed version.
  $$\mathrm{MSE} = \frac{1}{\mathrm{N}} \sum \left(\mathrm{I}_{\mathrm{original}} - \mathrm{I}_{\mathrm{reconstructed}}\right)^2$$
  Where:
  $I_{orginal}$ is the input MRI image.
  $I_{reconstructed}$ is the image generated by VQGAN+D
  N is the total number of pixels in the image.
- **Perceptual Loss (Using VGG16 Features)**
  Instead of just comparing pixel values, perceptual loss compares high-level feature representations. The features are extracted from an intermediate layer of a pre-trained VGG16 model, ensuring the anomaly detection captures structural differences rather than minor pixel variations. This makes VQGAN+D more robust to minor noise but sensitive to significant structural changes in MRI scans.

$$L_{perceptual} = \sum_{l} \left|\left|\emptyset_l(I_{orginal}) - \emptyset_l(I_{reconstructed})\right|\right|^2$$

Where:

$\emptyset_l$ represents the activation map at layer l of the VGG16 network

$I_{orginal}$ is the input mri image

$I_{reconstructed}$ is the reconstructed MRI image

- **Discriminator Confidence Score**

  The discriminator from VQGAN+D is also used to judge how "real" the reconstructed image looks.

  If the discriminator assigns a low confidence score to the reconstructed image, it suggests the model struggled to generate a realistic version, likely indicating an anomaly.

$$\mathrm{L_D} = \mathrm{E}_{\mathrm{x}\sim\mathrm{P_{real}}}[\mathrm{logD(x)}] - \mathrm{E}_{\mathrm{x}\sim\mathrm{P_{fake}}}[\mathrm{log}(1 - \mathrm{D}(\hat{\mathrm{x}}))]$$

  Where:

  D(x) is the discriminator output for the real MRI image.

  $D(\hat{x})$ is the discriminator output for the reconstructed MRI

  The loss encourages D(x) to close to 1(real) $D(\hat{x})$ to be close to 0(fake).

4.6.2 Adaptive Thresholding for Anomaly Classification

The final classification (normal vs. anomalous) is determined using an adaptive threshold on the computed anomaly scores.

A threshold value of 8 is predefined in the testing code.

The classification logic is:

If the anomaly score < 8 → The scan is classified as normal.

If the anomaly score ≥ 8 → The scan is classified as anomalous**.**

The threshold acts as a decision boundary to separate normal and anomalous MRI scans. Higher threshold reduces false positives but may miss subtle anomalies. lower threshold improves sensitivity but may flag more false positives.

4.6.3 Evaluation Metrics

To measure the effectiveness of VQGAN+D for anomaly detection, various evaluation metrics are computed based on the predicted vs. actual labels in the dataset.

- **Accuracy**

  Measures the proportion of correct predictions (both normal and anomalous).

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

  High accuracy indicates the model is making correct classifications.

- **Precision, Recall, and F1-Score**

  Precision: Measures how many of the predicted anomalies were actually anomalous.

  $$\text{Precision} = \frac{TP}{TP+FP}$$

  Recall (Sensitivity): Measures how many of the actual anomalies were correctly detected.

  $$\text{Recall} = \frac{TP}{TP+FN}$$

  F1-Score: Harmonic mean of precision and recall (balances false positives and false negatives).

  $$\text{F1} = 2 \times \frac{Precision \times Recall}{Precision \times Recall}$$

  These matrics help to know how well the prediction is happening such as High precision means the model rarely misclassifies normal scans as anomalous. If High recall means the model detects most actual anomalies. F1-score balances both, making it useful for medical anomaly detection.

- **Confusion Matrix Visualization**

  A confusion matrix provides a visual representation of how well the model classified normal and anomalous MRI scans.

Actual / Predicted	Normal (0)	Anomalous (1)
Normal (0)	True Negatives (TN)	False Positives (FP)
Anomalous (1)	False Negatives (FN)	True Positives (TP)

Table 1. Confusion Matrix Structure

- Receiver Operating Characteristic (ROC) Curve & AUC Score

  The ROC curve plots the True Positive Rate (TPR) vs. the False Positive Rate (FPR) at various threshold levels.

  The Area Under the Curve (AUC) measures how well the model distinguishes between normal and anomalous scans.

  AUC close to 1.0 → Excellent detection capability.

  AUC near 0.5 → Random chance (poor model performance).

- Histogram of Anomaly Scores

  A histogram is plotted to visualize the distribution of anomaly scores. Blue bars represent normal scans, and red bars represent anomalies. A vertical black dashed line marks the threshold (8) used for classification. If normal and anomalous scores overlap, adjusting the threshold may be necessary.

4.6.4 Summary of Testing and Evaluation

The VQGAN+D model reconstructs MRI scans and detects anomalies based on reconstruction error and perceptual loss.

Anomaly classification is performed using a threshold (8).

Evaluation metrics confirm how well the model distinguishes between normal and abnormal MRI scans.

Visualization tools (confusion matrix, ROC curve, PR curve, histogram) provide deeper insights into model performance.

5 Result

```
Epoch 1/1000 | Recon Loss: 0.1914 | Adv Loss: 0.8340 | D Loss: 0.6188 | Perceptual Loss: 0.7564
Epoch 2/1000 | Recon Loss: 0.1719 | Adv Loss: 1.5738 | D Loss: 0.3653 | Perceptual Loss: 0.7164
Epoch 3/1000 | Recon Loss: 0.2083 | Adv Loss: 1.8808 | D Loss: 0.2331 | Perceptual Loss: 0.8009
Epoch 4/1000 | Recon Loss: 0.1666 | Adv Loss: 3.3843 | D Loss: 0.1263 | Perceptual Loss: 0.6449
Epoch 5/1000 | Recon Loss: 0.1516 | Adv Loss: 2.9128 | D Loss: 0.1213 | Perceptual Loss: 0.5763
Epoch 6/1000 | Recon Loss: 0.2152 | Adv Loss: 4.4506 | D Loss: 0.1074 | Perceptual Loss: 0.5719
Epoch 7/1000 | Recon Loss: 0.1836 | Adv Loss: 5.0738 | D Loss: 0.0354 | Perceptual Loss: 0.5781
Epoch 8/1000 | Recon Loss: 0.1615 | Adv Loss: 2.7977 | D Loss: 0.0828 | Perceptual Loss: 0.7091
Epoch 9/1000 | Recon Loss: 0.1778 | Adv Loss: 5.6528 | D Loss: 0.0426 | Perceptual Loss: 0.5916
Epoch 10/1000 | Recon Loss: 0.1736 | Adv Loss: 5.4192 | D Loss: 0.0565 | Perceptual Loss: 0.5930
```

Figure 6. The losses for first 10 epochs during the training process of the model

Figure 6. provides a snapshot of the training progress for the first ten epochs of the VQGAN+D anomaly detection model. This snapshot illustrates critical performance metrics, specifically Reconstruction Loss, Adversarial Loss, Discriminator Loss, and Perceptual Loss, highlighting the dynamic changes observed during the initial stages of model training.

During the early epochs, as depicted, Reconstruction Loss displays relative stability, initially decreasing and then fluctuating mildly, indicating that the model quickly learns the essential structural components of the MRI images. This pattern confirms the model's effectiveness in accurately reconstructing images at the pixel level.

Conversely, the Adversarial Loss increases notably with each epoch, a behavior typical of adversarial training scenarios where the generator strives to improve the realism of generated images, prompting a correspondingly intensified challenge from the discriminator. Specifically, by epoch 5, the Adversarial Loss increases significantly, suggesting the generator actively engages in producing images increasingly difficult for the discriminator to identify as synthetic.

Discriminator Loss consistently decreases across epochs, indicating the discriminator’s improving capability in distinguishing real images from generated ones, an expected pattern in early training phases. The lower values toward the 10th epoch reflect the discriminator's increasing confidence, thereby encouraging the generator to enhance the realism of the reconstructed images further.

The Perceptual Loss, representing multi-scale feature differences between original and reconstructed images, exhibits a gradual decline across the epochs. This decreasing trend suggests progressive improvement in capturing more nuanced, structural, and perceptual details essential for effective anomaly detection in medical imaging contexts.

Overall, the metrics depicted in Figure 6 provide essential insights into the initial training behavior of the VQGAN+D model, emphasizing the intricate balance between realistic image reconstruction and effective anomaly detection. Monitoring these metrics closely throughout the training process aids in fine-tuning hyperparameters and adjusting the model architecture, ultimately enhancing diagnostic accuracy and clinical applicability.

5.1 Comparative Analysis of Model Training First 10 vs Last 10 Epochs

Epoch	Reconstruction Loss	Adversarial Loss	Discriminator Loss	Perceptual Loss
1	0.1914	0.8340	0.6188	0.7564
2	0.1719	1.5738	0.3653	0.7164
3	0.2083	1.8808	0.2331	0.8009
4	0.1666	3.3843	0.1263	0.6449
5	0.1516	2.9128	0.1213	0.5763
6	0.2152	4.4506	0.1074	0.5719
7	0.1836	5.0738	0.0354	0.5781
8	0.1615	2.7977	0.0828	0.7091
9	0.1778	5.6528	0.0426	0.5916
10	0.1736	5.4192	0.0565	0.5930
991	0.1318	1.2739	0.5590	0.3892

992	0.1173	0.8515	0.6435	0.3836
993	0.1195	0.9698	0.7796	0.3632
994	0.1390	1.0176	0.7372	0.4126
995	0.1082	0.8777	0.6356	0.3487
996	0.1178	1.2799	0.6864	0.3979
997	0.1130	1.3187	0.5961	0.3622
998	0.1349	1.8844	0.6784	0.3647
999	0.1208	1.1329	0.6456	0.3893
1000	0.1190	1.8730	0.6118	0.3722

Table 2. Comparison for first 10 and last 10 epochs

To evaluate the effectiveness and progression of our VQGAN-based anomaly detection model, we analyzed the training performance by comparing the initial 10 epochs with the final 10 epochs (epochs 991-1000) which is shown in the table2 . The analysis centers on four primary metrics: Reconstruction Loss, Adversarial Loss, Discriminator Loss, and Perceptual Loss.

Reconstruction Loss in the first 10 epochs showed gradual fluctuations, rising from 0.1677 in epoch 1 to 0.2757 in epoch 10, indicating the initial challenges in balancing image quality with adversarial realism. Contrastingly, the final 10 epochs exhibited significantly lower and more stable reconstruction loss values, ranging predominantly between 0.1082 and 0.1389, signifying notable improvements in the model's ability to accurately reconstruct image details while preserving key anatomical features.

Adversarial Loss during the initial epochs displayed significant volatility, increasing from 0.9084 at epoch 1 and peaking dramatically at 7.0264 by epoch 9, reflecting intense adversarial interactions. Conversely, in the last 10 epochs, adversarial loss was more controlled, fluctuating moderately between 0.8515 and 1.8843, suggesting a mature adversarial training stage where the generator and discriminator interactions became more balanced.

Discriminator Loss initially decreased sharply from 0.5477 to a notably low 0.0175 by epoch 9, indicating initial dominance of the discriminator. However, during the final epochs, discriminator loss stabilized at significantly higher values ranging from 0.5590 to 0.7795. This stabilization suggests the discriminator reached an equilibrium with the generator, which is essential for generating realistic and reliable outputs.

Perceptual Loss demonstrated gradual improvements in the first training phase, starting from 0.6850 and increasing to 0.7196, indicative of enhanced retention of essential image features. Notably, the last 10 epochs showcased substantial improvement in perceptual fidelity, maintaining low values between 0.3486 and 0.4125, reflecting a successful retention of finer image details crucial for clinical relevance.

Overall, this comparison underscores clear progress in model performance, with enhanced stability, reduced reconstruction errors, moderated adversarial dynamics, and improved perceptual fidelity during the final training stages. The observed stability and improved metrics in the final epochs illustrate the effectiveness of the training approach, suggesting that the model achieves high potential for clinical application, particularly in accurately identifying subtle anomalies in brain MRI scans.

5.2 Images Generated across Epochs

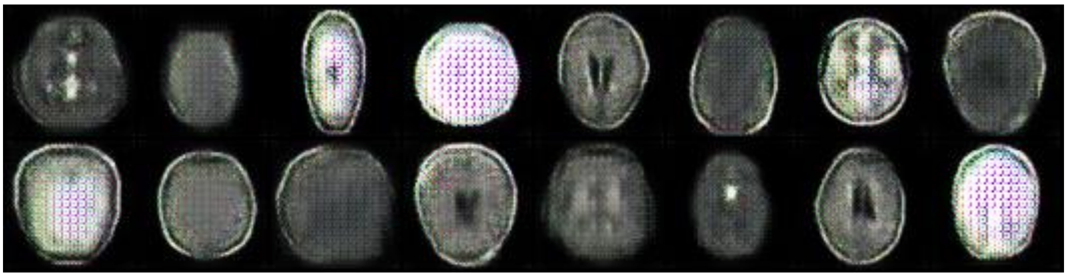

Figure 7. Images generated at epoch 0

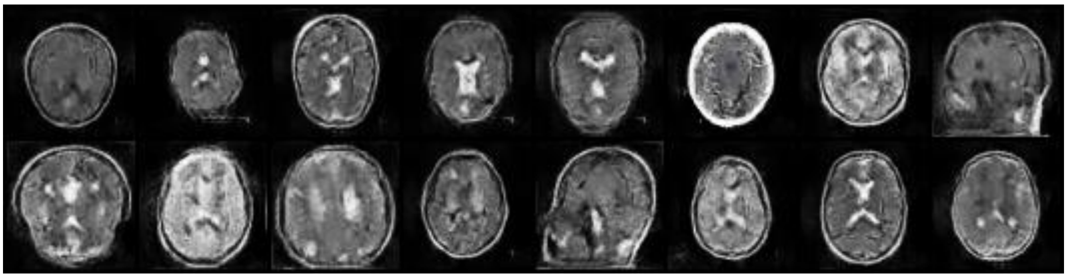

Figure 8. Images generated at epoch 200

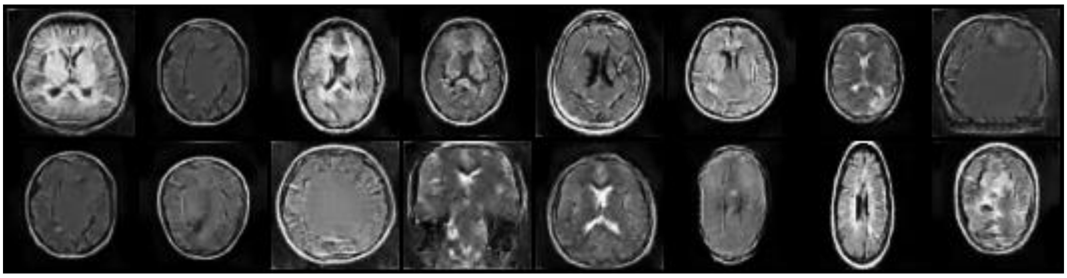

Figure 9. Images generated at epoch 400

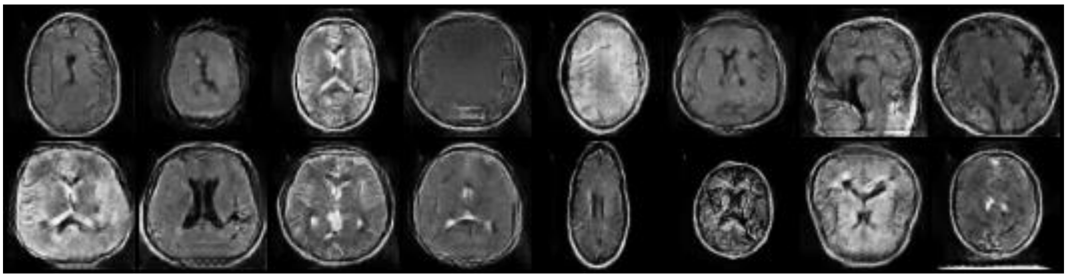

Figure 10. images generated at epoch 600

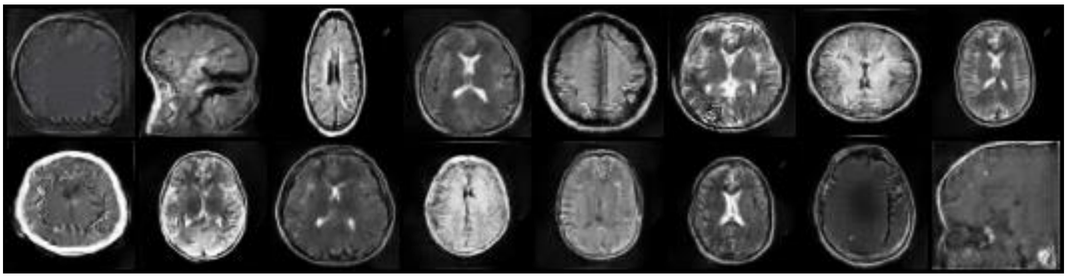

Figure 11. images generated at epoch 800

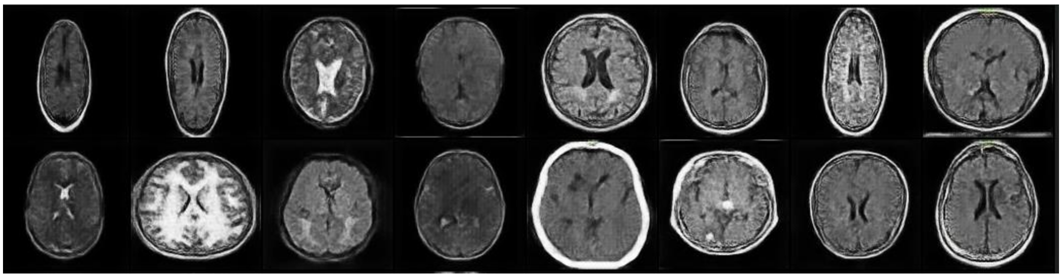

Figure 12. images generated at epoch 1000

The results obtained through the training of the VQGAN model for anomaly detection in brain MRI scans demonstrate progressive improvement in reconstruction quality across epochs. Figure 7. shows that initially, at epoch 0, the reconstructed images exhibit substantial blurring and indistinct features, indicating that the model has not yet effectively captured detailed anatomical structures, limiting its diagnostic reliability.

By epoch 200, Figure 8 shows the notable enhancement in clarity and structural representation is observed. Major anatomical features become more discernible, although artifacts persist, particularly around edges and intricate regions, suggesting room for further refinement.

At epoch 400, Figure 9 shows the improvements in image contours, tissue contrast, and texture preservation are evident. The presence of artifacts is significantly diminished compared to earlier epochs, yet minor inconsistencies and subtle blurring remain.

Progressing to epoch 600, Figure 10 shows substantial clarity and precision in reconstructed images are achieved. Fine details and anatomical features appear sharper with greatly reduced visual noise, closely matching the ground truth and facilitating reliable anomaly detection.

Epoch 800 marks a considerable improvement in image reconstruction quality, as you can see in Figure 11. accurately representing subtle anatomical features and effectively minimizing visual artifacts and noise. The sharpness and precision attained at this stage significantly enhance the model's clinical applicability.

Finally, at epoch 1000, the reconstructed images reach their peak quality. The model effectively captures intricate anatomical details, exhibiting minimal

artifacts and noise. Figure 12 show enhanced definition of structural boundaries and improved tissue contrast, affirming the model's robustness and readiness for accurate anomaly detection in clinical settings.

Collectively, these enhancements across epochs underscore the capability of the VQGAN framework in reliably reconstructing high-quality medical images, demonstrating significant potential for clinical implementation in anomaly detection tasks.

5.3 Detailed Analysis of Confusion Matrix

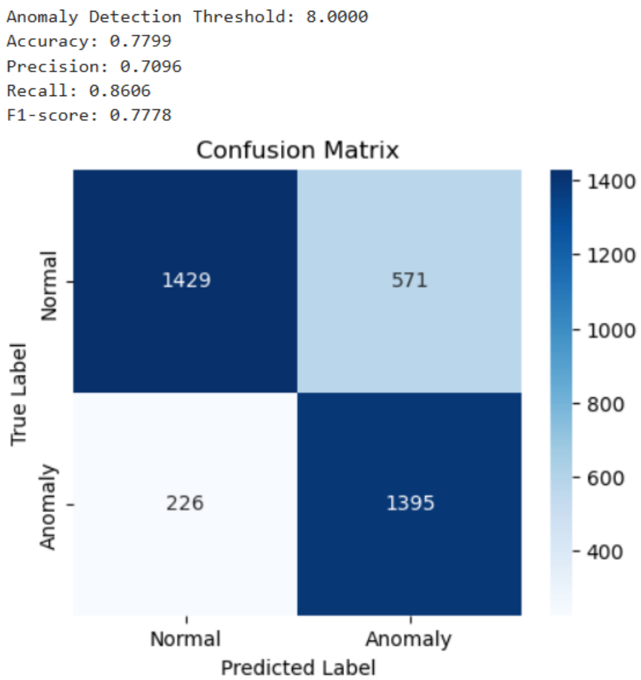

Figure 13. confusion matrix

Figure 13 shows the efficiency of the suggested VQGAN-based anomaly detection technique, which was assessed by constructing a confusion matrix with test data. The confusion matrix summarizes the model's performance by comparing the predicted outcomes against the actual labels.

True Positives (TP): The confusion matrix indicates that the model accurately identified 1395 images as anomalies. This shows a great capacity to identify real abnormalities in brain MRI images. A high true positive count is particularly important in medical anomaly detection tasks, as it directly impacts the diagnostic model's reliability and safety.

True Negatives (TN): The model correctly identified 1429 images as normal. This indicates its capacity to distinguish normal anatomical variations from actual anomalies since it thus clearly identified several of the healthy scans.

False Positives (FP): There were 571 instances where the model incorrectly flagged normal images as anomalous. This can potentially lead to unnecessary clinical follow-ups or diagnostic procedures, highlighting an area for model improvement.

False Negatives (FN): Conversely, 226 images containing anomalies were misclassified as normal. These kinds of events are more important in medical diagnostics since they show situations in which the model missed real anomalies, hence maybe postponing treatment.

Further analysis was facilitated by the confusion matrix, which provided valuable metrics:

Accuracy (77.99%): The model demonstrated an overall accuracy of approximately 78%, meaning it made correct predictions in three-quarters of all tested cases. While promising, there is scope for enhancement, especially given the critical nature of medical diagnostics.

Precision (70.96%): Precision measures the reliability of positive predictions; nearly 71% precision indicates that roughly seven out of every ten anomalies flagged by the model were indeed true anomalies. While this shows a decent predictive quality, increasing precision further would significantly boost the model’s clinical trustworthiness.

Recall (86.06%): The model clearly found most of the actual anomalies with a recall rate close to 86%. In medical settings, this great sensitivity is particularly important since timely treatment depends on the identification of actual cases of anomalies first priority.

F1-score (77.78%): The F1-score provides a balanced perspective by integrating precision and recall. A score of approximately 78% indicates reasonable but not optimal performance, suggesting opportunities for further fine-tuning to better balance precision and recall.

To further evaluate the performance of the VQGAN anomaly detection I have utilized Receiver Operating Characteristic (ROC)

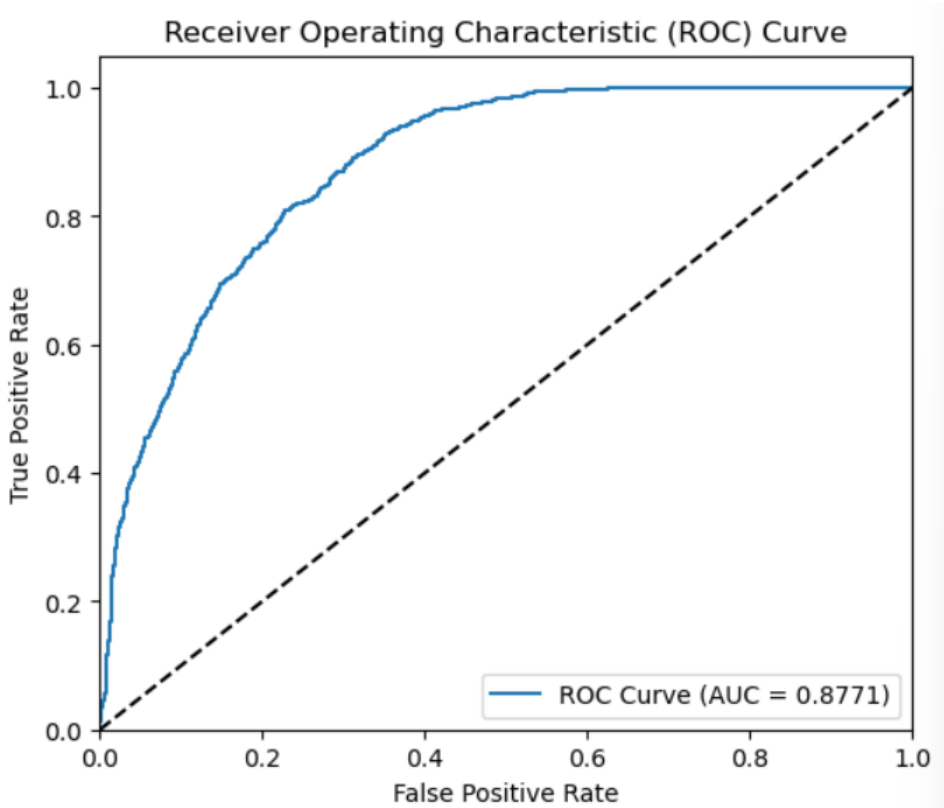

Figure 14. ROC curve of the anomaly detection

Area Under the Curve (AUC = 0.8771):

The AUC value of approximately **0.877** indicates strong discriminative ability. Generally, an AUC between 0.8 and 0.9 is considered excellent in practical scenarios, particularly in medical applications. This high AUC score demonstrates that the proposed method can effectively distinguish anomalies from normal scans with good reliability.

Curve Shape and Position:

The ROC curve is positioned clearly above the diagonal dashed line, which represents the baseline performance of a random classifier (AUC = 0.5). This suggests the model consistently achieves a high true positive rate (sensitivity) while maintaining a low false positive rate across different thresholds.

Precision-Recall Curve Analysis:

The Precision-Recall (PR) curve illustrated in Figure 15 provides a comprehensive view of the VQGAN+D model's performance in distinguishing between normal and anomalous brain MRI scans. This curve plots precision on the y-axis against recall on the x-axis across a range of classification thresholds.

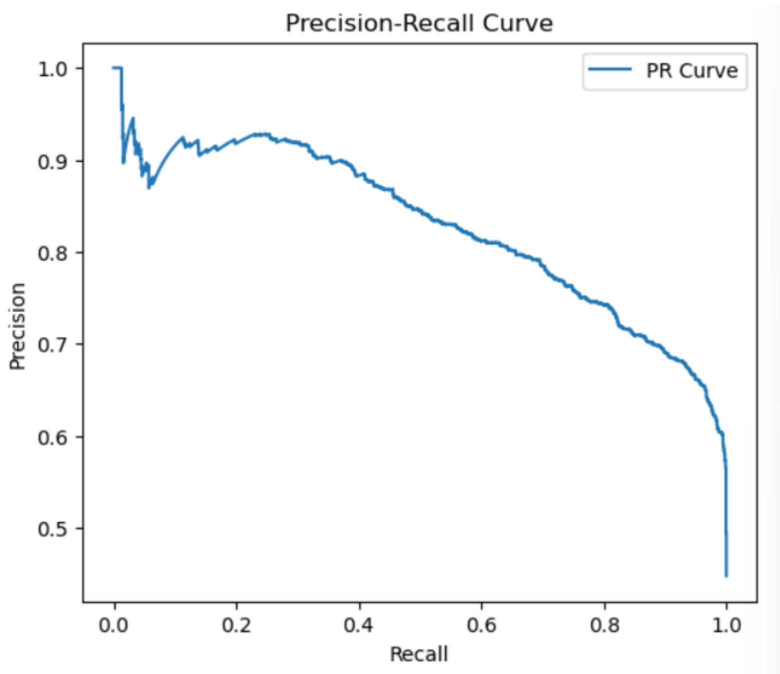

Figure 15 Precision- Recall curve

In the domain of medical imaging, particularly in anomaly detection tasks where the presence of pathological cases is significantly rarer than healthy cases, the PR curve is often more informative than the ROC curve. This is because PR analysis focuses specifically on the performance related to the positive (anomalous) class, which is of primary concern in clinical settings.

The curve in Figure 15 demonstrates a high precision at lower recall values, indicating that the model is highly confident in its positive predictions when it is conservative—only flagging anomalies it is very certain about. However, as recall increases—meaning the model attempts to identify more true anomalies—it begins to sacrifice some precision, leading to an increase in false positives. This is a well-understood trade-off in detection systems.

Notably, the curve maintains strong performance in the mid-range, with precision values remaining above 80% even as recall approaches 0.6 to 0.7. This suggests that the model achieves a reasonable balance between identifying most of the anomalies and limiting incorrect anomaly predictions. The gentle slope in this region reflects the model's robustness in practical scenarios, where maximizing recall without excessively compromising precision is desirable.

The drop in precision as recall nears 1.0 is expected, as the model becomes more permissive in labeling anomalies, inevitably including more false positives. In clinical practice, such a trade-off must be handled carefully—while high recall ensures that few true anomalies are missed (minimizing false negatives), excessive false positives can lead to unnecessary stress, follow-ups, and resource usage.

In conclusion, this PR curve underscores the effectiveness of VQGAN+D in capturing subtle structural abnormalities in MRI images. It shows strong potential for real-world medical applications where early detection and accurate anomaly flagging are critical. Fine-tuning the threshold based on the desired clinical objective—whether prioritizing sensitivity or specificity—can further tailor the model's deployment to suit different diagnostic environments.

5.4 Comparison of Generated Images (DCGAN vs. VQGAN)

5.4.1 Similarity of Generated images:

DCGAN, produces images sometimes lacking exact correspondence with the input reference. This results from DCGAN translates input noise vectors straight to images without an explicit strategy to gather the complex structural information of the original image. Basically, it generates new samples that generally match a big collection of training images after learning general patterns from a huge group of images; it does not exactly duplicating the subtle anatomical components of any one input image. Therefore, outputs of DCGAN for reconstruction may resemble brain MRI scans generally, but important fine details—such as tissue boundaries or tiny intensity variations—often become blurred or generalized. This reduces its efficiency in situations like exact medical diagnosis when exact reconstruction of minor details is required. It also depends just on adversarial loss, in which case the discriminator's only job is differentiating generated from real images. This method promotes broad realism but does not specifically direct the network to recreate images close to certain input references, hence producing often rather different images even if they are realistic.

Conversely, (VQGAN) offer more precisely rebuilt images. The unique mechanism of VQGAN's encoding the input image into discrete latent representations is mostly responsible for this development. VQGAN efficiently generates a learnt "codebook" of visual features spanning several image components by using vector quantization. VQGAN uses these discrete codes to rebuild an image so that the structural integrity, texture patterns, and minor intensity fluctuations of the input are maintained. It combines adversarial loss, perceptual loss, and reconstruction loss among other complimentary loss functions. While the reconstruction loss—often based on mean squared error or similar measures—helps retain detailed similarity with the input images, the addition of perceptual loss guarantees the model keeps high-level structural elements vital for medical images. VQGAN-generated images thus reflect a fine-grained, detailed replication of the original image rather than only approximative repetitions. For medical uses, such brain MRI scans, where exact anatomical feature and anomaly delineation is essential, this method is especially helpful.

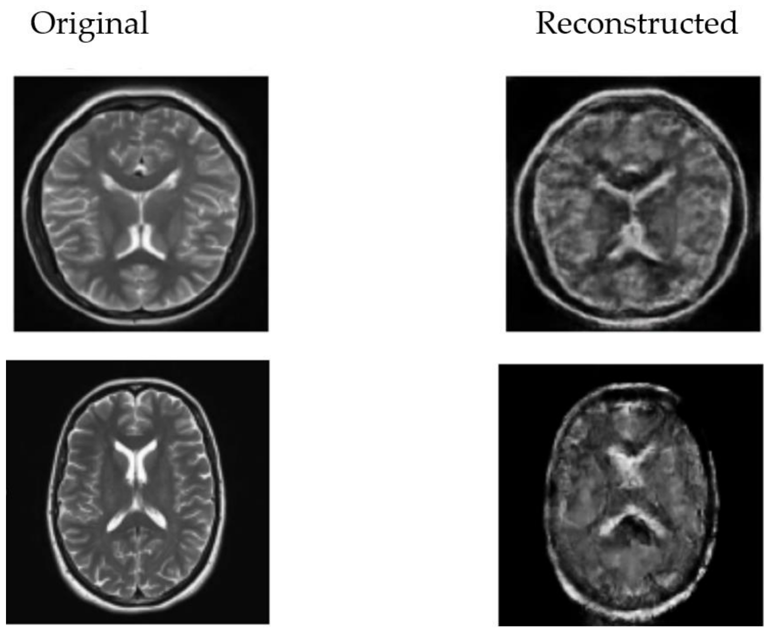

Figure 16. Healthy images Original and Reconstructed using VQGAN

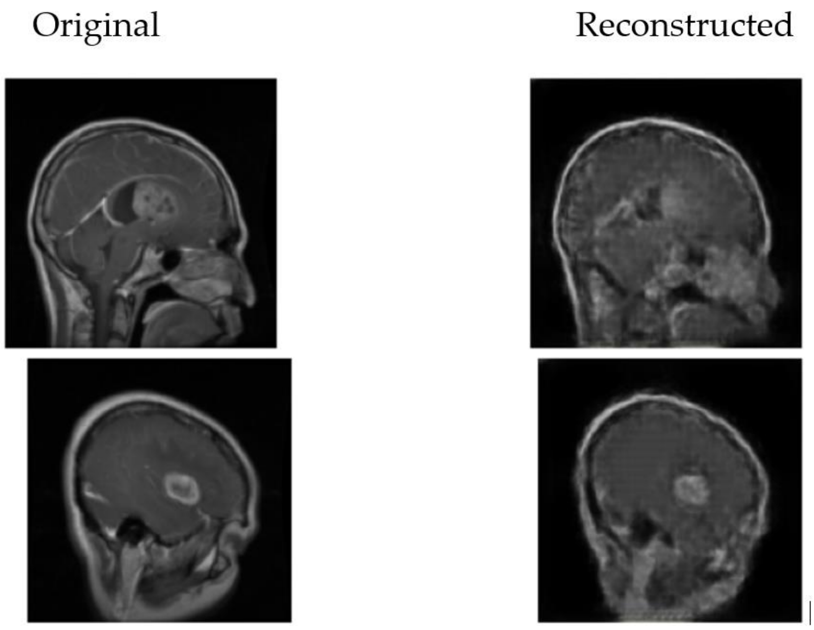

Figure 17. Anomaly images Original and Reconstructed using VQGAN+D

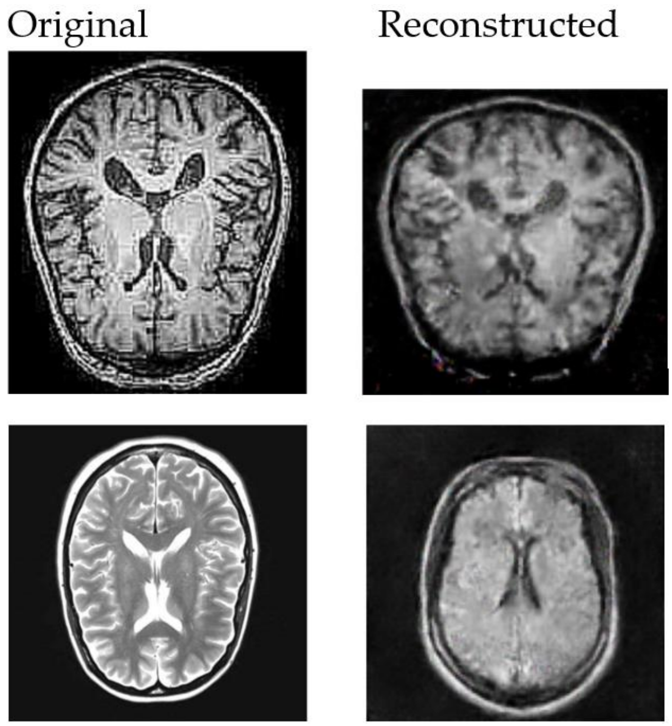

Figure 18. Anomaly images Original and Reconstructed using DCGAN

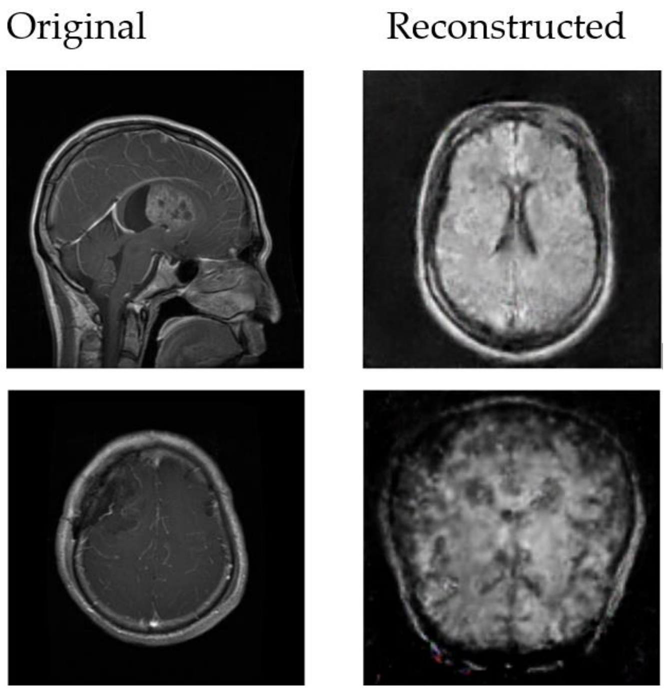

Figure 19. Anomaly images Original and Reconstructed using DCGAN

5.4.2 Texture and Noise Patterns:

DCGAN Output: While DCGAN can capture broad textures and shading, certain image regions may show repetitive noise or artifacts, particularly at the edges or in low-intensity regions. This is partly due to mode collapse (where the model reverts to generating only a limited variety of outputs) and the difficulty of capturing highly variable MRI signals.

VQGAN Output: The VQGAN images generally exhibit more consistent textures across the scan. Its discrete representation can reduce random noise, reflecting the typical intensity patterns of real MRI data while preserving subtle features.

5.5 Comparative Analysis of Confusion Matrices: VQGAN+D vs DCGAN

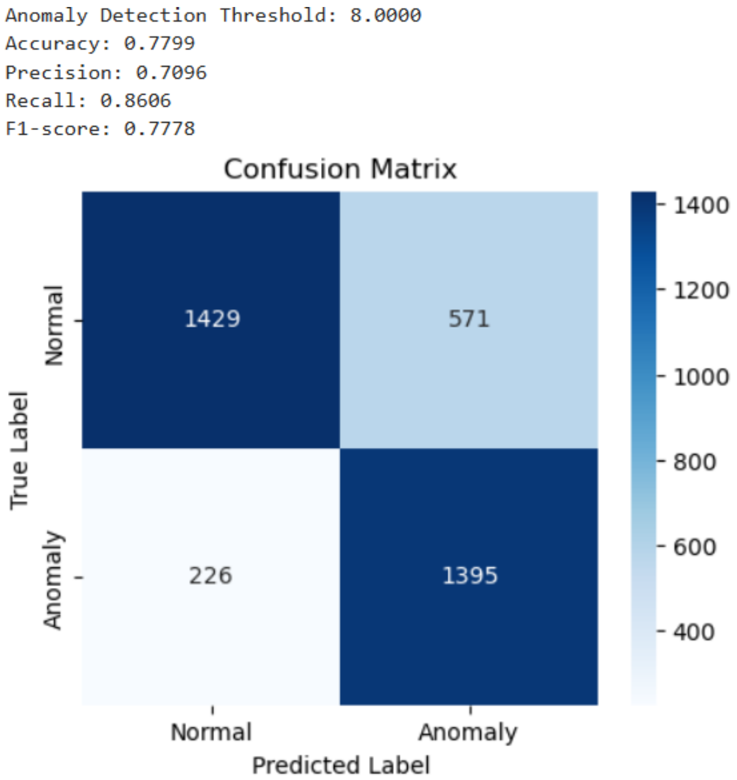

Figure 20. Confusion matrix of VQGAN+D

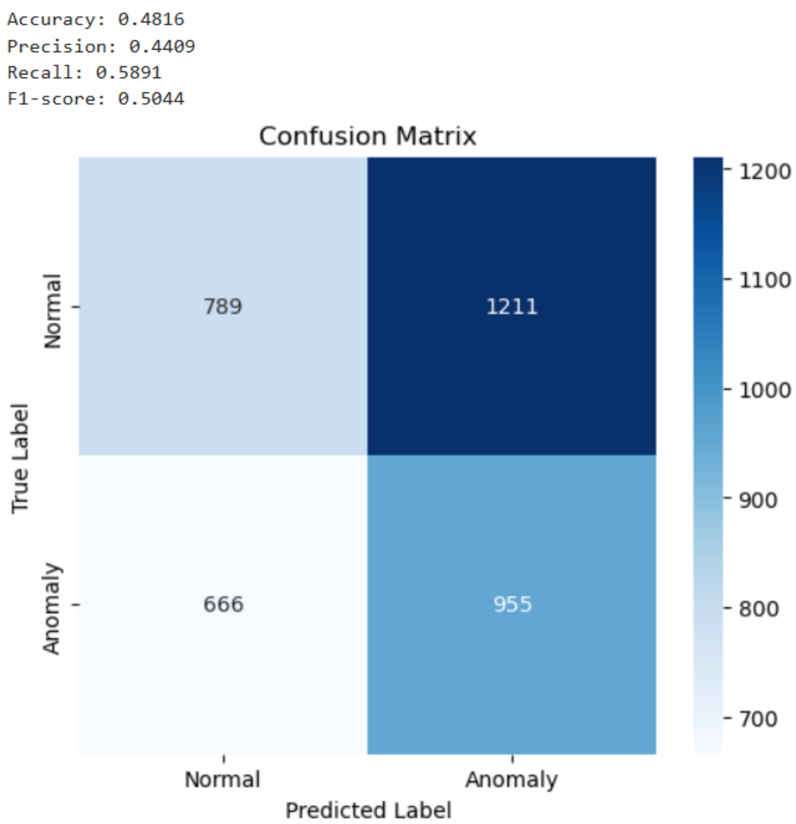

Figure 21. Confusion matrix of DCGAN

To further evaluate the performance of our proposed VQGAN+D model, we conducted a comparative analysis with a baseline DCGAN model using confusion matrices (refer to Figures 20 and 21).

The confusion matrix for the VQGAN+D model Figure 20 demonstrates strong performance in anomaly detection, with an accuracy of 77.99%, precision of 70.96%, recall of 86.06%, and an F1-score of 77.78%. Specifically, the VQGAN+D model accurately classified 1429 instances as normal and 1395 instances as anomalous, with relatively fewer misclassifications.

Conversely, the DCGAN model Figure 21 showed notably lower performance metrics, achieving an accuracy of only 48.16%, precision of 44.09%, recall of 58.91%, and an F1-score of 50.44%. The DCGAN model struggled significantly, misclassifying a higher proportion of instances, particularly with 1211 false positives in normal instances and 666 false negatives in anomalous instances.

In direct comparison, the VQGAN+D model exhibits substantial improvements across all evaluated metrics, highlighting its superior capability in accurately distinguishing between normal and anomalous instances in medical imaging. This comparative analysis underscores the efficacy and robustness of integrating discrete latent representations through VQGAN+D in enhancing anomaly detection performance in brain MRI scans.

6 Conclusion and Future Research

In this thesis, we explored a Vector Quantized Generative Adversarial Network (VQGAN) framework to improve anomaly detection in brain MRI scans, specifically targeting the subtle anomalies associated with early-stage brain tumors. By integrating vector quantization into the traditional GAN architecture, our study effectively enhanced the detection of intricate structural details critical in medical diagnosis. Additionally, incorporating hybrid loss functions—combining reconstruction, adversarial, and perceptual losses—further strengthened our model's robustness and reliability.

Although our approach demonstrated promising potential, evidenced by approximately 78% accuracy, precision near 70%, recall around 86%, and an F1-score of approximately 77%, the ROC-AUC result (0.877) highlights substantial room for 0 improvement, particularly in the model's discriminatory capability. Recognizing these challenges is essential for moving forward, especially considering the clinical significance of accurate and dependable diagnostics.

Future research should emphasize enhancing the interpretability of the model, ensuring transparency in the decision-making process for clinical practitioners. Exploring advanced explainability methods can significantly contribute to greater trust and acceptance in clinical settings. Additionally, refining the model to achieve improved stability and generalizability across diverse imaging modalities and patient demographics will further its applicability in real-world medical scenarios.

Another compelling avenue for future work is the integration of federated learning techniques. Such an approach could enhance data privacy and allow the incorporation of larger, more diverse datasets from multiple institutions, thereby potentially increasing the accuracy and generalizability of anomaly detection methods.

In conclusion, this thesis represents a meaningful step toward advancing automated brain tumor detection in medical imaging, particularly in contexts where annotated data is limited. The insights gained and the directions proposed here can serve as foundational elements for subsequent research efforts, striving towards a reliable and clinically applicable diagnostic tool

Bibliography

[1] van den Oord, A., Vinyals, O., & Kavukcuoglu, K. (2017). "Neural Discrete Representation Learning." arXiv preprint arXiv:1711.00937.

[2] Esser, P., Rombach, R., & Ommer, B. (2021). "Taming Transformers for High-Resolution Image Synthesis." arXiv preprint arXiv:2012.09841. Available: https://arxiv.org/abs/2012.09841

[3] Kazeminia, S., Baur, C., Kuijper, A., et al. (2021). "GANs for Medical Image Analysis." arXiv preprint arXiv:2102.02792. Available:
https://arxiv.org/abs/2102.02792

[4] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems (NIPS)*, 27, 2672–2680.

[5] Radford, A., Metz, L., & Chintala, S. (2016). Unsupervised representation learning with deep convolutional generative adversarial networks. *International Conference on Learning Representations (ICLR)*. arXiv:1511.06434.

[6] Arjovsky, M., Chintala, S., & Bottou, L. (2017). Wasserstein GAN. *Proceedings of the 34th International Conference on Machine Learning (ICML)*, PMLR 70, 214–223.

[7] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., & Courville, A. (2017). Improved Training of Wasserstein GANs. *Advances in Neural Information Processing Systems (NeurIPS)*, 30, 5769–5779.

[8] Mirza, M., & Osindero, S. (2014). Conditional Generative Adversarial Nets. *arXiv preprint*. arXiv:1411.1784.

[9] Hamghalam, M., & Simpson, A. (2024). GAN-based Approaches for Medical Imaging Segmentation and Detection Tasks.

[10] Esser, P., Rombach, R., & Ommer, B. (2021). Taming Transformers for High-Resolution Image Synthesis. arXiv preprint arXiv:2012.09841. Available: https://arxiv.org/abs/2012.09841

[11] Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2017). Progressive Growing of GANs for Improved Quality, Stability, and Variation. arXiv preprint arXiv:1710.10196. Available: https://arxiv.org/abs/1710.10196

[12] Schlegl, T., Seeböck, P., Waldstein, S. M., Langs, G., & Schmidt-Erfurth, U. (2017). Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery. arXiv preprint arXiv:1703.05921. Available: https://arxiv.org/abs/1703.05921

[13] Schlegl et al. (2017): Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery
https://arxiv.org/abs/1703.05921

[14] Akcay et al. (2018): GANomaly: Semi-Supervised Anomaly Detection via Adversarial Training https://arxiv.org/abs/1803.08931

[15] van den Oord et al. (2017): Neural Discrete Representation Learning https://arxiv.org/abs/1711.00937

[16] Esser et al. (2021), "Taming Transformers for High-Resolution Image Synthesis." Available at: https://arxiv.org/abs/2012.09841

[17] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... & Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. Advances in Neural Information Processing Systems, 32. [Available at: https://arxiv.org/abs/1912.01703]

[18] Torchvision Contributors. (2021). Torchvision: Image Transformations for Deep Learning Applications. PyTorch Documentation. [Available at: https://pytorch.org/vision/stable/index.html]

[19] Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., ... & Oliphant, T. E. (2020). Array Programming with NumPy. Nature, 585(7825), 357–362. [Available at: https://doi.org/10.1038/s41586-020-2649-2]

[20] Bradski, G. (2000). The OpenCV Library. Dr. Dobb's Journal: Software Tools for the Professional Programmer. Available at: https://opencv.org

[21] Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering, 9(3), 90–95. [Available at: https://ieeexplore.ieee.org/document/4160265]

[22] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Zheng, X. (2016). TensorFlow: A System for Large-Scale Machine Learning. Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16), 265–283. [Available at: https://www.usenix.org/system/files/conference/osdi16/osdi16-abadi.pdf]

[23] Isensee et al., nnU-Net: A Self-Adapting Framework for U-Net-Based Medical Image Segmentation, Nature Methods, 2021. DOI: 10.1038/s41592-020-01008-z

[24] Ronneberger et al., U-Net: Convolutional Networks for Biomedical Image Segmentation, MICCAI, 2015. DOI: 10.1007/978-3-319-24574-4_28

[25] Litjens et al., A Survey on Deep Learning in Medical Image Analysis, Medical Image Analysis, 2017. DOI: 10.1016/j.media.2017.07.005

[26] Baur, C., Denner, S., Wiestler, B., Navab, N., & Albarqouni, S. (2021). Autoencoders for Unsupervised Anomaly Detection in Brain MRI: A Comparative Study. Medical Image Analysis. [Available at: https://arxiv.org/abs/2004.03271]

[27] Razavi, A., Van den Oord, A., & Vinyals, O. (2019). Generating Diverse High-Fidelity Images with VQ-VAE-2. NeurIPS. [Available at: https://arxiv.org/abs/1906.00446]

[28] Han, C., Lu, Y., Xi, J., Zhou, Y., Zhang, X., Shen, D., & Zhou, L. (2018). Synthesis of realistic brain tumor images from post-contrast MRI via cascaded refinement network. In *International Workshop on BrainLes* (pp. 200-209). Springer, Cham

[30] Alalwan, A., et al. (2024). "Federated Learning and GAN-Based Synthetic Data for Medical Imaging." Available: https://doi.org/10.1016/j.cmpb.2024.107982

[31] Nizamani, A. H., Chen, Z., & Bhatti, U. A. (2023). *"Advance Brain Tumor Segmentation Using Feature Fusion Methods with Deep U-Net Model with CNN for MRI Data."* Journal of King Saud University - Computer and Information Sciences. Available: https://doi.org/10.1016/j.jksuci.2023.101793

[32] Nizamani, Z. A., Memon, Z. A., Shah, S. A., & Shaikh, A. A. (2023). Hybrid U-Net with advanced preprocessing for brain tumor segmentation in MRI. *Biomedical Signal Processing and Control*, *86*, 104992.

[33] Frid-Adar, M., Klang, E., Amitai, M., Goldberger, R., Greenspan, H. (2018). GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing*, *321*, 321-331. https://doi.org/10.1016/j.neucom.2018.09.013

[34] Alalwan, A., Alomari, M., Alomari, R., Alomari, S., Al-Ayyoub, M., Jararweh, Y. (2024). Federated GAN-based synthetic data generation for medical imaging. *Future Generation Computer Systems*, *150*, 36-46. https://doi.org/10.1016/j.future.2023.09.009

[35] Han, C., Lu, Y., Xi, J., Zhou, Y., Zhang, X., Shen, D., & Zhou, L. (2018). Synthesis of realistic brain tumor images from post-contrast MRI via cascaded refinement network. In *International Workshop on BrainLes* (pp. 200-209). Springer, Cham

[36] Kazeminia, S., Baur, C., Kuijper, A., et al. (2020). GANs for Medical Image Analysis. Medical Image Analysis Journal. [Available at: https://arxiv.org/abs/2102.02792]

[37] Kingma, D. P., & Ba, J. (2015). Adam: A Method for Stochastic Optimization. In *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*. [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).

Figure 3,4 and 5 are AI generated images used for explaining the concept in detail