Advanced Noise Reduction and Feature Enhancement in Medical CT Brain Imaging Using Deep Learning

by

Sahriar Wahid Galib
20301337
Mashrur Safir Shabab
20241037
Sabrina Rahman Mazumder
20301218
Shouvik Banerjee Argha
20301118
Zawadul Kafi Nahee
20301057

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

Department of Computer Science and Engineering
Brac University
June 2024

Declaration

It is hereby declared that

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

Student's Full Name & Signature:

Sahriar Wahid Galib
20301337

Mashrur Safir Shabab
20241037

Sabrina Rahman Mazumder
20301218

Shouvik Banerjee Argha
20301118

Zawadul Kafi Nahee
20301057

Approval

The thesis/project titled "Advanced Noise Reduction and Feature Enhancement in Medical CT Brain Imaging Using Deep Learning" submitted by

1. Sahriar Wahid Galib(20301337)
2. Mashrur Safir Shabab(20241037)
3. Sabrina Rahman Mazumder(20301218)
4. Shouvik Banerjee Argha(20301118)
5. Zawadul Kafi Nahee(20301057)

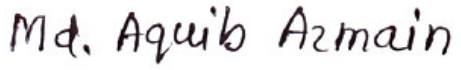
Of Spring, 2025 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on June 20, 2025.

Examining Committee:

Supervisor:
(Member)

Dr. Md. Golam Rabiul Alam
Professor
Department of Computer Science and Engineering
Brac University

Co-Supervisor:
(Member)



Md. Aquib Azmain
Lecturer
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Dr. Sadia Hamid Kazi
Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Abstract

Computed Tomography (CT) is a powerful diagnostic tool used in modern medicine, which helps a doctor provide accurate diagnostic reports and plan the individual's treatment. However, doing multiple CT scans on one individual poses health risks like cellular damage, deoxyribonucleic acid (DNA) mutation, tissue damage and lastly development of cancer over time. As a result, concerns regarding radiation exposure have led to the development of low-dose Computed Tomography (LDCT) techniques. However, LDCT suffers from low image quality with increased noise, causing the diagnostic to be less accurate. In this study, we provide a novel approach to enhance these LDCT images. Our work is leveraged upon deep learning techniques involving Generative Adversarial Networks (GAN). Thus, through rigorous experimentation on a diverse set of clinical CT images, we found the best model for enhancing LDCT images which holds details of anatomical structure and at the same time minimizes noise. This causes more accurate diagnostics, reducing the health hazard of patients. Thus, this approach paves the way for more accurate diagnostics while minimizing radiation exposure.

Keywords: ESRGAN (Enhanced Super-Resolution GAN), RRDB (Residual-in-Residual Dense Block), Self-Attention, Vision Transformer, PatchGAN Discriminator, Spectral Normalization, SSIM (Structural Similarity Index), VGG Perceptual Loss, Deep Learning

Table of Contents

Declaration i

Approval ii

Abstract iv

Table of Contents v

List of Figures vii

Nomenclature ix

1 Introduction 1
- 1.1 Background Studies . . . 1
- 1.2 Problem Statement . . . 2
- 1.3 Research Objective . . . 3
- 1.4 Report Organization . . . 4

2 Literature Review 5

3 Workplan 10

4 Dataset 13
- 4.1 Data Selection . . . 13
- 4.2 Data Collection . . . 14
- 4.3 Data Organization . . . 15
- 4.4 Data Visualization . . . 15
- 4.5 Data Preprocessing . . . 17
  - 4.5.1 Grey-scale Conversion . . . 17
  - 4.5.2 Cropping . . . 17

5 Methodology 19
- 5.1 Architectural Diagram . . . 19
- 5.2 Noise Encoding for Synthetic Noise Generation . . . 19
  - 5.2.1 Noise Extraction . . . 20
  - 5.2.2 Noise Encoding . . . 21
  - 5.2.3 Encoder . . . 22
  - 5.2.4 Latent Space . . . 23
  - 5.2.5 Decoder . . . 24
  - 5.2.6 Full Pipeline for Noise Encoding and LDCT Generation . . . 26

5.3 Model Architecture . . . 26
5.3.1 Generator Network . . . 27
5.3.2 Discriminator Network . . . 30
5.4 Loss Functions . . . 31
5.4.1 Content Loss . . . 32
5.4.2 SSIM Loss . . . 32
5.4.3 Adversarial Loss . . . 32
5.4.4 Perceptual Loss . . . 33
5.5 Training Strategy . . . 34
5.6 Semi-Supervised Labeling . . . 34
5.7 DenseNet-based Classification . . . 36
5.7.1 Label Encoding . . . 39
5.7.2 One-Hot Encoding . . . 39
5.7.3 Train-Test Splitting . . . 39

6 Experiments and Results **41**
6.1 Evaluation Metrics . . . 41
6.1.1 Peak Signal-to-Noise Ratio (PSNR) . . . 41
6.1.2 Structural Similarity Index Measure (SSIM) . . . 41
6.1.3 Signal-to-Noise Ratio (SNR) . . . 42
6.1.4 Sharpness . . . 42
6.1.5 Contrast . . . 42
6.2 Visual Results . . . 43
6.3 Quantitative Results . . . 44
6.4 Classification Results . . . 45
6.4.1 Classification with Original Dataset . . . 46
6.4.2 Classification with Hybrid Model Generated Dataset . . . 47
6.4.3 Comparison and Analysis . . . 49

7 Conclusion **51**
7.1 Contributions . . . 51
7.2 Limitations . . . 53
7.3 Future Work . . . 53
7.4 Social Impact . . . 55

Bibliography **57**

List of Figures

3.1 Workplan 12

4.1 Screenshot of DIACOM Viewer 14
4.2 Class Distribution of CT Scans 15
4.3 Average Pixel Intensity Distribution 16
4.4 Noise Level Distribution Across Classes 16
4.5 Gray-scale Conversion of Dataset 17
4.6 Cropping of Dataset 18

5.1 Full Architecture 19
5.2 Noise Extraction 21
5.3 Auto-Encoder 22
5.4 Encoder 22
5.5 Noise encoded to its latent representation 23
5.6 Latent space before and after introducing additional noise 24
5.7 1D Vector Plot 24
5.8 Comparison between HDCT and LDCT 25
5.9 Comparison between HDCT and LDCT 25
5.10 Decoder 26
5.11 Full Pipeline 26
5.12 Entire Architecture 27
5.13 Generator 28
5.14 Residual Dense Block (RDB) 29
5.15 Multi-Head Self-Attention (MSA) 30
5.16 Discriminator 31
5.17 Data Distribution 35
5.18 2D Visualization of 5 Clusters 35
5.19 Data Distribution before Semi-Supervised Labeling 36
5.20 After Semi-Supervised Labeling (Clustering) 36
5.21 Architecture 0f DenseNet121 37
5.22 DenseNet Classification Pipeline 40

6.1 Visual Comparison of $wr11\_159.bmp$ Image 43
6.2 Visual Comparison of $Wr02\_01.bmp$ Image 43
6.3 Visual Comparison of $wr41\_3.png$ Image 44
6.4 Visual Comparison of $CT05BACK5.bmp$ Image 44
6.5 Classification Using Original HDCT Images 46
6.6 Precision, Recall, F1-score by Class 46
6.7 Test Accuracy and Loss 47

6.8 Confusion Matrix on Test Set 47
6.9 Classification Using Hybrid Model Generated Images 47
6.10 Precision, Recall, F1-score by Class 48
6.11 Test Accuracy and Loss 48
6.12 Confusion Matrix on Test Set 49
6.13 Classification 1 & 2 49
6.14 Test Loss and Test Accuracy Comparison 50
6.15 Number of Misclassified Images 50

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

$ADAM$ Adaptive Moment Estimation

AE Autoencoder

BCE Binary Cross Entropy

CNN Convolutional Neural Network

CT Computed Tomography

$ESRGAN$ Enhanced Super-Resolution Generative Adversarial Network

GAN Generative Adversarial Network

$HDCT$ High-Dose Computed Tomography

$LDCT$ Low-Dose Computed Tomography

MOS Mean Opinion Score

MRI Magnetic Resonance Imaging

MSE Mean Squared Error

NNC Neural Network Classifier

PII Personally Identifiable Information

$ReLU$ Rectified Linear Unit

RNN Recurrent Neural Network

SGD Stochastic Gradient Descent

SNN Spike Neural Network

$SSIM$ Structural Similarity Index Measure

SVM Support Vector Machine

VGG Visual Geometry Group

ViT Vision Transformer

Chapter 1

Introduction

1.1 Background Studies

In the field of advanced medical imaging, there have been evolutionary breakthroughs that later on went through radical changes. Computed Tomography (CT) scanning has gone through significant changes from 1990 to 2024. MRI and X-ray are the two kinds of techniques for medical imaging other than CT scans. Each method has its benefits and downsides. Computed Tomography(CT) tends to be chosen for specific cross-sectional images of the body to provide every particular detail for examining bones and large descriptive images that are clearer than X-rays. These CT scans give the doctors smooth illustrations, which can help doctors correctly diagnose and treat different kinds of diseases, including brain hemorrhages, tumors, and accidents. These scans are beneficial when examining soft internal tissues of the body, bones, or organs with great details to provide an immediate diagnosis and are beneficial for identifying malignancies and also for recommending operations such as biopsies and surgeries. This is revolutionary because it allows for the detection of problems at an early stage, while MRI can be more expensive and time-consuming. On the other hand, X-rays do not provide that many details compared to CT scans.

CT scans are of multiple types. Low-dose CT scans use lower radiation exposure, while high-dose CT scans have more radiation exposure. High doses of radiation are harmful to the patient's body and health. Risks with many complications arise from the higher-dose exposure to radiation. Therefore, limiting radiation exposure is necessary to minimize the danger of cancer and other health issues caused by radiation. So LDCT scans have become preferable for patients. Brenner and Hall (2007) point out that cumulative radiation exposure from medical imaging is a rising problem, particularly for patients receiving many scans [1]. Furthermore, to minimize radiation doses and keep the picture quality up to mark, LDCT makes use of cutting-edge methods using sophisticated algorithms and technologies that preserve highly qualitative image quality. In addition to highlighting the trial of National Lung Screening [2], illustrated the value of LDCT, which lowers death rates by 20% compared to chest X-rays.

When it comes to the reduction of radiation exposure as well as bringing down prolonged health hazards for patients by integrating CT Scan devices with Low Dose CT Image (LDCT) capturing functionality and GAN-based deep learning and state-

of-the-art image processing techniques for denoising and up-scaling LDCT images can be remarkably promising. This approach towards CT scanning in clinical sectors both guarantees a patient's lower risk of health threats as well as allows for preserving diagnostic benefits with substantial radiation dose reduction. A paper published by Montoya et al. (2017) in 'The American Journal of Neuroradiology' concludes that dosage reductions of 75%–90% were achieved for pediatric head CT used to evaluate craniosynostosis without affecting observer performance. In addition to this, despite LDCT being a promising component by itself, combining deep learning models and image processing techniques allows for noise reduction and generating high-resolution CT images to reap the benefits of high-dose CT images (HDCT) without exposing patients to unwanted radiation. In the field of LDCT image denoising, for instance, Tan et al. (2022) presented a novel unsupervised learning model based on CycleGAN that has shown the best experimental results in terms of suppressing LDCT picture noises and preserving edge and detail. [3]

The potential for image quality enhancement in LDCT scans using Generative Adversarial Networks is promising. Models based on GAN take the form of Deep Convolutional GAN or Super Resolution GAN, which can produce high-resolution target images based on low-resolution source images. Kang et al. (2017) proved through research studies that by using the GAN model, image quality is enormously improved so that all organ parts' details are made visible and made LDCT scans similar to HDCT scans, without high-dose radiation use [6] As a result, these models have better diagnostic accuracy while using reduced doses of imaging agents than earlier models. LDCT, although assuring low radiation exposure, provides noisy images with numerous peculiarities. Therefore, the extraction and classification of the noise patterns are shown to be important aspects toward generating an accurate clear view, which is upscale of LDCT to that of HDCT. A new variation of the generative adversarial network was proposed for noise extraction and noise pattern creation as well as transformation of noise characteristics from source image to target image along with conventional image denoising methods by Li et al. In their work, the GAN-NE model demonstrated quite good performance for the given task. [10]

Patient safety/health should be the major consideration of any technology implemented in the medical sector. Even though using lower doses of radiation assures patient safety, the disadvantages of low-dose CT imaging may be sorted by technologies such as deep learning and image processing. This brings out the aspect of how prospective modern technologies are and how they can be used in healthcare. The evolution of such technologies is indeed ongoing, and They help enhance patient outcomes, increase efficiency of delivery, and reduce disparities in healthcare. [9].

1.2 Problem Statement

As Computed Tomography (CT) scans provide detailed images of internal organs, they are now a widely used tool in the medical sector. However, cumulative exposure to CT scans increases the likelihood of cancer. Thus, to prevent this, people have delved into the idea of low-dose CT scans (LDCT) to decrease the ionizing radiation associated with CT scans. However, LDCT images are difficult for medical profes-

sionals to make a judgment since they have increased noise and decreased image details. All the image enhancement techniques so far come with a trade between suppression and the prevention of fine anatomical details. A balance is necessary between the two, as if the image is left too noisy it becomes difficult for analysis and if it goes through, over-smoothing can cause the fine details of the diagnostic to be less clear.

This research aims to enhance the image quality of LDCT images without compromising the fine details that may lead to an incorrect diagnosis for the patient. We seek to enhance the image quality of LDCT images using a class of deep-learning model which is known as Generative Adversarial Network (GAN). Our approach focuses on enhancing the image details of LDCT images while reducing the noise effectively. Therefore, it will become easier for medical professionals to conclude an accurate diagnosis. On top of that, the radiation risk for patients decreases.

1.3 Research Objective

The fundamental aim of this paper encompasses the exploration of novel methodologies to strengthen the purpose of using high-dose CT Images in clinical practices. The idea is to utilize noise patterns extracted from low-dose noisy CT images and generate noise patterns through the application of Generative Adversarial Network (GAN) based models to manipulate high-dose CT Images to perform a comparison study. The specific objectives that lie in front of us are delineated as follows:

- **Collection and Preprocessing of Low Dose CT Images Dataset:** We have collected dataset images from multiple hospitals as well as performing necessary preprocessing phases to the accumulated images to assure consistency in resolution, format, and features.

- **Noise Pattern Extraction:** Building a large-scale dataset of Low Dose noisy CT images from original HDCT dataset images and developing robust algorithms for noise pattern extraction from the collected Low Dose CT images to classify precision and accuracy in noise classification.

- **GAN-based Noise Generation:** Implementing cutting-edge GAN-based networks for learning realistic noise patterns, with the subsequently extracted noise patterns used as input to train GAN models that can generate artificial noise patterns like that of reality within Low Dose CT images are validated through the generated noise patterns are then validated using both quantitative and qualitative indicators, and their close resemblance to real-world noise features is ensured.

- **Implementation of Noise into High-Dose CT Images:** The characteristics of Low-Dose CT images by adding created and validated noise patterns to High-Dose CT scan images. Also, implementing the trained models to add simulated noise patterns to the source dataset's high-dose CT images without causing anatomical structures beneath to deteriorate.

- **Denoising and Upscaling Resultant Images:** Denoising the resulting images after implementing the noise patterns using up-to-date denoising algorithms using deep learning-based approaches and conventional image processing techniques. In addition, the same specific characteristics of noise, introduced in the process of implementing the noise pattern, are aimed to be visualized and restored into the image. Furthermore, a further developing upscaling technique to increase spatial resolution on denoised high-dose CT images yet retaining or near retaining all diagnostic information should be pursued.

- **Comparative Analysis and Evaluation:** Complete quantitative and qualitative comparison between processed images (whether upscaled or denoised) and the original high-dose CT scan images to establish that the diagnostic features and clinical relevance are maintained, conducting comprehensive evaluation.

1.4 Report Organization

In our paper titled 'Advanced Noise Reduction and Feature Enhancement in Medical CT Brain Imaging Using Deep Learning', the deep learning model will enhance Low-Dose Computed Tomography (LDCT) images to cope with the health hazards of multiple CT scans or the usage of High-Dose Computer Tomography (HDCT), such as the cellular damage and cancer. Traditional LDCT methods are associated with increased noise and lowered image quality, which are both negative factors affecting diagnostics performance and patient health. In this study, HDCT images, which are generated from noise-riddled LDCT, are enhanced by GAN technology to have the majority of the noise eliminated while keeping the anatomical structure at a proper level to enable more accurate diagnostic decisions. The work includes data acquisition and preprocessing of HDCT and LDCT datasets, as well as the generation and extraction from the LDCT datasets, and much more. Once LDCT images are artificially created and fed to the GAN algorithms for upsampling to HDCT, realistic noise patterns will be developed for high-quality dose images. The next step is then to compare that with the original high-dose CT. So, we have performed clustering method for unlabeled data for classification to check the authenticity and accuracy. Both qualitative and quantitative assessments, in totality, will evaluate to validate the performance GAN-based approach, provide useful insight, and verify our claims and research objective.

Chapter 2

Literature Review

The paper "Asymmetric Convolution-based GAN Framework for Low-Dose CT Image Denoising" by Saidulu and Muduli (2025) discusses an innovative denoising approach which is designed specifically for low-dose CT (LDCT) scans that often suffer from excessive noise due to reduced radiation exposure [11]. For addressing this problem, a GAN-based framework was proposed by the authors where the generator uses asymmetric convolutional blocks that have filters with varied shapes like 1×3 or 3×1 for capturing structural information from multiple directions as well. Compared to traditional square kernels, this asymmetry enables better extraction of features as it helps for reducing redundant computational complexities. Moreover, an attention mechanism is incorporated by this model which guides the network for focusing more on the most informative regions of the CT images by improving the denoising quality better. A standard architecture is followed by the discriminator in their framework in which the discriminator learns for distinguishing between real high-dose CT images and denoised outputs by refining the generator's learning. The authors acknowledge some limitations in their model, particularly in capturing long-range dependencies and deeper semantic relationships even after showing promising improvements over conventional methods. Additionally, as their model is limited to local asymmetric kernels and standard adversarial training, it may lead to over-smoothing or loss of fine anatomical details, especially in highly noisy regions or at edges and tissue boundaries. Although our proposed model is built based on their GAN framework, it introduces several significant advancements to overcome these challenges.

First of all, we have used a self-attention mechanism with a single head in our model that allows the generator to learn relationships across distant pixels by helping capture more global context within an image that is something in which asymmetric convolution cannot achieve. This is mainly useful in CT images where structural continuity and subtle textures are vital. Secondly, our generator is adapted on ESRGAN, instead of a shallow architecture which utilizes deep Residual-Dense Blocks (RDBs) for the enhancement of feature representation and it enables better learning of both fine and coarse details. Our attention operates over normalized, downsampled features to save memory and maintain efficiency, unlike the previous model's single attention block. Moreover, we have used a combination of L1 content loss, adversarial loss, SSIM loss, and perceptual loss functions which ensures that the output not only matches pixel-wise values but also preserves perceptual quality and maintains anatomical integrity. To summarize, the earlier model's limitations have

been overcome effectively by our model's contributions that improves texture clarity, reduces artifacts, maintains details consistency across large regions as well as it enables more stable training that makes our model more powerful as well as a practical solution to denoise LDCT image in clinical applications.

Moving on, the paper "SRTransGAN: Image Super-Resolution using Transformer based Generative Adversarial Network" introduces another innovative approach to single-image super-resolution (SISR) by integrating transformer architectures within a Generative Adversarial Network (GAN) framework [7]. The authors propose a transformer-based encoder-decoder generator capable of producing 2x and 4x upscaled images by recognizing the limitations of traditional Convolutional Neural Networks (CNN) in terms of capturing global contextual information. By enhancing the reconstruction of high-resolution images from their low-resolution counterparts, this generator leverages self-attention mechanisms to model long-range dependencies. By complementing the generator, a Vision Transformer (ViT) architecture is employed by the discriminator that processes images as sequences of patches, facilitating effective differentiation between real and synthesized images. A superior performance is demonstrated by the SRTransGAN framework which achieves an average improvement of 4.38% in PSNR and SSIM scores over existing methods as it provides saliency map analysis for elucidating the learning dynamics of the model. Our model introduces several enhancements to address the limitations and further improve super-resolution performance based on the foundational concepts which are presented in SRTransGAN. SRTransGAN may encounter challenges in preserving fine-grained local details and managing computational complexity even though it effectively captures global features through transformer architectures. Our model integrates a Multi-Head Self-Attention mechanism that operates on downsampled feature maps which strikes a balance between capturing long-range dependencies and maintaining computational efficiency in terms of addressing these problems. Furthermore, our generator architecture is inspired by the Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) which utilizes Residual Dense Blocks (RDBs) for facilitating deeper network training and better feature representation. The model's capacity for reconstructing intricate textures and structural details is enhanced by this design choice. In contrast to the two-stage approach of SRTransGAN, our model adopts a single-stage pipeline that streamlines the training process as well as reduces potential error propagation between stages. Moreover, a comprehensive lost function strategy was employed by our model which combines content loss, adversarial loss, structural similarity index measure (SSIM) loss and perceptual loss which ensures that both pixel-level accuracy and perceptual quality were maintained by the generated images. Our model addresses the limitations identified in SRTransGAN through these advancements by offering a more efficient and robust solution for image super-resolution tasks.

The paper named 'Medical Images Super-Resolution Using Generative Adversarial Networks' by Yuchong Gu et al. presents a deep learning method that uses GANs to improve the quality of medical images such as low dose CT scans and low field MRI scans which make the images blurry, sharper and more detailed [4]. There the models generate high resolution images that look realistic and preserve important texture details. To do this use a Generative Adversarial Network (GAN) where it helps

the model learn to generate high-resolution images that look more like real medical scans.Next the authors developed new generator models called Residual Whole Map Attention Network (RWMAN) which uses special attention layers that help to focus on important parts of the image and ignore less useful areas as keeping the final details and generating realistic patterns in the reconstructed super resolution images. However, this model can not handle long distance features well where it is limited also long range dependencies. In the training process incorporates a multi task loss function that combines content loss, adversarial feature loss which helps the MedSRAGAN models to generate images that visualize close to the original image. Also, the author did a Mean Opinion Score (MOS) test, where five experienced radiologists judged the image quality and the model got good ratings by them. Next the model uses two stage method which can sometimes lead to erroses passing one stage to the next. To address this, our model allows for better modeling of long range dependencies by incorporating a Multi-Head Self-Attention mechanism which improve the network to focus on different parts of the images simultaneously and it leads to improved super resolutions performance on especially medical CT images. Moreover, we use Residual Dense Block (RDB) allowing deeper network training and better feature presentation. Unlike MedSRGAN's two stage method, our model uses a single stage way which is simple and avoids potential error transfers propagation between stages. Furthermore, we use better set of loss functions the combination of content loss, adversarial loss, SSIM loss and perceptual loss to balance pixel level accuracy with helps make sure the perceptual quality of images. Lastly our models address the limitations from MedSRGAN by offering more efficient and more robust solution in medical images super resolution by integrating these advanced techniques.

The paper "Super-Resolutions using GANs for Medical Imaging" by Gupta et al.(2020) discusses how Generative Adversarial Networks (GAN) can be used to improve the quality of medical images such as MRI scans [5]. The authors build their model based on the SRGAN framework which includes a generator that creates high-resolution images and a discriminator which differentiates between real high resolution images with those generated images. Moreover their generator is trained not only to reduce pixel-level loss differences but also to match high level visual features using a VGG based perceptual loss for aiming the output images more realistic. Although this method performs better than simple basic interpolation and basic CNN based upscaling, it fails to keep original structure and fine details in clinical images. One major reason is that the generator is shallow and relies on simple residual design which limits its ability to capture complex features or fine grained textures. Additionally, the absence of attention mechanisms and limited receptive field hampers the model's capacity to capture long range dependencies or global anatomical coherence which are important requirements in clinical diagnostics where continuity of structure like blood vessels or tumors and soft tissues is critical which needs to be seen clearly. To solve these limitations, our approaches bring in several important upgrades to the network while still keeping the GAN-based framework. Firstly, we use 23 residual blocks in the generator network which improves the depth and nonlinearity of the network. It also helps the models to extract the high or low levels features from medical images. Secondly, we incorporated a Multi-Head Self-Attention mechanism after the RDB stack within the generator allowing the model to capture long range

dependencies by attending to relationships between distant pixels which are absent in the original SRAGN architecture. This feature is impactful for global context to preserve anatomical realism and preserve the correct shape of organs or tissues of medical images. Moreover, to perverse the texture and structural integrity, we used a combination of weighted loss like Content Loss, SSIM Loss, Perceptual Loss and Adversarial Loss. On top of that, we applied spectral normalization in the discriminator network to make training more stable. All these changes help our model solve the problems Gupta et al.'s by going deeper into small images with fine details and understanding the whole images better to look real and ensure global coherence and producing outputs that are both visually and diagnostically superior and demonstrate significant advancements in medical super resolution.

A 2023 study titled "Low-Dose CT Image Synthesis for Domain Adaptation Imaging Using a Generative Adversarial Network With Noise Encoding Transfer Learning" presents a GAN-based approach to enhance low-dose CT image quality [8]. The goal of their research was to effectively reduce noise present in LDCT images while preserving anatomical details and ensuring diagnostic utility across domains. Their proposed GAN-NETL model integrates noise encoding and transfer learning into a generative adversarial network (GAN) to synthesize LDCT images with realistic noise styles. To begin with, the problem statement of their study addresses the upsides of Low-dose computed tomography (LDCT) due to lower radiation exposure in clinical settings, yet how these LDCT images contain a significant amount of noise and artifacts due to fewer photons or undersampling, ultimately compromising image quality and diagnostic reliability. Additionally, they focus on the domain adaptation problem that indicates the challenges regarding deep learning models with a view to generalizing LDCT images from various scanners, protocols, or simulations that differ in noise style. Many studies have attempted to use synthetic noise or simulation to train denoising models but such models often fail to capture the real clinical noise characteristics due to domain shifts. The proposed GAN-NETL model in this study introduces a Noise Encoding Network into the GAN generator and combines it with a Transfer Learning (TL) strategy to enable domain adaptation. Basically how it works is that the framework synthesizes LDCT images with noise distributions that match those of the target domain. Therefore, it facilitates the training of denoising models using synthetic paired datasets.

A brief overview of the framework of their proposed model includes two parallel encoders and a two-stage training using generative adversarial network. The first encoder extracts anatomical structures from pre-processed CT images and the second encoder extracts noise features from original CT images in order to capture high-frequency noise components. Output from the second encoder fused with the content features extracted by the first encoder is then passed through the decoder, synthesizing the final LDCT image. To address the issue of domain specific noise, their two-stage training strategy comes into play. The 'Source Domain Training' trains both encoders using a public dataset that provides simulated LDCT images from original CT scans. In 'Target Domain Fine-tuning' training, the first encoder is frozen and the network is fine-tuned using a private dataset that shares noise properties with the target clinical data. This transfer learning step ensures that the noise encoding reflects the true noise distribution of real world LDCT scans and

performs generalization for domain adaptation. Their ablation studies confirm that when the noise encoder is removed synthesized images fail to replicate the realistic noise patterns observed in clinical data. Their GAN-NETL framework validates the effectiveness of the noise encoding and domain transfer mechanisms. To conclude, the GAN-NETL model used in this paper is an effective approach to separate noise from actual image content in CT scans with the usage of dual-encoder structure and two-stage transfer learning that help the model to effectively bridge the domain gap in LDCT image processing. This makes GAN-NETL a significant advancement in the literature of CT image synthesis and a practical tool for training deep learning models in real-world clinical settings with limited paired data.

Chapter 3

Workplan

We have done a vast review of the literature to get a grasp of the aspects of high and low dose CT imaging, noise and artifacts related to CT images, and proposed our model in terms of image processing and CT image denoising, especially in medical imaging [3]. The major subjects of discussion for this research will include low-dose CT imaging and their benefits and drawbacks, particularly the abundance of noise and lower picture quality. The acquisition of HDCT images from reputable hospitals will be needed greatly. The study's image processing part will contain Generative Adversarial Networks or GAN's creative applications in enhancing medical images. The goal of the current study is to close conceptual holes in present techniques and set up the conditions for further practical research and the potential benefits for the medical industry. This phase includes collecting and going through primary data that are vital to the study. The primary dataset will consist of high-dose CT images. Our entire workflow focuses on addressing a critical challenge in modern medical imaging, that is, reducing patient exposure to harmful radiation during CT scans while preserving diagnostic image quality. The central goal of our work is to prove that, although low-dose CT (LDCT) images are noisy and degraded in terms of quality, they can be utilized to mimic the quality of high-dose CT (HDCT) images as closely as possible using image processing and deep learning techniques. This allows healthcare providers to safely perform scans at lower radiation levels without compromising on diagnostic accuracy.

With that in mind, we will begin by collecting a large set of primary CT scans from several reputed hospitals. Unlike many existing studies that rely on secondary CT image datasets, our work is grounded in real clinical data. The chances are that we will acquire high dose CT images from different sources. However, those HDCT images will exhibit varying degrees of noise. Instead of discarding those noises, we will use it for our cause. Using noise extraction, noise encoding and transfer learning technique, we will isolate and extract reusable noise patterns present in the HDCT images. We will then amplify these noise patterns and apply to our HDCT images to synthetically create LDCT-like noisy versions of the original images. This will give us paired data, noisy images aligned with its clean and corresponding high dose images. This will later be beneficial for training a supervised deep learning model based on ESRGAN.

Before moving into training a deep learning model, we will apply a series of preprocessing steps to the HDCT images to improve their quality and consistency. These

will include some nonlinear filtering methods and denoising techniques that will be particularly important to ensure that the HDCT images serve as clean and reliable ground truths during the learning phase. Then, with the dataset consisting of high-quality HDCT images and their corresponding noise-amplified LDCT variants, we will proceed to train a deep learning model based on ESRGAN with our own modification and improvements in terms of model architecture. After training the deep learning model, we will be able to generate denoised versions of the LDCT images that closely resemble the original HDCT images. This part is crucial since we are trying to prove that we can achieve similar or closer diagnostic results by utilizing these generated denoised versions of noisy LDCT images compared to those of the HDCT images.

Moving further, we will turn to the second part of the study that focuses on classifying the original HDCT images based on their labels that we acquired during data collection. There is a high chance that we might have to deal with a mix of labeled and unlabeled data and therefore we might need to perform a semi-supervised labeling technique based on visual features and assign labels to unlabeled images consistent with the labeled subset. We then will train a classification model like DenseNet-121 on the labeled HDCT dataset to assess its diagnostic performance. The model's effectiveness will be evaluated using standard metrics such as accuracy, precision, recall, and F1 score. Finally, we will test the classifier on the generated, denoised LDCT images we acquired using the deep learning model from earlier. This will allow us to make a comparison between the classification performances of generated images and original HDCT scans, supporting the viability of using denoised LDCT images in clinical settings. Moreover, it will help us to come to a conclusion whether or not our deep learning model not only produced visually similar images but also retained diagnostically relevant information.

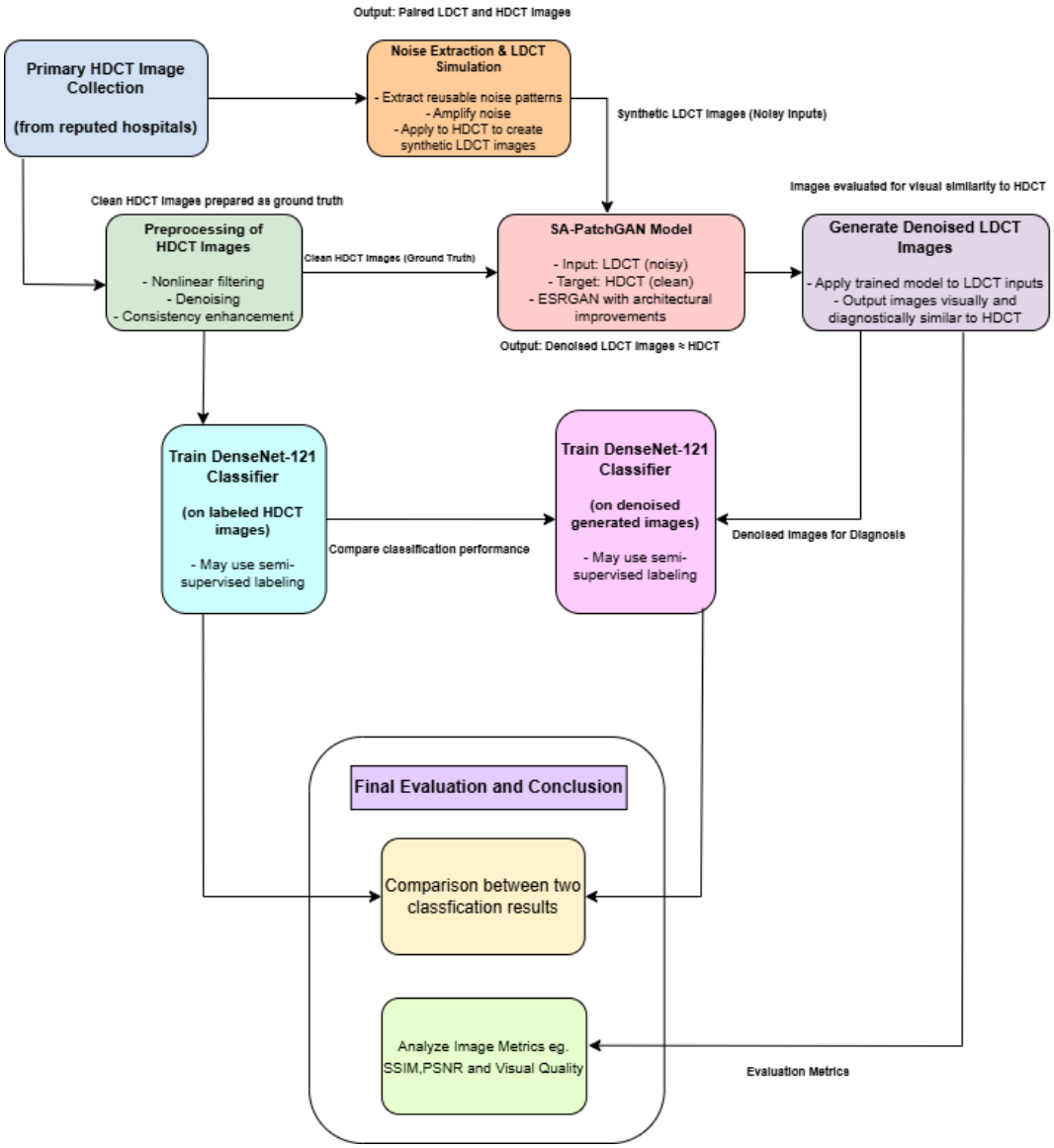


Figure 3.1: Workplan

Chapter 4

Dataset

4.1 Data Selection

CT scans are the second most common medical imaging technique used in the world, behind X-rays. CT (computer tomography) scans provide better accuracy and more details of their intended view of the organs. Additionally, CT scans provide a vast amount of perks in a lot of medical situations that other high-resolution techniques, such as MRI's, cannot provide. They offer better speed and better availability and, in certain aspects, offer better detailed images for bone structures, etc. But the main problem with CT scans is the increased need for radiation and the unfortunate amount of noise due to the modalities way of imaging. Many of those downsides come from the innate desire of patients to move during the scans. In addition, if we were to consider in the perspective of Bangladesh, a significant amount of CT scan and imaging machines are either slightly defective or cannot function fully to its supposed level of photon attenuation, beam hardening, etc. These unfortunate factors contribute to noise degradation and repeated scans. This increases the patient's body to multiple doses of radiation. If the foretold problems can be solved using machine learning, the repeated scans and the health of the patients would be solved. Thus, computer tomography fits perfectly with our research objectives.

Since we chose to work with CT images, the next step was which organ we choose for our area of focus. CT can be performed on various types of organs in the human body. The kidneys, stomach, pelvic region, etc. are the most common areas that are scanned for abnormalities by CT scan. But of all the organs of the human body, we chose the brain to focus on the research. Brain CT images provide us with a vast amount of difficult yet doable challenges due to the complexities of the brain. The brain structure and its muscle through the CT images are inherently marred by noise. Due to the brain and muscle creases, working on the subtle differences of gray matter, small lesions, or in extreme cases, brain hemorrhages can prove to be difficult. But through image refinement and resolution upscalling, we can remove these obstacles and improve the qualitative aspects for deeper image analysis and abnormalities detection. Also, by increasing the image resolution and making the anatomical details much better, it will vastly help the medical professionals and radiologists in identifying problems. Thus, we have chosen brain CT images for our primary data set to be worked with and researched for this project.

4.2 Data Collection

CT images of the brain can be very difficult to source for research. Since medical images highly confidential, especially in regards to sensitive data such as CT scans and MRI's. Since Bangladesh has the same types of issues that plague other fields for the acquirement of datasets. We have obtained the CT scans of the brain for this research from two esteemed medical and hospital institutions in Bangladesh. One of the first ones was from Anwer Khan Modern Medical College Hospital(AKMMCH). The hospital is well-regarded in the country for their advanced radiology departments, primarily the CT scan imaging. Also because of their significant patient throughput, which made them suitable hospital for acquiring the required datasets. The entire journey started with providing authorization documents to the hospital for their permission. There, we have worked with the radiologists preceding in the department. Due to the delicate nature of patients personal medical history and their data, we ensured to both parties that no personally identifiable information (PII) were included in the dataset. Every image was taken and cleaned of any identifiable characteristics or metadata that could breach patient confidentiality and put the aforementioned patients in trouble.

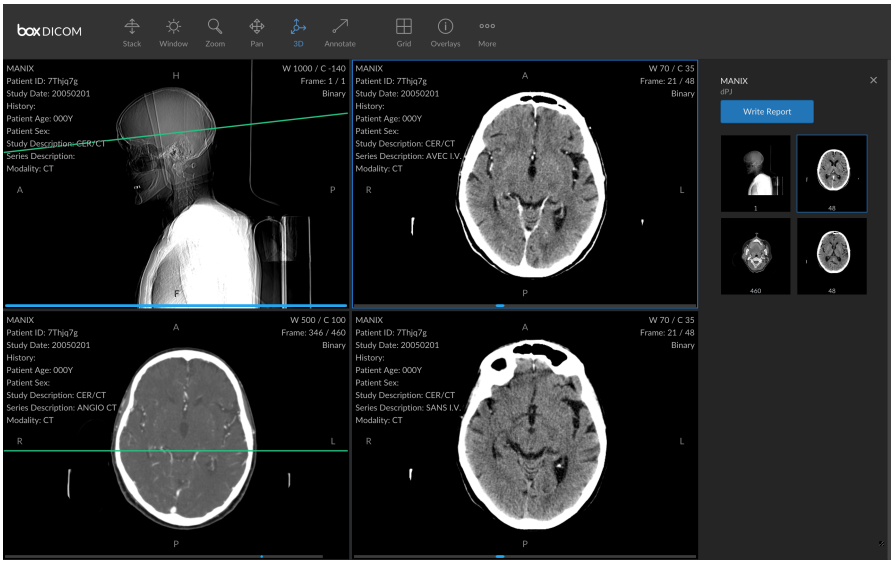


Figure 4.1: Screenshot of DIACOM Viewer

However, the data collected from the AKMMCH was flawed in many ways. Due to the way how the radiologist department was organized, we were not able to use a lot of the data. Since our research requires the report and CT scan images from the exact same patient, we had to forego a lot of valuable data that we had received. Since the vast majority of them did not have matching patients. Despite that, we managed to match 3 patients' CT scans and their corresponding reports for our dataset collection. The dataset had a variety of CT scans of the patients' brain, from routine check-up imaging to potential head injuries. Besides the images, we were also able to get the reports associated with those patients. These reports provides us the necessary ideas and more importantly, context behind each patients CT scans. They also include the radiologists findings and overall medical impressions in regards to the patients situation. On the other hand, we were also able get CT scans and

16 reports from Shaheed Suhrawardy Medical College (SSMC) hospital, thanks to their sonography department which precedes over the CT scan imaging sector. But just like AKMMCH, their data providing method was also quite scattershot. We received well over 20 patients and over 200 reports from brain CT scan patients. However, the majority of them were not the same patients. Thus proving a difficult task of recovering the needed data and its' corresponding reports of the heap. At the end, only 4 patients had their CT scans and matching data for us to work on. Overall, we were able to collect 7 patients with their reports, totaling in 2670 images. Afterwards, we collected more than 90 patients CT scans without their corresponding reports. Thus we put them the unlabeled class.

4.3 Data Organization

After we have gathered all the patients' CT scans and their corresponding reports, we decided to sort the dataset into different, distinct classes. The reason being the proposed models for learning and applying the appropriate upscalling algorithms. We sorted the unsorted dataset into individual classes, each representing a particular ailment or anatomical deformity. We also added a "Normal" class, where the patients have no abnormalities. After reviewing all the reports of the patients, we came to the conclusion that 6 classes must be created and all the patients must be categorized into these six classes. The classes are **cerebral infract**, **cortical atrophy**, **lipoma**, **normal**, **ischemic changes** and **unlabeled**.

Class	Cerebral Infarct	Cortical Atrophy	Lipoma	Normal	Ischemic Changes	Unlabeled	Total
Patients	2	1	1	5	1	91	100
Images	702	71	647	1277	9	9372	12078

Table 4.1: Distribution of patients and images across classes.

4.4 Data Visualization

Here, we can see the distribution of the classes of patients, sorted by their CT scan impressions.

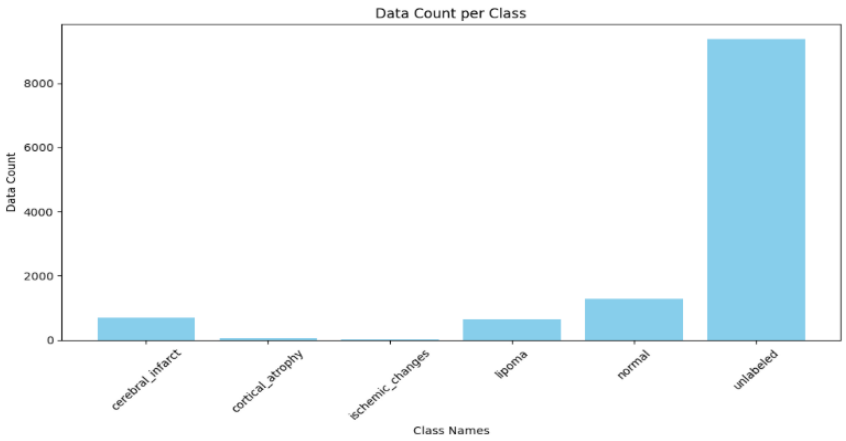


Figure 4.2: Class Distribution of CT Scans

The class distribution can be seen being very lob-sided, where most the images are **unlabeled**. The rest being moderate in numbers except for **ischemic changes** which only account for 9 images.

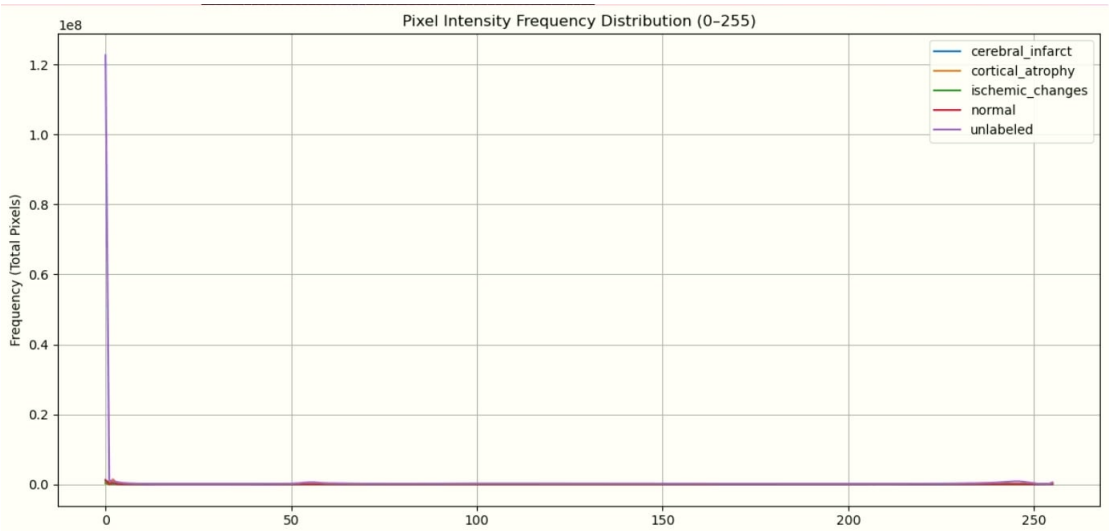


Figure 4.3: Average Pixel Intensity Distribution

Unsurprisingly, the pixel intensity of the images, regardless of class are more or less aligned with one another. The vast majority of the images and their pixel intensity all along the x and y axis.

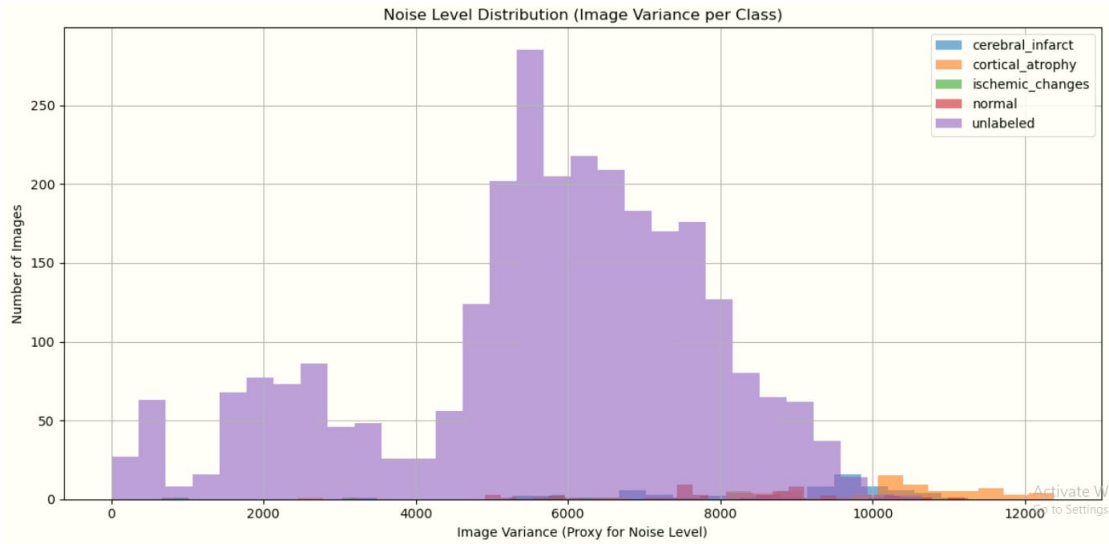


Figure 4.4: Noise Level Distribution Across Classes

The graph above shows the variance of noise levels among images of the dataset. Of the five classes, the vast majority of the noise levels are primarily from the **unlabeled** class and the rest being the other 4. The **unlabeled** images clustering around a variance of 5000 to 7000. The rest are hovering over the 8000 to 12000.

4.5 Data Preprocessing

The preprocessing stage is an incredibly important step for any machine learning research. Especially for those in regards to computer vision and image processing. When there is a large set of data which are in different sizes, color values and unmanageable, preprocessing takes those data and transforms them into manageable, standard and easier to work with data. Very important steps such as scaling and cropping, gray-scale conversion, and normalization are crucial for making the model work and getting the best performance out of it. It also ensures that we can analyze the data more easily since the dataset is much simpler and uniform. The steps are detailed below:

4.5.1 Grey-scale Conversion

Gray-scale conversion is usually done for datasets that require the images to be gray. The process is taking the images' RGB values and turning them into gray, single-channel images. Since the images are now single-channel and gray, brightness and the intensity of said brightness determine the performance of the model. For tri-channel image datasets, this poses a problem and degrades the model's performance. However, our entire dataset is CT scans of the brain. Since some models, especially GAN-based models, focus more on shapes and patterns rather than colors, making gray-scale images makes them more effective for specific tasks. Thus, all the images must have uniform gray-scale values.

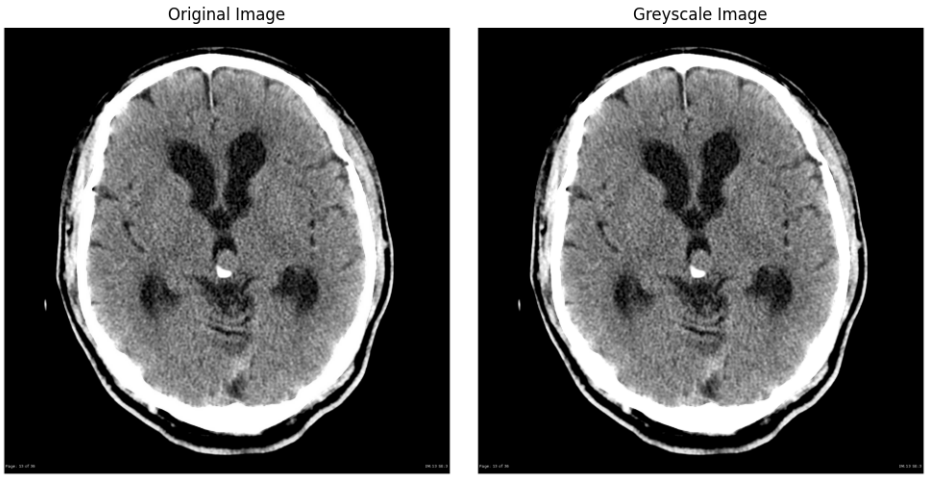


Figure 4.5: Gray-scale Conversion of Dataset

4.5.2 Cropping

Cropping refers to the crucial process of removing unnecessary and unusable parts of the image to focus more on the region of interest. In the case of CT scans, this typically involves cropping the anatomical region, such as the brain, while eliminating irrelevant areas like background or other body parts that might have been captured during the scan.

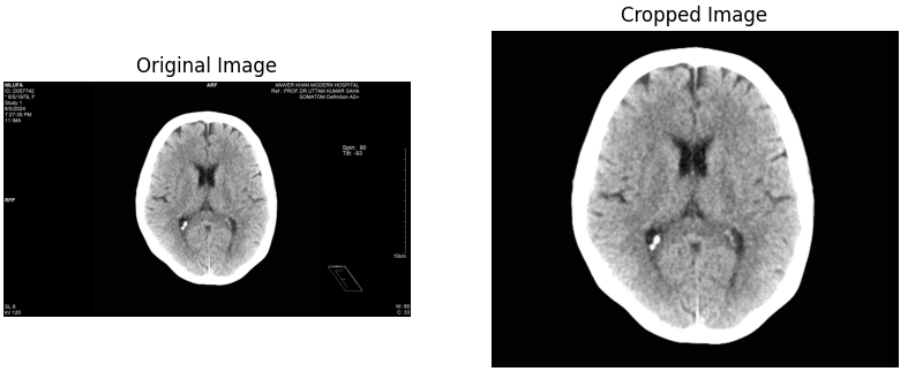


Figure 4.6: Cropping of Dataset

Chapter 5

Methodology

5.1 Architectural Diagram

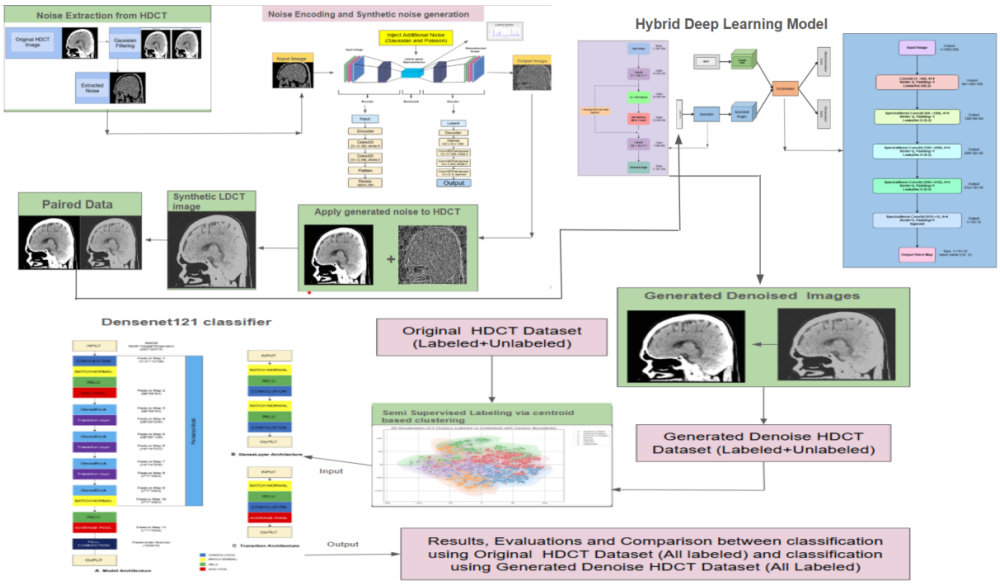


Figure 5.1: Full Architecture

5.2 Noise Encoding for Synthetic Noise Generation

'Noise Encoding' is a technique that involves introducing random variations of 'noise' to data or models during training to improve the robustness and adaptability in terms of deep learning. This method can be advantageous because the addition of noise makes our data or model resilient to real-world imperfections and variations, improving its ability to perform well despite noisy or imperfect data and to transfer knowledge to related data. In our context, we are working with medical CT imaging dataset. However, Low-dose CT scans are often more difficult to obtain in large quantities. Moreover, obtaining High-Dose CT images and their corresponding Low-dose CT images are even more difficult when our goal is to construct a paired dataset. Although HDCT images are of higher resolution, they are not exactly free of noise due to the noise and sensitivity of medical imaging equipment and real

world artifacts. However, noise is not necessarily a bad thing and can be a valuable feature in terms of learning noise patterns in order to generate corresponding Low-dose CT images to construct a paired dataset. This is where Noise Encoding becomes beneficial.
Approaching this method of 'Noise Encoding' involves several steps. Firstly, it's crucial to identify noises and separate them from structural components of the CT-Scan images. This will help us visualize the noise patterns and give us an insight regarding noise intensity and other important factors. Once we are done working with noise extraction from the original HDCT images, we can move to the next step of encoding the noise patterns and perform necessary modifications to encoded noises. Doing so, we can generate random and intensified noise patterns using the original encoded noises and decode them in order to apply them to our original HDCT images and generate synthetic LDCT images. In our approach, we transform high-dose CT images into synthetic low-dose images by utilizing Autoencoder(AE) for latent space representation of encoded noises.

5.2.1 Noise Extraction

Noise extraction is a crucial part of image processing, particularly for medical images like CT scans, as noise can hide important details and impact the accuracy of diagnoses or analyses. In our case, we need to extract the noises from original HDCT images in order to obtain noise patterns that will be used later in our workflow. We can achieve this using 'Gaussian filtering' in which we generate a filtered and smoother version of the original and noisy HDCT image and subtract it from the original image to extract noise. 'Gaussian filtering' is a denoising technique that filters an image by averaging pixel values according to a Gaussian distribution. Pixels that are closer to the center have a greater impact than those farther away. Assigning weight to each pixel based on its distance from the center is performed by the Gaussian Kernel. The Gaussian function G(x,y) that defines the weights of the filter is given by:

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

Where:

- $G(x, y)$ is the weight for the pixel at position (x, y).
- σ is the standard deviation, controlling how much smoothing is applied (larger σ = more smoothing).
- x and y are the coordinates of the pixel relative to the center.

Then the Gaussian Kernel performs convolution in order to multiply pixel values by the assigned weights, and the weighted sum of these pixels is then assigned as the new value for the center pixel. After that, by averaging the pixel values in the neighborhood using Gaussian weights, random pixel variations are smoothed out to remove noise. Therefore, pixel at (x,y) in the new image I (x,y) is computed by:
The Gaussian filter can be applied as follows:

$$I'(x, y) = \sum_{i=-k}^{k} \sum_{j=-k}^{k} I(x + i, y + j)G(i, j)$$

Where:

- $I(x, y)$ is the original image.
- $I'(x, y)$ is the filtered image.
- $G(i, j)$ is the Gaussian kernel.
- The sums go over a neighborhood of pixels (determined by the kernel size, typically k).

Finally, we get the noise pattern as, `Noise = (Original Image-Filtered Image)`.

$$I'(x, y) = \sum_{i=-k}^{k} \sum_{j=-k}^{k} I(x + i, y + j)G(i, j)$$

Where:

- $I(x, y)$ is the original image.
- $I'(x, y)$ is the filtered image.
- $G(i, j)$ is the Gaussian kernel.
- The sums go over a neighborhood of pixels (determined by the kernel size, typically k).

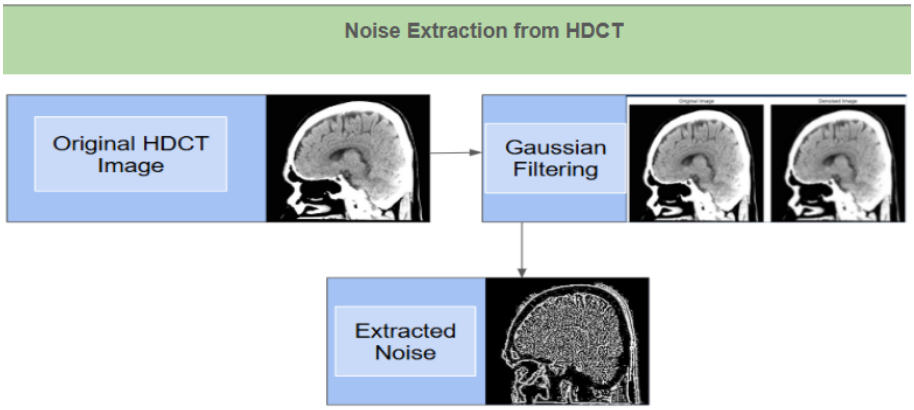


Figure 5.2: Noise Extraction

5.2.2 Noise Encoding

Noise extraction from images is followed by our next step, which is 'Noise Encoding'. Essentially, the goal is to use a neural network that learns to encode image data into a lower-dimensional latent space and then reconstruct the data back from this compressed representation. This is crucial for our approach because lower-dimensional latent representations allow us to make changes to noise vectors, for instance, introducing additional noise or intensifying existing noise etc. We need to decode the encoded noises after performing modifications to generate noise patterns and apply the noise patterns to our original HDCT images. Our used neural network for this

task is 'AutoEncoder (AE)'. With a view to encoding, decoding or generating noise patterns, 'AutoEncoder (AE)' is structured with encoder, latent space and decoder.

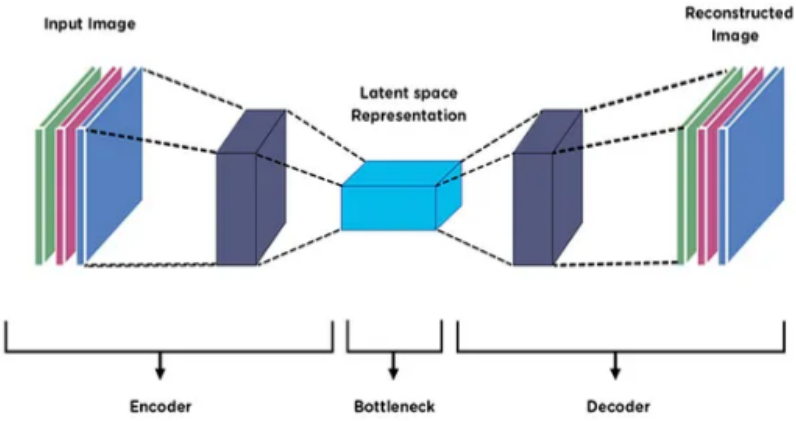


Figure 5.3: Auto-Encoder

5.2.3 Encoder

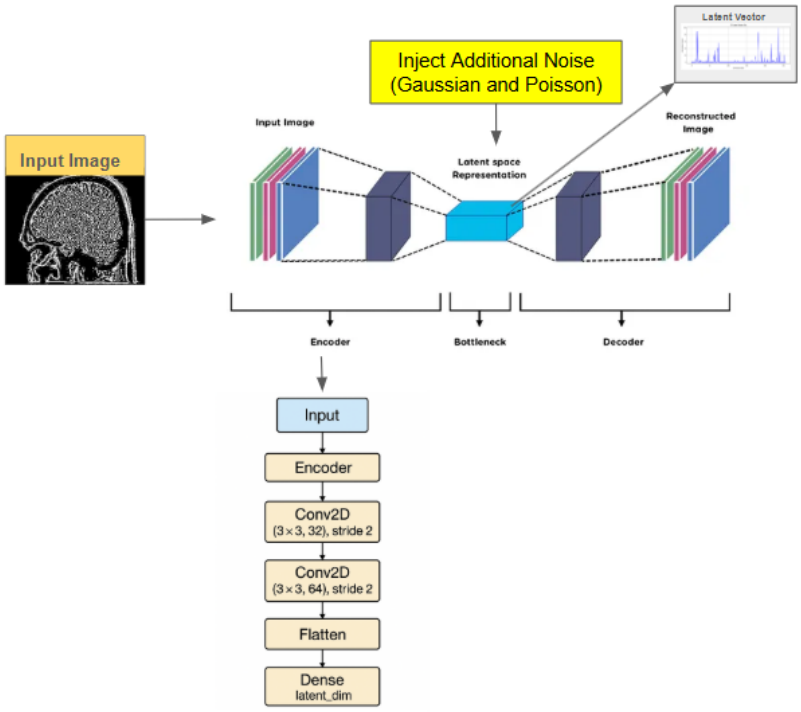


Figure 5.4: Encoder

The encoder learns a function h=f(x), where x is the input image and h is the latent vector. The latent vector or latent space is a compressed representation of image data. The image is passed through the encoder network, where each layer progressively reduces the dimensions. For example, an image represented as a 2D matrix of

pixel values converts into a 1D array as its corresponding latent representation. The input image is transformed into a latent vector by applying multiple transformations (linear or nonlinear) in the hidden layers of the encoder.

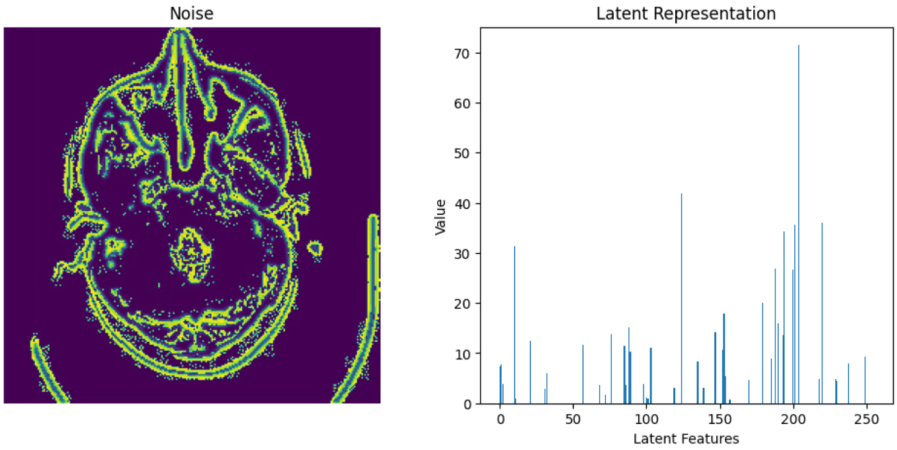


Figure 5.5: Noise encoded to its latent representation

5.2.4 Latent Space

The encoder transforms noises into latent representations of the original noises. Although the dimensionality of the input is reduced in latent space to make the noise information more compact, latent vectors contain the most important information and essential features about the input. Therefore, we introduced 'Gaussian noise' and 'Poisson noise' to the latent space to produce more intense noise patterns by modifying the original latent vectors. Gaussian noise follows a Gaussian distribution, altering each pixel with a random fluctuation that has a continuous distribution. This type of noise is unbiased, as it is centered around zero. If X_{original} represents the original signal and X_{noisy} represents the signal with added noise, Gaussian noise modifies or intensifies the original signal.

$$X_{\text{noisy}} = X_{\text{original}} + N$$

Where N follows a Gaussian (normal) distribution:

$$N \sim \mathcal{N}(\mu, \sigma^2)$$

Here, μ is the mean of the Gaussian distribution that is often zero for centered noise. σ^2 is the variance, which decides how much the noise fluctuates around the mean. A higher variance means stronger noise. Introducing Poisson noise to the original input could provide a realistic noise pattern because these types of noises are commonly found in images and related to photon counts in imaging systems. Since we are using medical imaging data, adding Poisson noise to latent space can create realistic and intensified noise patterns. If the noisy value of a pixel is represented as X_{noisy}, it is the sum of the original signal X_{original} and the Poisson noise. Poisson noise follows a Poisson distribution with a mean λ, where the pixel intensity of X_{original} is proportional to λ.

$$X_{\text{noisy}} = X_{\text{original}} + \text{Poisson noise}$$

Modifying the latent space with Gaussian noise and Poisson noise results in producing an intensified but realistic latent space of the encoded noises that we can later decode and generate random noise patterns to apply to our original HDCT images to generate their Low-Dose version.

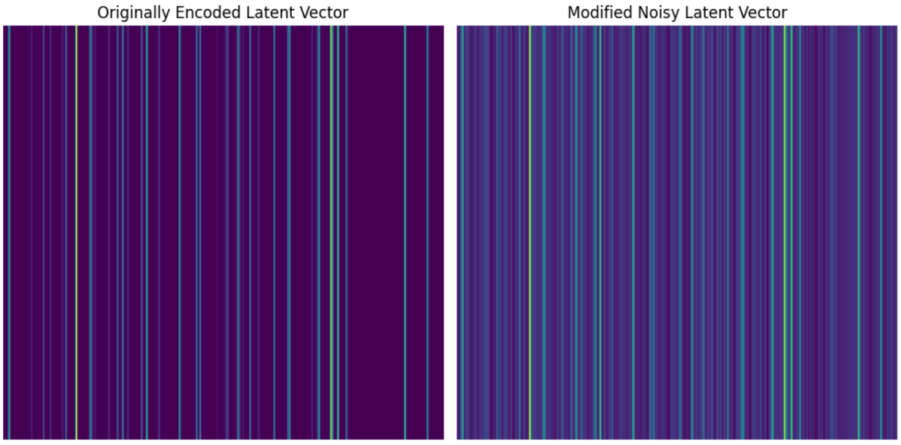


Figure 5.6: Latent space before and after introducing additional noise

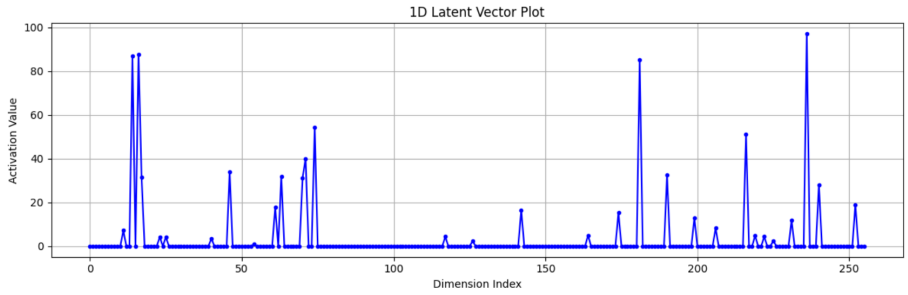


Figure 5.7: 1D Vector Plot

5.2.5 Decoder

The decoder performs the reverse operation of the encoder. The decoder is trained to map the latent vector h back to the original image x =g(h). The modified latent vector is passed through the decoder, and it performs the reverse operation of the encoder, which transforms the latent vector back to image representation. When the decoder returns the intensified noise patterns, we take random noise patterns from the decoded noises and apply them to our original HDCT images to produce a low-dose version of the original HDCT images.

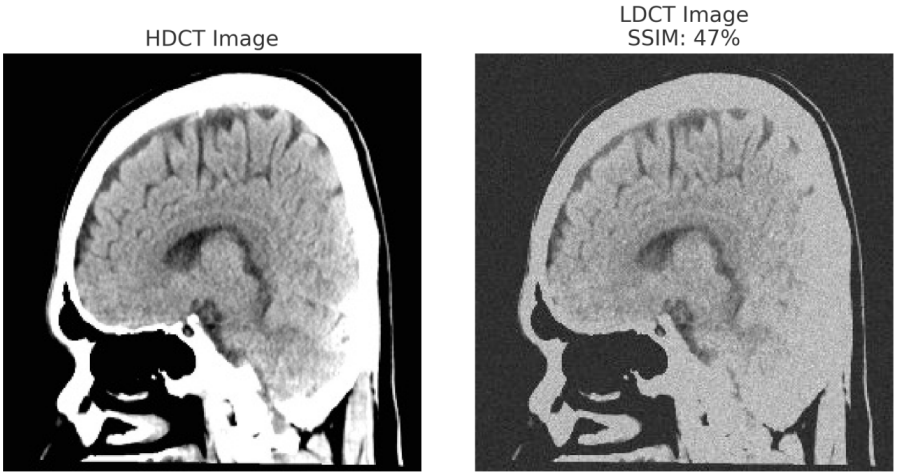


Figure 5.8: Comparison between HDCT and LDCT

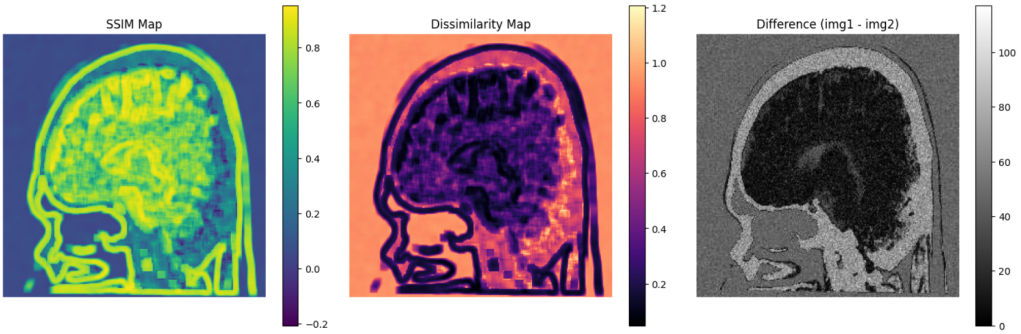


Figure 5.9: Comparison between HDCT and LDCT

The autoencoder is trained to minimize the difference between the input image and the output reconstructed image. This difference is quantified by a loss function, for example, mean squared error. The goal is to reduce the reconstruction error as much as possible during training. That's why we can see the generated LDCT image has SSIM (Structural Similarity Index Measure) very close to the original HDCT image.

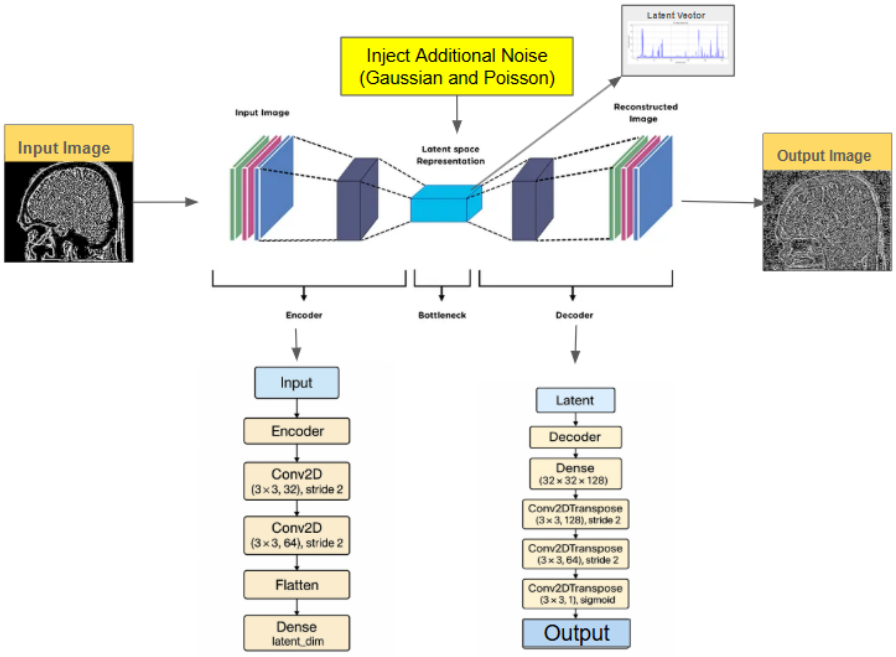


Figure 5.10: Decoder

5.2.6 Full Pipeline for Noise Encoding and LDCT Generation

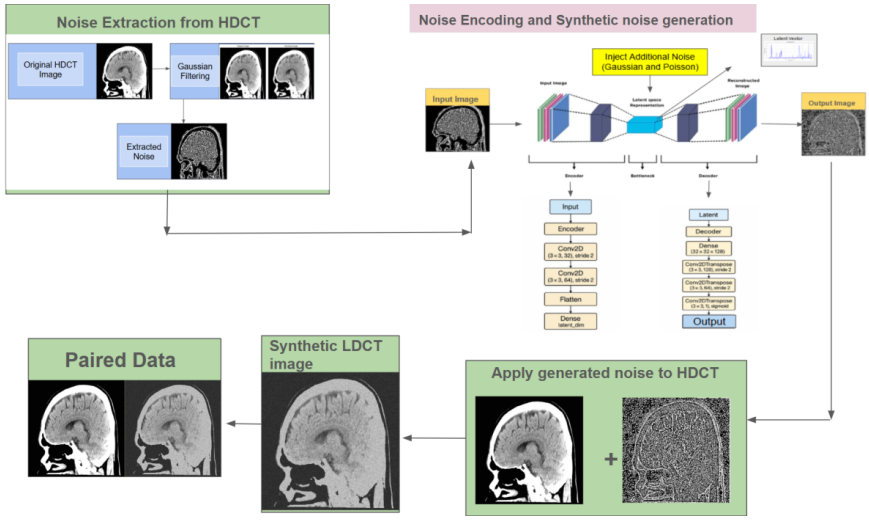


Figure 5.11: Full Pipeline

5.3 Model Architecture

For denoising low-dose CT (LDCT) images, our proposed work utilizes an ESRGAN-inspired generator and a PatchGAN-style discriminator. Medical images, in general, require the preservation of structure and fine details for accurate diagnosis. Thus, both networks have been designed to preserve these fine details and textures while denoising.

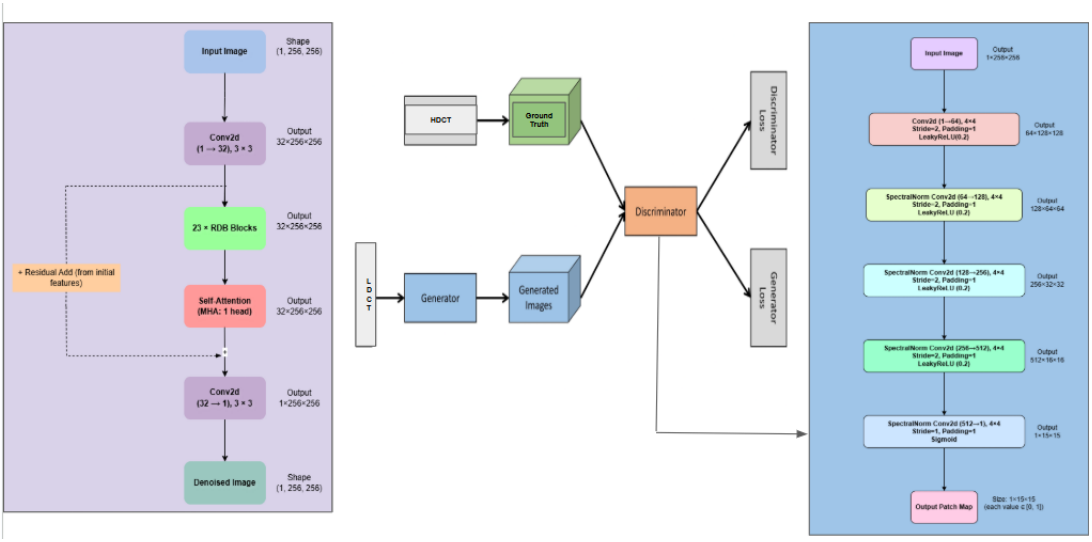


Figure 5.12: Entire Architecture

5.3.1 Generator Network

The implemented generator is adapted from the ESRGAN architecture where an attention mechanism is incorporated after the Residual Dense Blocks (RDB) stack. We replaced the original Residual-in-Residual Dense Blocks (RRDB) with the simplified Residual Dense Blocks (RDB). This reduced architectural complexity and memory consumption and improved training efficiency. The input LDCT images are processed via the following components:

- **Initial Convolution Layer:** The single-channel input LDCT images are converted into an embedded feature space of 32 channels.

- **Residual Dense Blocks(RDB):** In order to extract deep features and enable efficient gradient flow, 23 RDB blocks were used. Each RDB block contains three convolution layers along with LeakyReLU as the activation function. By residual connection, the output from each RDB block is added to its own input. This design ensures that the original content of the images remain intact along with learning new and relevant features. Furthermore, this encourages stable training and helps preserve low level information.

$$F_{\text{rdb}} = F_{\text{in}} + \text{RDB}(F_{\text{in}})$$

  where F_{in} is the feature map inputting Residual Dense Block (RDB), $RDB(F_{\text{in}})$ is the output from the Residual Dense Block (RDB) (3 conv layers + activations), F_{rdb} is the sum of input feature of RDB and the processed feature out of RDB.

- **Multi-Head-Self-Attention Module:** After all the Residual Dense Block (RDB) stacks, a Multi-Head Self-Attention (MSA) module is applied. This not only helps to enhance global context but also model long-range dependencies. Just before applying attention, the input feature map is first compressed using bilinear interpolation only if the input feature map exceeds 128×128. This helps to reduce computational overhead and memory usage, making the

model more efficient and scalable for high-resolution image. Self-attention is then applied and resulting feature maps are upsampled back to their original dimensions. This module F_{attn} is computed as:

$$F_{\text{attn}} = \text{MSA}(F_{\text{rdb}})$$

where F_{rdb} is the output feature map from the Residual Dense Block (RDB) stack. This mechanism helps the model to retain both local details and global structures for the input image. Convolutional Neural Networks (CNN, which in our case RRDB), are only effective at capturing local details but not global context. Thus, incorporating MSA helps the model by enabling global context awareness.

- **Final Convolution Layer:** To preserve low-level structural information (e.g., edges, boundaries) a skip connection is used since these features may be lost deeper within the network. The skip connection employed adds the early-layer features F_{initial} from the first convolution layer to the output of the attention-enhanced features. The resulting denoised (generated) image I is given by:

$$I = \text{Conv}(F_{\text{attn}} + F_{\text{initial}})$$

where F_{attnl} is the output of the attention layer, F_{initiall} is the feature map from the first convolutional layer.

The final Convolutional layer combines the output of refined feature map with the structural information. This design enables the generator to learn both fine details and structural information which are crucial for accurate interpretation of CT images.

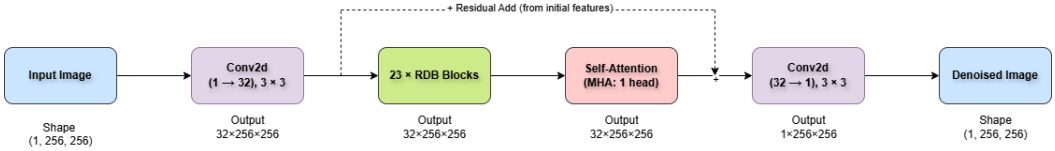


Figure 5.13: Generator

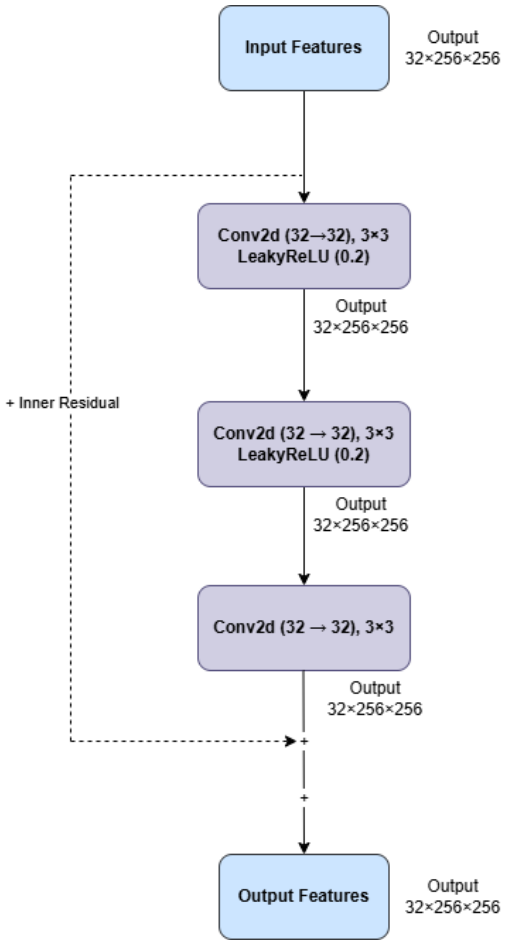


Figure 5.14: Residual Dense Block (RDB)

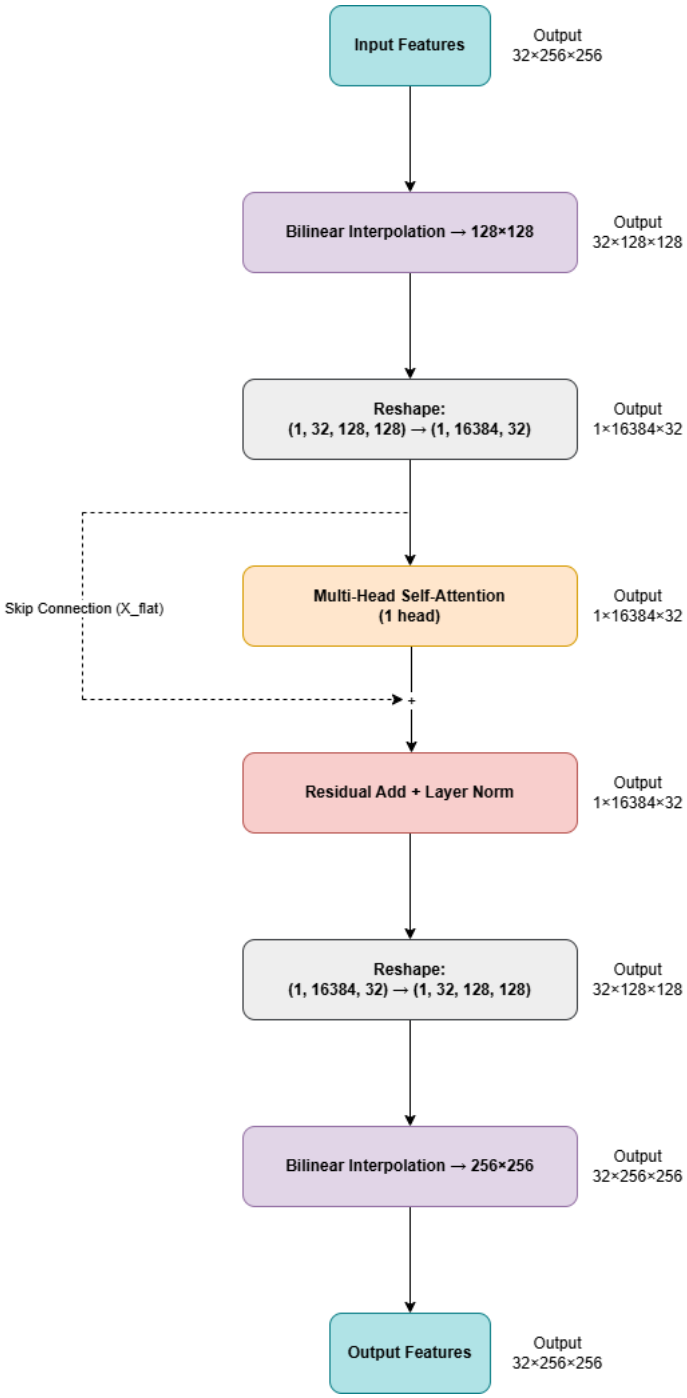


Figure 5.15: Multi-Head Self-Attention (MSA)

5.3.2 Discriminator Network

The implemented discriminator follows the PatchGAN style discriminator. Instead of evaluating the whole image, it classifies local patches as real or fake. This helps to enforce texture realism.

- **Convolution Layers:** The discriminator consists of several convolutional layers. Each of these convolutional layers is followed by LeakyReLU as the activation function. Spectral normalization stabilizes adversarial training. It also helps to prevent exploding gradients.

- **Final Convolution Layer:** The final layer uses sigmoid activation function to output probabilities of each patch.

$$D(I) = \sigma(\text{Conv}_n(\ldots(\text{Conv}_1(I))))$$

These patch-wise probabilities enable the generator to produce perceptually and structurally realistic images. The lightweight discrimination improves generalization and accelerates training, making the system more robust. In medical images, subtle features can turn out to be significant from a diagnosis perspective. This discriminator design helps to focus on texture and structural fidelity and preserve these critical details.

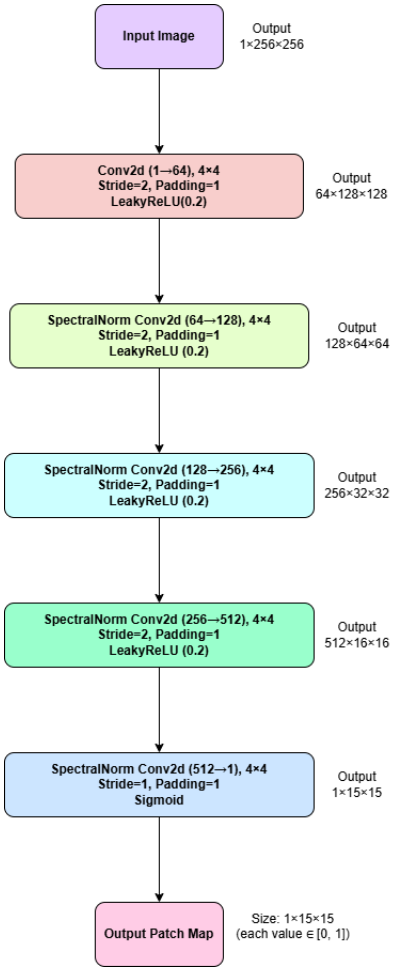


Figure 5.16: Discriminator

5.4 Loss Functions

During training, a combination of loss functions are employed which encouraged the generator to produce outputs that not only align with the ground truth (high dose CT) image but also are visually clear, structurally accurate and medically reliable. Each loss function used focused on a specific aspect of image quality like pixel accuracy, texture, structure, and perceptual appearance. Each loss function used is described in detail in the following subsections.

5.4.1 Content Loss

The content loss is used to ensure that the denoised image matches to the ground truth image at the pixel level. We computed this using the L1 norm. The L1 loss calculates the average absolute difference between each corresponding pixel of the generated and the ground truth images. L1 loss tends to preserve sharp edges and details (which are important for medical images). Compared to L2 loss, L1 loss is less sensitive to outliers, which encourages the generator to produce (denoised CT) images that are closer to the ground truth (high dose CT) images in terms of structure and intensity at the pixel level.
The content Loss L_{Content} is defined as:

$$L_{\text{content}} = \|\hat{I} - I_{\text{HDCT}}\|_1$$

where $\hat{I}$ is the generated (denoised CT) image and I_{HDCT} is the ground truth (high dose CT) image.

A smaller difference means that the generated (denoised CT) image is similar to the ground truth (high dose CT) image.

5.4.2 SSIM Loss

Preserving anatomical structures is a crucial in the reconstruction of medical CT image. Pixel-wise losses ensures the numerical similarity of the corresponding pixels between the ground truth (high dose CT) image and the generated (denoised CT) image. To address this issue, we implemented Structural Similarity Index (SSIM) Loss which measures the perceptual difference between the ground truth (high dose CT) images and the generated (denoised CT) images based on luminance, contrast, and structure. This is very similar and aligns more closely with human visual perception. We used SSIM-based loss as a similarity score and tuned our models hyper-parameters to get higher SSIM score. SSIM score ranges from 0 to 1, where 0 corresponds to no similarity and 1 corresponds to perfect similarity. The SSIM Loss L_{SSIM} is defined as:

$$L_{\text{SSIM}} = 1 - \text{SSIM}(\hat{I}, I_{\text{HDCT}})$$

where $\hat{I}$ is the generated (denoised) image and I_{HDCT} is the ground truth (high dose CT) image.

5.4.3 Adversarial Loss

Content Loss and SSIM focus on pixel-wise and structural similarity. But a high Content Loss and SSIM does not guarantee that the image will look visually realistic. As a result, Adversarial Loss was introduced through the discriminator to encourage the generator to produce visually realistic denoised CT images. The model was trained within a typical GAN framework where the generator and discriminator are trained in competitive settings. By employing Adversarial Loss, the generator not only learned to denoise an image but also to enhance the fine textures and visual realism. The discriminator learns to distinguish between the ground truth (high dose CT) image and the generated (denoised) image, while the generator learns

to fool the discriminator. Adversarial Loss punishes the generator whenever the discriminator correctly identifies an image as fake.
The Adversarial Loss L_{loss} for the generator is defined as:

$$L_{\text{adv}} = -\log D(\hat{I})$$

where Î is the generated (denoised) image from the generator and *D(Î)* is the discriminators probability of $\hat{I}$ being real.

The generator is able to fool the discriminator if *D(Î)* is close to 1. So the generator generated a good image. But if *D(Î)* is close to zero then the discriminator knows that the image is fake. Thus, the generator tries to maximize *D(Î)*. Or in Loss terms, minimize the negative log of *D(Î)*. In practice, we use Binary Cross-Entropy (BCE) Loss. We write the same idea using BCE.

$$L_{\text{adv}} = \text{BCE}(D(\hat{I}), 1)$$

As explained before, the generator wants *D(Î)* to be 1 i.e. real. Here, BCE compares the discriminator's output [*D(Î)*] with the label 1. If *D(Î)* and 1 are close, the loss is small indicating a realistic image. But if it deviates then the loss increases. So BCE Loss guides the generator to produce outputs that are indistinguishable from the ground truth (high dose CT) image in terms of both realism and fine textures.

This is how Adversarial Loss is different from the other losses mentioned above. Pixel-level losses (like L1 of Content Loss) only focus on pixel level difference which lead to blurry outputs when trained alone. On the other hand, Adversarial Loss encourages the generator to produce realistic textures.

5.4.4 Perceptual Loss

Pixel-wise and structural similarity losses ensure local accuracy but may fail to capture higher-level features like edges, contours and textures i.e. the overall visual appearance of a CT image. As a result, images may be closer pixel-wise but they may not appear perceptually correct. To address this issue, we employed Perceptual Loss, which compares the feature representations of the generated (denoised) and ground truth (high dose CT) images extracted from pretrained Convolution Neural Network (CNN) over comparing raw pixels. We used the VGG16 network pretrained on ImageNet dataset which learned to recognized high-level features. Our generated (denoised) and ground truth (high dose CT) images were passed through VGG16 where features were extracted from an early layer i.e., the layer after the first few convolutions. The Perceptual Loss then compares the VGG feature maps of the generated (denoised) image and the ground truth (high dose CT) image, encouraging the generator to match high-level features of real images. VGG16 expects 3-channel RGB images. However, our images were of grayscale so we had to replicate them to three channels from one to make them compatible. The perceptual Loss L_{perc} is defined as:

$$L_{\text{perc}} = \|\varnothing(\hat{I}) - \varnothing(I_{HDCT})\|_2^2$$

where Ø is the feature extraction function from the VGG16 network. It is the squared difference between VGG features of the generated (denoised) and the ground

truth (high dose CT) image. $\hat{I}$ is the generated (denoised) image and I_{HDCT} is the ground truth (high dose CT) image.

5.5 Training Strategy

We followed an adversarial setup while training our proposed GAN model. Both the generator and discriminator were trained simultaneously where the generator was trained to optimize the minimization of the weighted sum of Content Loss, SSIM Loss, Adversarial Loss and Perceptual Loss. On the other hand, the discriminator is trained so that it can successfully distinguish between the ground truth (high dose CT) image and the generated (denoised CT) image. With each epoch, both the networks improve by challenging one another; a key fundamental idea in GAN based learning.
While training, Adam optimizer was used for both the networks. The learning rate was set to 2×10^{-4} and momentum parameters were set as $\beta_1 = 0.5$ and $\beta_2 = 0.999$. In order to prevent gradient explosion, gradient clipping with a maximum norm of 1.0 was applied to both the networks which helped to maintain training stability. As regularizer, label smoothing was used in the discriminator which helped preventing the discriminator from becoming overconfident and allowed smoother gradient flow back to the generator.

After each epoch, model checkpoints were saved so that if any interruption occurs, training can be restarted from the same position. Also, a side by side visual comparison between the ground truth, generated and noisy image was generated after each epoch in the model train and the evaluation metrics of generated images are saved in a CSV file which allowed to track the progress of the model over time. Generator are are also check pointed at every epoch for result analysis. Early stopping mechanism is implemented but was not used, allowing the model to train for a fixed number of epochs.

The total Generator Loss L_{G} is the weighted sum of all the individual loss components.

$$L_G = L_{\text{Content}} + \lambda_{\text{adv}}.L_{\text{Adv}} + \lambda_{\text{SSIM}}.L_{\text{SSIM}} + \lambda_{\text{perc}}.L_{\text{perc}}$$

In the final training run, the weights were set as follows:

$$L_G = L_{\text{Content}} + 0.12.L_{\text{Adv}} + 0.05.L_{\text{SSIM}} + 0.16.L_{\text{perc}}$$

5.6 Semi-Supervised Labeling

Our dataset contains approximately 12,000 brain CT scan images collected from reputed hospitals which are both labeled and unlabeled. Among these, around 2,700 images are associated with patient reports and thus considered labeled, while the remaining 9,000 images are unlabeled. To assign labels to the unlabeled images, we implemented a semi-supervised labeling strategy based on centroid-based clustering. Using a pretrained DenseNet model for feature extraction, we generated feature vectors for all labeled images. These vectors were grouped by class and the mean vector

for each class was computed to serve as its centroid. Next, feature vectors for the unlabeled images were extracted and compared to each class centroid using cosine similarity. If an unlabeled image showed a similarity score equal to or above a set confidence threshold (e.g., 0.80), it was assigned to the corresponding class. Images falling below the threshold were moved to an "unknown" folder to avoid potential mislabeling. In cases where an image matched multiple centroids closely or had conflicting similarities, it was placed in an "overlap" folder to prevent classification ambiguity. Furthermore, we visualized the distribution of both labeled and pseudo-labeled data using t-SNE for dimensionality reduction and KDE plots to highlight class density regions.

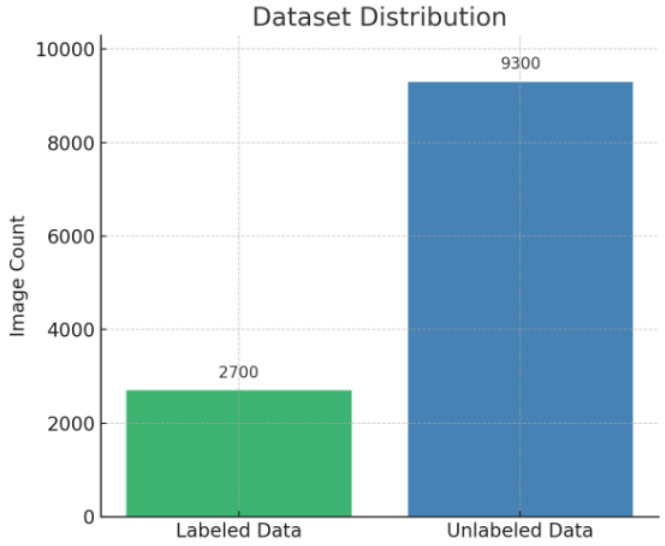


Figure 5.17: Data Distribution

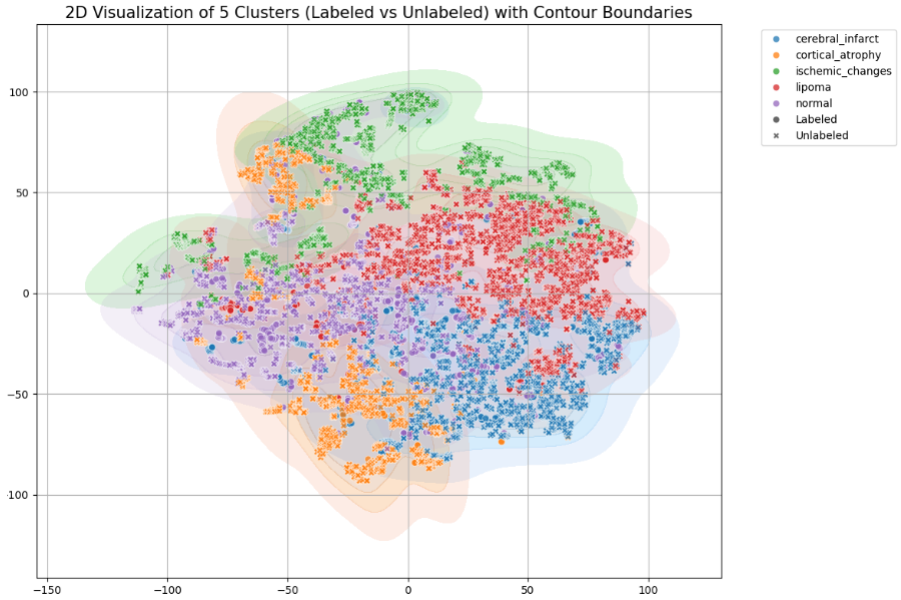


Figure 5.18: 2D Visualization of 5 Clusters

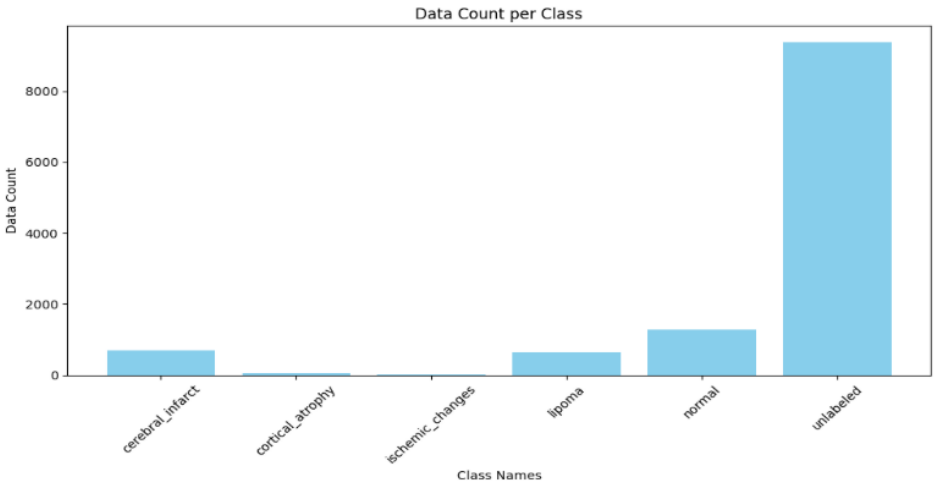


Figure 5.19: Data Distribution before Semi-Supervised Labeling

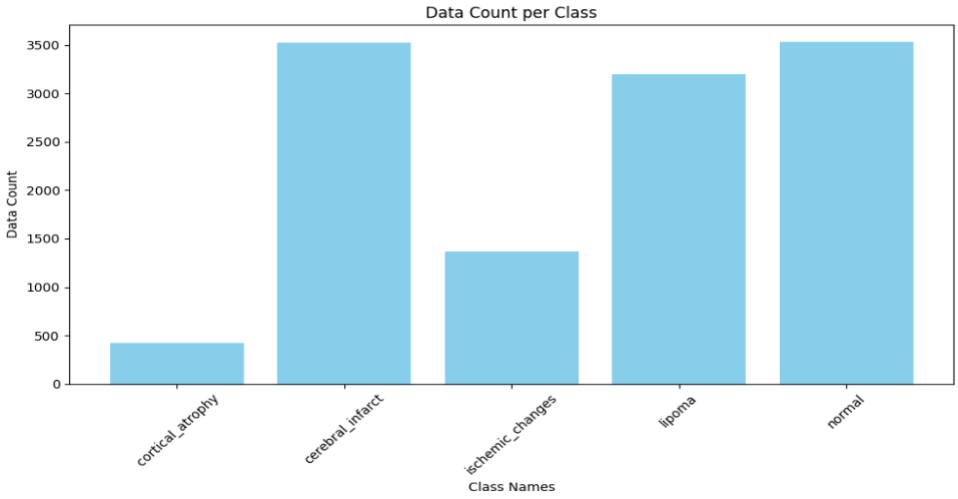


Figure 5.20: After Semi-Supervised Labeling (Clustering)

5.7 DenseNet-based Classification

DenseNet121 as the name suggests consists of dense blocks and this is a popular Convolution Neural Network (CNN) architecture. In this model, each of the layers receives inputs from all the previous layers. This allows efficient gradient flow, as features learned in previous layers are preserved. This pretrained model of DenNet121 was trained on over 14 million images, which had over 1000 classes. Since this is a pre-trained model, the weights are already initialized while being trained on image net dataset. For our specific task, the pre-trained layers of the model were frozen so that the weights do not get updated when it is trained using our dataset. This also prevents the model from being overfitted.

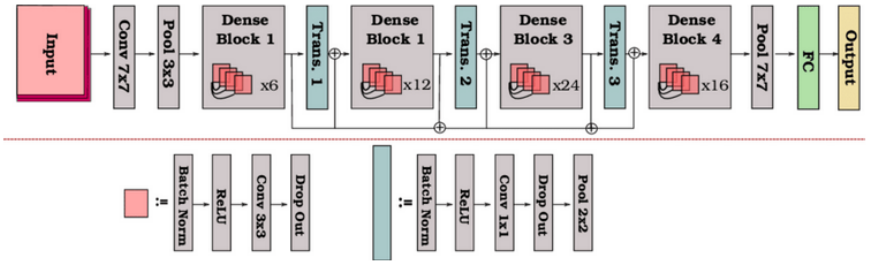


Figure 5.21: Architecture 0f DenseNet121

So, when we trained this model using our own dataset, only the weights of the newly added layers got updated. These new sequential and custom layers were added so that this model gets adapted to our specific classification tasks.

GlobalAveragePooling2D: This was used in place of a traditional fully connected dense layer in order to reduce the number of parameters, avoid overfitting and summarize spatial information from features maps efficiently.

Dense(1024, activation='relu'): To introduce non-linearity, Rectified Linear Unit (ReLu) was applied in this dense layer of 1024 neurons. This activation function outputs the unit directly if it is positive, or 0 otherwise. This helps the model learn complex patterns.

$$f(x) = \max(0, x)$$

$$f(x) = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$$

Dropout(0.5): This is a regularization technique which is introduced to prevent overfitting. So, during each iteration, 50% of the neurons in this layer are randomly turned off. This forces the model to generalize better. On top of that, this forces the model to use multiple neural pathways throughout the network instead of relying on over specific neurons. This also forces the network to become robust and further encourages them to learn new independent features.

Dense(4, activation='softmax'): Since we have 4 classes, the final dense layer comprises 4 units. Each of these neurons corresponds to one different class. Softmax is used as an activation function. This converts the output of each neuron into a probability score which if added will be equal to 1.

Mathematically, for each output z_i, softmax transforms it as:

$$\text{softmax}(z_i) = \frac{e^{z_i}}{\sum_j e^{z_j}}$$

So, each neuron gives a probability score between 0 and 1. Then the class which has the highest score is chosen as the model's final prediction.

Optimizer: As an optimizer, Adaptive Moment Estimation (ADAM) was used and this combines both RMSProp and Stochastic Gradient Descent (SGD). This automatically adjusts the learning rate during training and helps the model to converge by minimizing the loss function.

Loss Function ('categorical_crossentropy'): As loss function, categorical cross entropy was used. This measures the difference between the predicted probability distribution and the actual class distribution. In short, this measures how well the network's predictions match with that of the actual class labels.

$$L(y, \hat{y}) = -\sum_{i=1}^{C} y_i \log(\hat{y}_i)$$

Where:

- y is the true label (one-hot encoded).
- $\hat{y}$ is the predicted probability for each class output by the softmax layer.
- C is the total number of classes.
- The logarithmic term penalizes incorrect predictions (since $\log(1) = 0$ but $\log(0)$ is undefined and penalizes highly).

The model is trained so that this loss function minimizes and adjusts its weight so that in the next iteration it produces a higher probability for the correct class.

ReduceLROnPlateau: If the validation loss plateaus, and the model stops improving for two consecutive epochs then ReduceLROnPlateau callback dynamically reduces the learning rate by half. But the learning rate will not go below 0.00001

EarlyStopping: If the validation loss does not improve for 3 consecutive epochs, then EarlyStopping call back stops the model from training any further which prevents the model from being overfitted to the training data. In addition, after the training stops, the model's weight gets restored to the point where the best validation loss was achieved during training.

Training the Model: The model will train for a maximum of 50 epochs. However, if validation loss stops improving and does not improve for 3 consecutive epochs, training will stop earlier as early stopping is in place. Batch size 16 was used, so the model updates its weight after every 16 samples. A greater batch size was not used due to RAM crashing issues. Also, this smaller batch size helped the model to generalize better.

Enconding Class Labels: Here, we are doing multi class classification where the goal is to classify the input image into one of the four categories. For this, we encoded the class labels into a numerical format and then applied one hot encoding to prepare the labels for the classification.

5.7.1 Label Encoding

- `cortical_atrophy` $\rightarrow 0$
- `cerebral_infarct` $\rightarrow 1$
- `lipoma` $\rightarrow 2$
- `normal` $\rightarrow 3$

5.7.2 One-Hot Encoding

$$\text{Label } 0 \rightarrow [1, 0, 0, 0]$$
$$\text{Label } 1 \rightarrow [0, 1, 0, 0]$$
$$\text{Label } 2 \rightarrow [0, 0, 1, 0]$$
$$\text{Label } 3 \rightarrow [0, 0, 0, 1]$$

Here, each of the binary vectors represents one of the classes. The element corresponding to the correct class is set to 1 while all others are set to 0. This one hot encoded vector makes it easy to compare the model's output with that of the true labels during training.

5.7.3 Train-Test Splitting

The dataset is split into two parts:

- 80% for training
- 20% for testing

The training set is further split into two parts:

- 80% for pure training
- 20% for validation during training

To sum up:

- 64% of the data is used for training
- 16% for validation while training
- 20% for testing

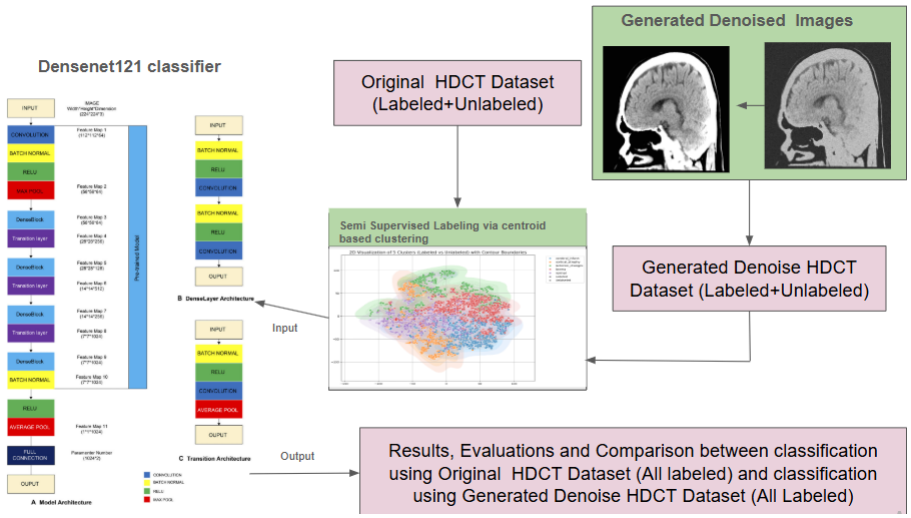


Figure 5.22: DenseNet Classification Pipeline

Chapter 6

Experiments and Results

6.1 Evaluation Metrics

In order to evaluate the quality of the generated (denoised CT) images, we used several quantitative metrics. To track the model's improvement and convergence behavior, we used one single image to calculate evaluation metrics after each epoch. After training completes, evaluation metrics were carried out on the whole test set to evaluate the model's generalization capabilities across the entire test set.

At the pixel level, MSE provides an interpretable measure of image reconstruction. Nevertheless, it does not account for perceptual or structural differences that is visually significant. MSE computes the average squared difference between the generated (denoised CT) image and the ground truth (high dose CT) image.

$$\text{MSE} = \frac{1}{HW} \sum_{i=1}^{H} \sum_{j=1}^{W} (I_{SR}(i,j) - I_{GT}(i,j))^2$$

where HW is the product of image height and width, I_{SR} is the generated (denoised CT) image and I_{GT} is the ground truth (high dose CT) image.

6.1.1 Peak Signal-to-Noise Ratio (PSNR)

PSNR is derived from MSE and is expressed in decibels (dB). PSNR is the ratio between the maximum possible signal i.e. pixel intensity and the power of noise i.e. reconstruction error. A higher PSNR value indicates a better image quality. However, it does not correlate with how humans visualise an image.

$$\text{PSNR} = 20.\log_{10}\left(\frac{\text{MAX}_I}{\sqrt{\text{MSE}}}\right)$$

where MAX_1 represents maximum possible pixel intensity which is 255 for 8-bit images.

6.1.2 Structural Similarity Index Measure (SSIM)

SSIM score tells the preservation of anatomical structures. Compared to MSE or PSNR, SSIM models the human visual system more closely. SSIM compares luminance, contrast and structural patterns between two corresponding generated

(denoised CT) images and the ground truth (high dose CT) image. The SSIM score ranges from 0 to 1, where 0 corresponds to no similarity and 1 corresponds to a perfect structural match.

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

where C_1 and C_2 are stability constants, μ_x^2 and μ_y^2 the local means, σ_x^2 and σ_y^2 is the variance and σ_{xy} is the covariance between the images.

6.1.3 Signal-to-Noise Ratio (SNR)

SNR computes a value that represents the ratio of the signal power to the noise power. For our tasks, it helps us to assess how much noise is removed from the generated (denoised CT) image. A higher SNR value indicates that useful information was preserved while suppressing the noise. Thus, high SNR value represents a clean image whereas a low SNR value represents an image with noise.

$$\text{SNR} = 10.\log_{10}\left(\frac{\sum I_{GT}^2}{\sum (I_{GT} - I_{SR})^2}\right)$$

Here, I_{GT} is the ground truth (high-dose CT) image, and I_{SR} is the generated (denoised CT) image.

6.1.4 Sharpness

For medical CT imaging, accurate representation of anatomical boundaries is essential, which depends upon the retention of the edge sharpness. When sharpness is retained, the edges and fine details of an image are preserved. To calculate sharpness, we use variance of the Laplacian of the image. Laplacian highlights areas of the image where the intensity changes quickly. We essentially calculate the variance of this Laplacian response. A high variance indicates a sharper image, whereas a low variance indicates an image with less defined edges, i.e,. less sharp.

$$\text{Sharpness}(I) = \text{Var}\left(\overline{V}^2 I\right)$$

Here, $\overline{V}^2 I$ represents the laplacian of the image upon which variance is calculated. I represents the input image.

6.1.5 Contrast

Medical images are grayscale images and thus contrast plays an essential role in distinguishing different anatomical structures in CT images. Contrast is calculated as the variance of the pixel intensities. A high contrast implies a stronger difference in intensity across the pixels of the image.

$$\text{Contrast}(I) = \frac{1}{HW}\sum_{i,j}\left(I(i,j) - \mu\right)^2$$

Here, HW is the product of image height and width, $I(i, j)$ represents the pixel intensity at the position (i, j) and is the mean intensity of all pixels.

6.2 Visual Results

Visual comparison was carried out to evaluate the output of the proposed GAN-based denoising model. The comparison involved three types of images: ground truth (high dose CT) image, generated (denoised CT) image and lastly the noisy (low dose CT) image. The test data consisted of diverse test samples to evaluate if the model generalized across all types of noise levels and anatomical areas, mimicking real world scenarios. The comparison images in sequence consist of:

- **LDCT** - which shows noise and loss of structural information.
- **Generated Image** - which shows that the generator reduced noise, restored anatomical structure as well as enhanced contrast.
- **HDCT** - This is the ideal output which we are taking as a benchmark.

Figure 6.1: Visual Comparison of $wr11\_159.bmp$ Image

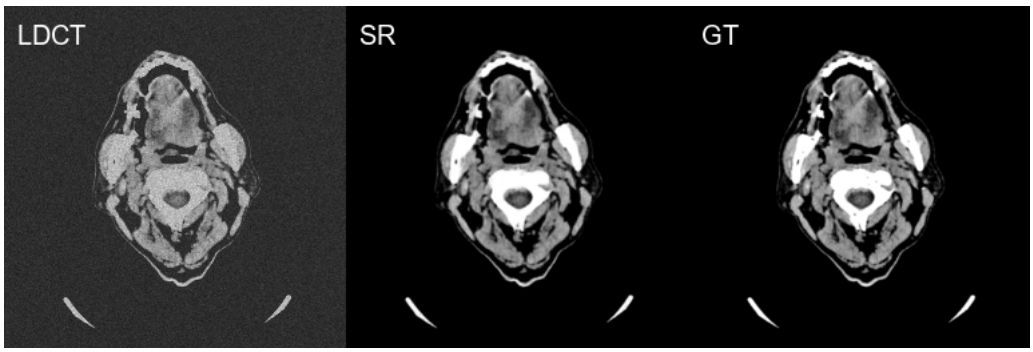


Figure 6.2: Visual Comparison of $Wr02\_01.bmp$ Image

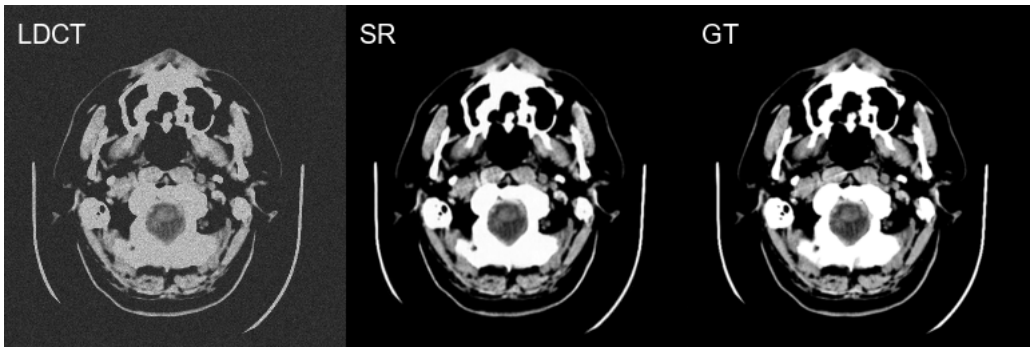


Figure 6.3: Visual Comparison of $wr41\_3.png$ Image

Figure 6.4: Visual Comparison of $CT05BACK5.bmp$ Image

The visual evidence suggests that the generator was able to restore and retain the structural features and suppress noise.

6.3 Quantitative Results

On the entire test set, quantitative evaluation was carried out to assess the performance of the proposed GAN model. The low dose CT images are fed into the trained generator, which generates the denoised CT image. Each output of the generator was compared with its corresponding ground truth (high dose CT) image. The evaluation metrics used for comparison are:

- Mean Squared Error (MSE)
- Peak Signal-to-Noise Ratio (PSNR)
- Structural Similarity Index (SSIM)
- Signal-to-Noise Ratio (SNR)
- Sharpness
- Contrast

For each metric, the score was not only computed image-wise but an average score for each of these metrics was computed for the entire test set which provided us a statistical figure to measure the model's denoising performance.

Image	MSE	PSNR	SSIM	Sharpness_GT	Sharpness_Gen	Contrast_GT	Contrast_Gen	SNR
Wr01_297.bmp	16.30	36.01	0.9332	1075.86	642.60	72.46	71.67	3.50
Wr02_01.bmp	11.40	37.56	0.9720	1285.08	714.91	60.45	59.53	2.07
Wr03_444.bmp	11.61	37.48	0.9717	1172.91	688.04	66.68	65.82	2.08
Wr04_01.bmp	13.71	36.76	0.9671	873.66	556.03	75.96	75.33	2.51
Wr05_121.bmp	15.61	36.20	0.9428	1242.10	751.41	67.77	66.87	3.45
Wr06_01.bmp	14.53	36.51	0.9613	1455.33	866.01	67.67	66.58	2.04
wr07_540.bmp	20.44	35.03	0.9432	1832.47	1209.01	79.63	81.93	1.82
wr08_134.bmp	20.83	34.94	0.9363	2182.69	1370.28	84.22	83.87	2.51
wr09_510.bmp	14.24	36.60	0.9611	1116.78	719.42	77.62	77.56	2.38
wr10_290.bmp	17.19	35.78	0.9471	1085.53	705.72	83.94	84.46	2.42
wr11_159.bmp	21.97	34.71	0.9286	1937.04	1231.58	75.28	75.03	2.48
wr12_87.bmp	22.53	34.60	0.9308	2353.48	1543.27	93.13	92.45	2.20
CT01FRONT12.jpg	39.02	32.22	0.7276	3117.01	2419.32	99.81	101.18	1.27
CT02FRONT04.jpg	19.09	35.32	0.7441	2417.82	1755.01	93.64	93.52	0.97
CT04BACK5.bmp	18.52	35.46	0.8995	1423.83	1239.06	75.06	74.24	2.64
CT05BACK5.bmp	23.52	34.42	0.8876	2693.52	2291.51	79.28	78.16	2.29
CT06TOP02 (102).bmp	18.81	35.39	0.9418	2836.25	2167.36	73.78	72.99	1.16
CT07TOP02 (134).bmp	22.45	34.62	0.9240	4753.22	3251.94	91.16	90.16	0.93
wr23_42.png	20.67	34.98	0.9288	2054.58	1135.11	74.80	73.68	2.29
wr28_1.png	14.55	36.50	0.9575	2574.96	1537.06	80.63	79.20	1.88
wr29_7.png	20.69	34.97	0.9317	1064.23	558.76	81.40	80.13	2.11
wr41_28.png	8.43	38.87	0.9776	1183.48	796.28	79.10	77.82	3.07
wr41_3.png	15.58	36.21	0.9672	2373.55	1502.19	78.90	77.48	1.79
wr50_1.png	14.92	36.39	0.9678	2387.30	1362.24	68.50	66.69	1.58
wr54_9.png	15.59	36.20	0.9631	2300.95	1305.88	72.46	72.16	1.33
wr73_12.png	13.29	36.89	0.9642	1803.89	1022.00	91.06	89.50	1.99
wr77_15.png	14.93	36.39	0.9661	2569.12	1416.02	73.57	72.45	1.40
Average	**17.79**	**35.82**	**0.9313**	**1969.13**	**1287.33**	**78.44**	**77.80**	**2.08**

Table 6.1: Comparison of Image Quality Metrics for Different Images

The results clearly demonstrate that the model outperformed the low-dose CT inputs by generating improved (denoised CT) images. Higher SSIM and PSNR account for improved structural and perceptual quality whereas sharpness and contrast signify the enhanced visual quality.

6.4 Classification Results

We performed two separate classification tasks using Densenet121. The first classification is done using the original HDCT images. The second classification is done using our hybrid model generated images.

6.4.1 Classification with Original Dataset

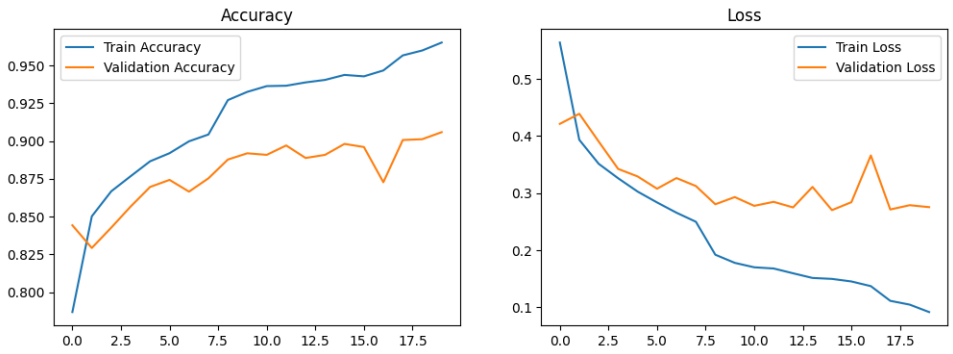


Figure 6.5: Classification Using Original HDCT Images

Class	Precision	Recall	F1-score	Support
Cerebral Infarct	0.90	0.87	0.88	768
Cortical Atrophy	0.80	0.91	0.85	95
Ischemic Changes	0.94	0.93	0.93	265
Lipoma	0.88	0.91	0.90	591
Normal	0.91	0.90	0.91	697
Accuracy			0.90	2416
Macro Avg	0.88	0.90	0.89	2416
Weighted Avg	0.90	0.90	0.90	2416

Table 6.2: Classification report with precision, recall, F1-score, and support per class

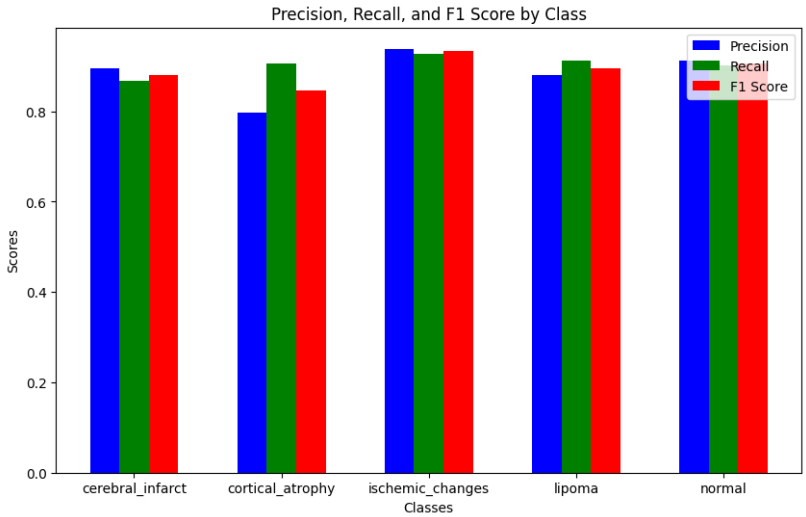


Figure 6.6: Precision, Recall, F1-score by Class

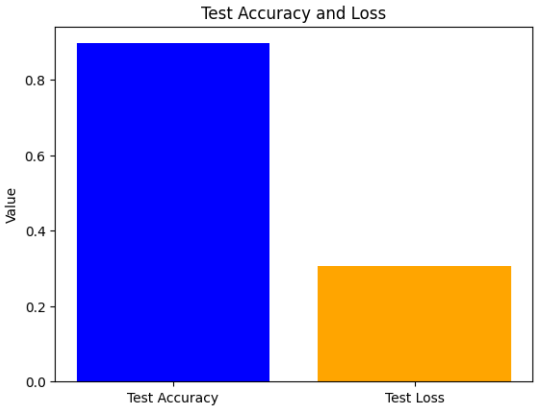


Figure 6.7: Test Accuracy and Loss

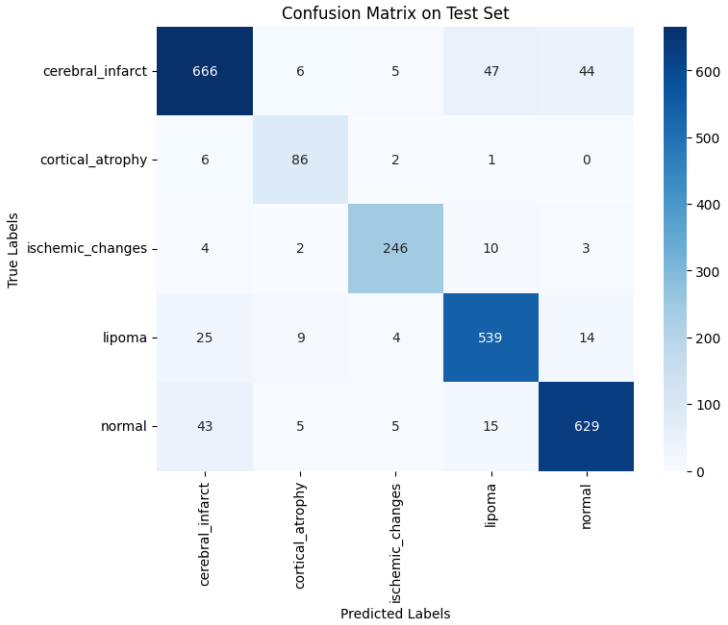


Figure 6.8: Confusion Matrix on Test Set

6.4.2 Classification with Hybrid Model Generated Dataset

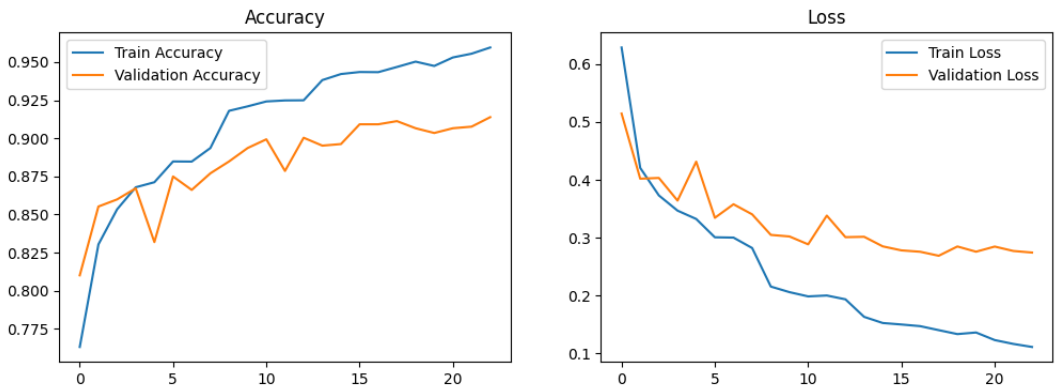


Figure 6.9: Classification Using Hybrid Model Generated Images

Class	Precision	Recall	F1-score	Support
Cerebral Infarct	0.85	0.86	0.85	84
Cortical Atrophy	0.91	0.88	0.90	705
Ischemic Changes	0.93	0.93	0.93	273
Lipoma	0.91	0.93	0.92	640
Normal	0.92	0.93	0.93	707
Accuracy			0.91	2409
Macro Avg	0.90	0.91	0.91	2409
Weighted Avg	0.91	0.91	0.91	2409

Table 6.3: Updated Classification report with precision, recall, F1-score, and support per class

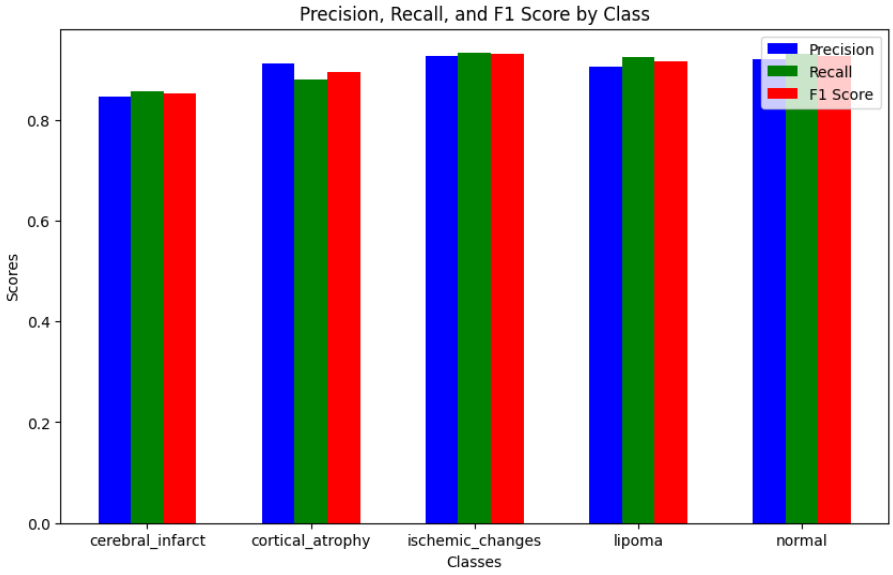


Figure 6.10: Precision, Recall, F1-score by Class

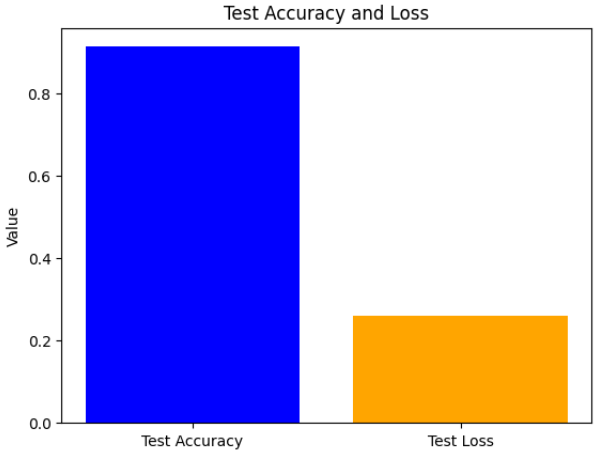


Figure 6.11: Test Accuracy and Loss

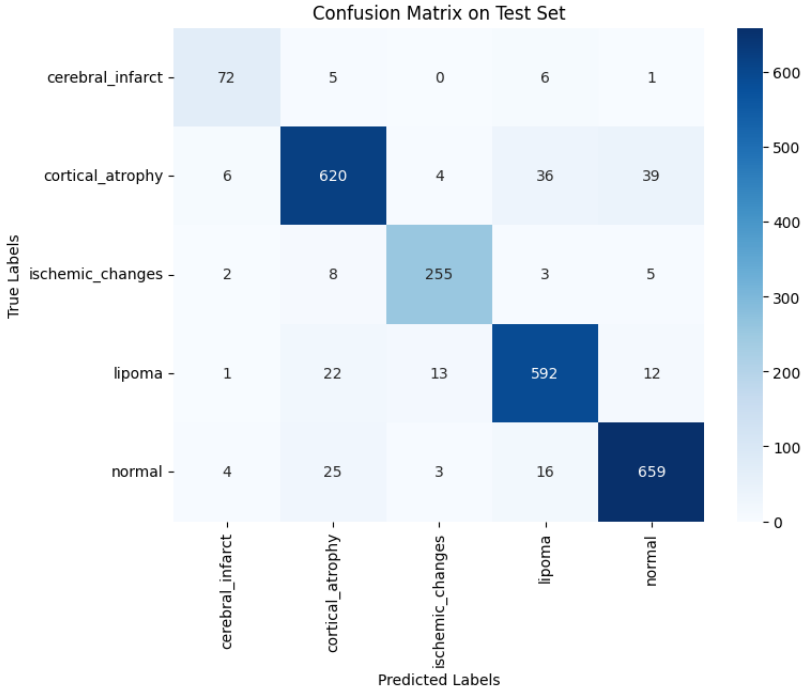


Figure 6.12: Confusion Matrix on Test Set

6.4.3 Comparison and Analysis

In the train and validation accuracy-loss graph of classification 2, It has slightly better validation accuracy and less noisy validation loss. It also has more consistent validation performance. Validation loss is smoother and more stable, suggesting better generalization and less overfitting.

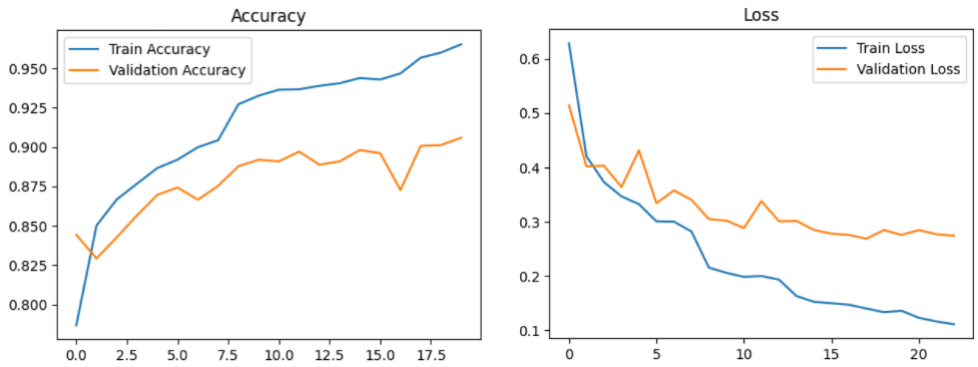


Figure 6.13: Classification 1 & 2

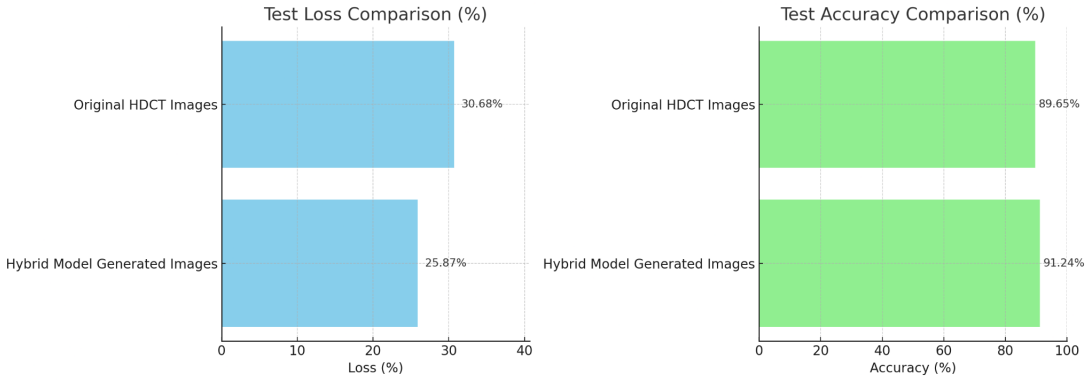


Figure 6.14: Test Loss and Test Accuracy Comparison

Figure 6.15: Number of Misclassified Images

Chapter 7

Conclusion

7.1 Contributions

A key contribution of our work is the acquisition of a unique and clinically sourced dataset of brain CT images that we collected from multiple hospitals. Unlike commonly used publicly available CT image datasets, ours is first hand collected consisting of good quality high dose brain CT images, making it both primary and novel. The labeled portion of our data includes radiologist verified diagnostic reports, while the remaining unlabeled data was effectively utilized through strategies that support semi-supervised learning. This approach allowed us to fully leverage all available data, enhancing the realism and applicability of our training process while maintaining the integrity of our dataset.

This study introduces a framework for simulating low-dose CT images by learning and injecting noise representations in the latent space of a convolutional autoencoder. As mentioned earlier, our proposed pipeline for CT image denoising using paired data required us to collect both high dose CT images and their corresponding low dose CT images. However, due to lack of availability of paired CT data, we leaned towards synthetically generating low dose CT images by utilizing the noises present in our own original high dose CT images. Unlike traditional denoising or noise simulation methods that operate in the pixel space, our work trains an autoencoder specifically on noise-only images to learn compact latent representations of noise. First of all, we used a convolutional autoencoder, but instead of training it to clean up images like typical denoising tasks, we trained it purely on noise images. This allowed the model to learn how noise looks when encoded into a latent space. After we acquired the latent representations of the noises, we injected Gaussian and Poisson noise into the latent space to replicate realistic CT image degradation where noise arises both from electronic signal distortion (Gaussian) and quantum photon variations (Poisson). The decoder then maps synthetic latent noise back to realistic image-space noise that we can apply on our original high dose images to degrade their quality to replicate their corresponding low dose version. Our key contribution in this framework lies in the latent-space modeling of noise which allows us to rely on noise-only images and help us avoid the need for paired clean-noisy datasets based training. This method of latent-space modeling also allows for controlled and measurable degradation of original images and to create different variations of synthetic low dose CT images because we can control how much additional noise

we inject into the latent space. Our proposed method leverages a custom trained model rather than relying on pre-trained networks, thus preserving the integrity of the dataset and making it lightweight and domain-specific. Therefore, this method can be a practical tool for generating synthetic and reusable noise patterns and its application in scenarios where collecting real low-dose CT scans is impractical or difficult.

Lastly, our research proposes a robust and novel approach in terms of denoising low dose CT (LDCT) images based on ESRGAN that we customized by introducing a hybrid architecture. Our key contribution towards this deep learning network includes integrating multiple advanced components like Residual Dense Blocks (RDBs), spectral-normalized PatchGAN discriminator and a custom-configured Multi-Head Self-Attention mechanism. To begin with, our core contribution is introducing a self-attention layer into our proposed GAN. While many existing works either overlook the attention layer or indiscriminately insert full-scale attention across the network, we introduced a memory-efficient, spatially-downsampled MHSA layer after the RDB stack. This design not only reduces computational overhead but also retains contextual understanding which allows the network to model long-range dependencies that is particularly beneficial in distinguishing real anatomical features from noise. Additionally, we modified the discriminator by replacing the conventional full-image discriminator with a PatchGAN-style discriminator. While traditional adversarial frameworks learn global image realism, our discriminator focuses on smaller patches which improves the model's capacity to capture local texture patterns and small-scale anatomical details essential in medical diagnosis. This design also makes training more stable by preventing the model from overfitting. Since medical datasets are usually small, we added spectral normalization to every convolution layer. This keeps the learning process smooth and avoids sudden spikes or unstable gradients during training. We also improved the way our model learns by designing a custom loss function that balances several goals at once. Our generator is trained using four types of losses, among them the L1 loss to ensure the output matches the original pixel-wise, SSIM loss to preserve structure, perceptual loss using VGG16 to keep important features and adversarial loss to sharpen textures. We also introduced a custom training system with several features to keep the training stable and efficient. For instance, we used gradient clipping to avoid sudden spikes during learning, especially helpful when using perceptual loss and attention layers. Also, we added label smoothing, which gives the discriminator softer targets helping it learn better and avoid becoming overconfident. Lastly, we put our proposed model through rigorous training, trials and errors and hyperparameter tuning with a view to achieve optimal performance and desired results. To conclude, LDCT images often suffer from high noise, blurred structures, and loss of fine details. Traditional denoising models struggle with over-smoothing or generating unrealistic textures. GAN-based methods face training instability, especially with small medical datasets. Existing models also lack efficient attention mechanisms and proper handling of medical images. Our attempt towards architectural innovation, loss formulation, training stabilization, attention integration and discriminator reconfiguration promises solutions to these problems.

7.2 Limitations

While working on this research, we encountered several limitations that affected the results of this project. To start with, we anticipated collecting at least 20,000 images but could only gather 12,000 images. On top of that, the dataset lacked diversity since many of the slices came from the same patient. Even though few of the patients had 9-100 images approximately, most had 500-600 images reducing the variability of data. Furthermore, we focused only on brain CT images, and the class distribution of our data was highly skewed. To give an example, "normal" class consisted of the highest number of images, whereas "ischemic changes" had only nine in total. This likely impacted the training and made the model biased towards certain classes. In terms of hardware, despite using RTX4090 GPU we ran into memory issues with the attention mechanism at 256×256 image resolution. To work around this, we had to downsample each image before attention was applied, which may have affected spatial details of the images. Since training time was a concern for us and we were limited by memory constraints, we were unable to experiment with deeper attention modules. Additionally, while we were in the process of acquiring CT images, most hospitals declined to share data, and even those who did were unable to provide corresponding diagnostic labels. We were able to source data from only a few hospitals in Dhaka, but if we were able to collect data from all across Bangladesh, this could have significantly increased the variability of the dataset.

7.3 Future Work

While our clinically sourced novel dataset with proposed frameworks promises an effective approach for what we wanted to achieve, it also provides a strong foundation for further research, and several promising avenues remain open for future exploration. To begin with, a potential path in the future could be expanding our dataset. For instance, our current dataset consists of a moderate number of CT images, however, by collaborating with additional clinical partners in future we might be able to acquire a large dataset. Additionally, we can acquire images of other anatomical regions other than the brain, such as the chest or abdomen that can perhaps help us develop models that are more generalizable and robust across a wider range of clinical scenarios. Another possible area of future research could include longitudinal studies that would require us to collect follow-up scans of the same patients over time, allowing us to conduct clinical studies where monitoring over time is essential. On the other hand, improving data labeling could also help us fine tune our research in the future. While a portion of our current dataset is labeled with diagnostic reports, increasing the number of annotated data would allow us for stronger supervision and more accurate model evaluation. Moreover, since our dataset contains both labeled and unlabeled CT images, we can expand our work in future by replacing the current semi-supervised learning approach with more advanced self-supervised or unsupervised techniques that would enhance performance in tasks like classification. Our proposed framework for simulating low dose CT images using latent space noise modeling is promising but the noise injection process itself can also be refined. For instance, future models could learn optimal noise distributions directly from real low-dose scans, allowing for more accurate simulation. This could involve adversarial training (GAN based architectures) where a discrim-

inator ensures that synthetic noise mimics real low-dose characteristics. Moreover, the framework could be extended into a two-stage training pipeline: first simulating low-dose images using our current approach, and then training denoising or diagnostic models using these simulated images. This would create a fully synthetic data generation and utilization pipeline, especially valuable in data-scarce environments. future studies could involve clinical collaborations to validate the realism and diagnostic utility of the synthetic low-dose images. Expert radiologists could evaluate whether the simulated noise is representative of actual low-dose conditions and whether the degradation affects diagnosis or interpretation accuracy.

Furthermore, while our proposed hybrid GAN based deep learning model with self-attention and PatchGAN discriminator has demonstrated promising results in denoising LDCT images, there is scope for future enhancements, improvements, and extensions. One possible area for future work involves experimenting with the self attention layer implementation. Furthermore, our future work will explore other GAN architectures e.g. StyleGAN, CycleGAN, conditional GAN's etc which may turn out to be more compatible with our work. In our proposed architecture, we introduced a single memory-efficient, downsampled MHSA layer after the RDB stacks, however, in future we could try experimenting with stacked attention layers or attention at multiple stages and increased number of heads and embedding dimensions that was not possible due to time or resource constraints. Additionally , in future we could experiment with deeper or more flexible structures like residual-in-residual dense blocks (RRDBs), residual hybrid attention blocks or attention-guided skip connections instead of using a simplified generator with Residual Dense Blocks (RDBs) only. This would allow us to evaluate the model's trade-off between performance and memory usage compared to our current architecture. Moreover , our future research could explore multi-scale discriminators or dual-discriminator systems, where one discriminator focuses on global consistency and another on local textures. Our current PatchGAN base discriminator has proven effective in our current work but combining it with a secondary global discriminator could allow the model to better balance high-level anatomical realism with low-level texture accuracy. Another possible goal for our future work can be extensive generalization and validation across multiple datasets. Our current model has been trained and tested on a fixed pair of LDCT–HDCT dataset, however, we can evaluate our model on external validation datasets to apply domain adaptation techniques by using a mix of public and private datasets to improve generalization. Also, evaluating model performance on real-world LDCT images with no ground truth and using radiologist feedback will provide insights and ensure clinical usability. Another exciting path for future work could include adapting our model to support 3D volumetric data rather than 2D slices. 3D CT scans promise better continuity between slices, therefore, future versions of our model may incorporate 3D convolutional layers, 3D self-attention blocks, and other smarter methods to use our model for 3D data. In conclusion, although our model lays a solid base for denoising low dose CT images, we aim to improve in terms of architecture, domain adaptation capability and real world usability. Another research direction could be experimenting with state of the art diffusion models. If trained on a large corpus, their probabilistic nature could allow for more realistic image simulations.

7.4 Social Impact

Our current research has various impacts on society as a whole. Our upscaling work for CT images of the brain, by removing noise and enhancing their resolution and features has resulted in social impacts for both Bangladesh and the rest of the world. Those impacts are:

- Upscaled CT images can vastly assist doctors to detect tumors, hemorrhages etc more accurately due to our feature enhancement. Due to how the CT scan imaging works, the machine has a limit on how much detail that they can extract from a single subject. Thus using AI upscaling, it can give more clarity to the often less featured images.

- Repeated usage of CT scans can be reduced due to upscaling. If a patient gets repeated exposure to high dose x-rays, it can be very detrimental for them. Thus our model can lower the amount of CT scans the patient must take.

- Bangladesh is a country where the vast majority of the population are living below the poverty line. Thus many rural clinics are mainly equipped with older or lower-quality CT machines. Using AI upscaling, it can give them better CT scan images without purchasing more expensive equipment.

- Since in Bangladesh, there is no universal healthcare and the more advanced CT scanning can be expensive, it lessens the burden of the patients who are not financially stable enough.

Bibliography

[1] D. J. Brenner and E. J. Hall, "Computed tomography—an increasing source of radiation exposure," *New England journal of medicine*, vol. 357, no. 22, pp. 2277–2284, 2007.

[2] N. L. S. T. R. Team, "Reduced lung-cancer mortality with low-dose computed tomographic screening," *New England Journal of Medicine*, vol. 365, no. 5, pp. 395–409, 2011.

[3] J. Montoya, L. Eckel, D. DeLone, *et al.*, "Low-dose ct for craniosynostosis: Preserving diagnostic benefit with substantial radiation dose reduction," *American Journal of Neuroradiology*, vol. 38, no. 4, pp. 672–677, 2017.

[4] Y. Gu, Z. Zeng, H. Chen, *et al.*, "Medsrgan: Medical images super-resolution using generative adversarial networks," *Multimedia Tools and Applications*, vol. 79, Aug. 2020. DOI: 10.1007/s11042-020-08980-w.

[5] R. Gupta, A. Sharma, and A. Kumar, "Super-resolution using gans for medical imaging," *Procedia Computer Science*, vol. 173, pp. 28–35, 2020, International Conference on Smart Sustainable Intelligent Computing and Applications under ICITETM2020, ISSN: 1877-0509. DOI: https://doi.org/10.1016/j.procs.2020.06.005. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050920315076.

[6] C. Tan, M. Yang, Z. You, H. Chen, and Y. Zhang, "A selective kernel-based cycle-consistent generative adversarial network for unpaired low-dose ct denoising," *Precision Clinical Medicine*, vol. 5, no. 2, pbac011, 2022.

[7] N. Baghel, S. R. Dubey, and S. K. Singh, *Srtransgan: Image super-resolution using transformer based generative adversarial network*, 2023. arXiv: 2312.01999 [eess.IV]. [Online]. Available: https://arxiv.org/abs/2312.01999.

[8] M. Li, J. Wang, Y. Chen, *et al.*, "Low-dose ct image synthesis for domain adaptation imaging using a generative adversarial network with noise encoding transfer learning," *IEEE Transactions on Medical Imaging*, vol. 42, no. 9, pp. 2616–2630, 2023. DOI: 10.1109/TMI.2023.3261822.

[9] S. Majumder, S. Roy, A. Bose, and I. R. Chowdhury, "Understanding regional disparities in healthcare quality and accessibility in west bengal, india: A multivariate analysis," *Regional Science Policy & Practice*, vol. 15, no. 5, pp. 1086–1114, 2023.

[10] M. Meng, Y. Wang, M. Zhu, *et al.*, "Ddt-net: Dose-agnostic dual-task transfer network for simultaneous low-dose ct denoising and simulation," *IEEE Journal of Biomedical and Health Informatics*, 2024.

[11] N. Saidulu and P. R. Muduli, "Asymmetric convolution-based gan framework for low-dose ct image denoising," *Computers in Biology and Medicine*, vol. 190, p. 109 965, 2025, ISSN: 0010-4825. DOI: https://doi.org/10.1016/j.compbiomed.2025.109965. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482525003166.