M.Tech Dissertation Preliminary Report

titled on

GENERATIVE ADVERSARIAL NETWORKS-BASED DATA AUGMENTATION FOR ROBUST DEEP LEARNING MODELS IN MEDICAL IMAGES

Submitted in partial fulfilment towards the award of the degree of

MASTERS OF TECHNOLOGY

in

Computer Science and Engineering

by

Tejal Rohidas Khade

P23CS019

Supervisor

Dr. Chandra Prakash

2024 – 2025

Department of Computer Science And Engineering

SARDAR VALLABHBHAI NATIONAL INSTITUTE OF TECHNOLOGY, SURAT

SARDAR VALLABHBHAI NATIONAL INSTITUTE OF TECHNOLOGY, SURAT

DECLARATION

I hereby declare that the work being presented in this dissertation preliminary entitled "Generative Adversarial Networks-Based Data Augmentation for Robust Deep Learning Models In Medical Images" by me i.e. Tejal Rohidas Khade, bearing Roll No: P23CS019 and submitted to the Computer Science And Engineering Department at Sardar Vallabhbhai National Institute of Technology, Surat; is an authentic record of my own work carried out during the period of July 2024 to December 2024 under the supervision of Dr. Chandra Prakash. The matter presented in this report has not been submitted by me to any other University/Institute for any cause.

Neither the source code there in nor the content of the project report have been copied or downloaded from any other source. I understand that my result grades would be revoked if later it is found to be so.

____

(Tejal Rohidas Khade)

CERTIFICATE

This is to certify that the dissertation preliminary report entitled "Generative Adversarial Networks-Based Data Augmentation for Robust Deep Learning Models In Medical Images", prepared and presented by Tejal Rohidas Khade, bearing Admn. No: P23CS019 of MTech. - II, Semester - III in Computer Science and Engineering, at Department of Computer Science and Engineering of the Sardar Vallabhbhai National Institute of Technology, Surat is satisfactory.

Certified By

Dr. Chandra Prakash
Assistant Professor,
Department of Computer
Science and Engineering,
Sardar Vallabhbhai National
Institute of Technology,
Surat – 395007, Gujarat
India

Jury's Signature

PG Incharge,
M.Tech in CSE
SVNIT - Surat

Head,
Department of Computer
Science and Engineering,
SVNIT - Surat

Acknowledgments

I am grateful to Dr. Chandra Prakash, who has been a great advisor from the very beginning. I am thankful to him for his valuable discussions and the numerous contributions that he has provided to this work. Without his support and guidance, this work could not have been accomplished. Besides my advisor, I would like to thank my research progress committee members for their encouragement, insightful comments, and suggestions.

I want to thank Dr. Mukesh A. Zaveri, Head of Computer Science and Engineering, SVNIT, Surat, for allowing me to explore research aspects of deep learning and providing infrastructural facilities for my work. I want to thank all the faculties and staff members of the Computer Science and Engineering Department, SVNIT, Surat.

Tejal Rohidas Khade
P23CS019

Abstract

Generative Adversarial Networks (GANs) have become increasingly significant across various fields, including medical imaging, due to their capability to produce high-quality synthetic data. GANs have been employed in tasks such as image classification, segmentation, detection, denoising, and reconstruction, contributing to disease diagnosis. Among the severe medical complications is Diabetic Foot Ulcer (DFU), a condition caused by diabetes that can lead to limb amputation if not addressed promptly. Early and accurate detection and classification of DFUs are crucial to ensuring timely treatment and preventing severe outcomes. However, medical image datasets often face challenges such as limited data and imbalanced class distributions, which can hinder the performance of deep learning models in DFU classification. GANs offer a robust solution to these challenges by enhancing the quality and diversity of training datasets. Recent advancements in GAN variants have further improved data augmentation techniques. In addressing the DFU classification problem, we propose a GAN-based data augmentation strategy that generates realistic and diverse DFU images, coupled with Convolutional Neural Networks (CNNs) for classification. This approach aims to overcome dataset limitations, capturing the complex patterns and variations associated with different ulcer stages. By enriching the dataset, our method improves the model's classification performance in terms of accuracy, precision, recall, and F1 score. The proposed framework effectively preserves the realism and distribution of medical images, offering a promising solution for DFU classification and early diagnosis.

Keywords: *Generative Adversarial Networks, GANs, Diabetic Foot Ulcer, DFU classification, medical imaging, data augmentation, deep learning, Convolutional Neural Networks, CNNs, synthetic data, imbalanced datasets*

Table of Contents

List of Figures

List of Acronyms

GAN Generative Adversarial Networks

CNN Convolutional Neural Networks

DFU Diabetic Foot Ulcer

AUC Area Under Curve

CGAN Conditional Generative Adversarial Networks

DCGAN Deep Convolutional Generative Adversarial Networks

ACGAN Auxiliary Classifier Generative Adversarial Networks

WGAN Wasserstein Generative Adversarial Networks

Chapter 1

Introduction

Generative Adversarial Networks (GANs), introduced by Ian Goodfellow in 2014, have revolutionized neural network architectures, enabling two competing networks—Generator and Discriminator—to iteratively improve performance. Yann LeCun described GANs as one of the most impactful ideas in machine learning. GANs are widely used for self-supervised data augmentation, generating high-quality, diverse synthetic data. Variants like ACGAN and InfoGAN enhance data quality by leveraging class structures and attributes.Although GANs have shown promising results, they still face challenges such as goal inconsistency and difficulties in simultaneously achieving both high quality and diversity in generated data, pointing to the need for further improvements. Nonetheless, Generative Adversarial Networks (GANs) have become valuable tools in medical imaging, particularly for addressing issues like limited and imbalanced datasets. By generating realistic, high-quality synthetic data, GANs improve the diversity of training datasets, which is essential for enhancing the performance of deep learning models. In diabetic foot ulcer (DFU) classification, GANs help capture intricate patterns and variations across different stages of ulcers, resulting in improved model generalization. GANs can also be used in other medical use cases like brain tumor detection,

1.1 Need of Generative Adversarial Networks

We need GANs because traditional augmentation techniques, such as rotation, flipping, translation, Gaussian noise, and color jitter, rely solely on variations of the original dataset, which limits the diversity of generated data. These methods are constrained in addressing the challenge of imbalanced datasets, particularly for underrepresented classes, and may fail to capture the complex variations necessary for robust model training. In medical applications, GANs have shown promise for addressing data insufficiency and imbalance, particularly in diabetic foot ulcer (DFU) classification. DFUs, a severe complication of diabetes, can lead to limb amputation if untreated. Effective diagnosis is critical but hindered by limited and imbalanced datasets. This paper proposes a GAN-based augmentation strategy to generate diverse and realistic DFU images, integrated with Convolutional Neural Networks (CNNs) for accurate classification. By enriching datasets and preserving medical image realism, this approach aims to improve classification performance, offering a robust solution for early DFU detection and treatment.

1.2 Types of Generative Adversarial Networks(GANs)

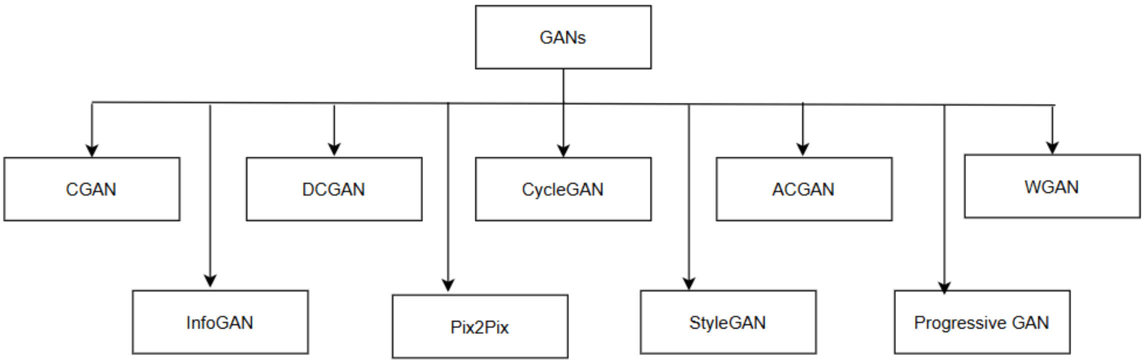


Figure 1.1: Types of GAN Models

The fig 1.1, [5] represents a hierarchical categorization of various types of Generative Adversarial Networks (GANs), showcasing the diversity of GAN architectures designed for medical imaging tasks.

1. CGAN (Conditional GAN): Introduces conditional information, like class labels, to both the generator and discriminator, helping generate more specific features and tackle class imbalance.

2. DCGAN (Deep Convolutional GAN): Uses a deep convolutional network architecture for both the generator and discriminator, enhancing the synthesis of high-quality images.

3. CycleGAN: Utilizes two GANs to enable image-to-image translation between two domains without requiring paired data. It ensures that transformations are reversible by incorporating a cycle-consistency loss.

4. ACGAN (Auxiliary Classifier GAN): Extends Conditional GAN by enabling the discriminator to classify images based on their class labels in addition to determining their authenticity.

5. WGAN (Wasserstein GAN): Implements the Wasserstein loss function to improve training stability and quality of generated samples.

6. InfoGAN: Introduces latent codes that allow the generation of interpretable and meaningful image variations, such as rotation or size.

7. Pix2Pix: A specialized form of Conditional GAN used for paired image-to-image translation tasks, where the discriminator ensures the generated image corresponds to the input image.

8. StyleGAN: Enables control over fine-grained features in the generated images, such as textures and styles, without affecting other aspects of the image.

9. Progressive GAN (Progressive Growing GAN): Increases the complexity of the generator and discriminator gradually during training, enabling the network to learn finer details over time.

1.3 Applications of GAN

Generative Adversarial Networks (GANs) have emerged as powerful tools in deep learning, enabling innovative solutions across various domains.

1. **Medical Imaging:** GANs have revolutionized medical image analysis by addressing challenges such as data scarcity, class imbalance, and noise reduction.

2. **Image Synthesis and Data Generation:**GANs are widely used to create high-quality synthetic images for training and simulation purposes.

3. **Video and Audio Applications:**GANs are used to create, enhance, and modify video and audio content.

4. **Text-to-Image Synthesis:**GANs like DALL-E and others can convert textual descriptions into high-quality images, aiding industries such as advertising, education, and content creation.

5. **Gaming and Virtual Reality (VR):** GANs are used to generate realistic textures, environments, and characters in gaming.

6. **Security and Surveillance:** GANs help improve facial recognition accuracy by generating diverse facial datasets.

7. **Education and Training:** Creating educational content, such as synthetic medical images for training radiologists. Developing virtual labs and simulations for STEM education.

1.4 Motivation

The application of Generative Adversarial Networks (GANs) in medical image analysis has gained significant attention due to their ability to generate high-quality and realistic synthetic data. This capability addresses several critical challenges inherent to medical image datasets, including scarcity, imbalanced class distribution, and diversity in representations. These challenges are particularly pronounced in the case of Diabetic Foot Ulcer (DFU) classification, which often suffers from limited and imbalanced datasets due to the difficulty in collecting and annotating medical images.

Chronic conditions, such as diabetic complications, pose significant challenges to healthcare systems due to their potential for severe outcomes if not managed promptly. Medical imaging plays a critical role in early detection and classification of such conditions, enabling timely intervention and reducing the risk of serious complications. However, delayed identification or misclassification can lead to severe consequences, including prolonged hospital stays, increased costs, and adverse patient outcomes. Accurate and early diagnosis is essential for effective treatment, highlighting the need for robust imaging solutions to address these challenges.

The combination of deep learning models, especially Generative Adversarial Networks (GANs), with deep neural networks for precise classification of diabetic foot ulcer (DFU) images presents a promising solution to overcome existing challenges. This research introduces a novel approach that utilizes the advantages of GANs to improve the accuracy and effectiveness of DFU image classification, enabling timely interventions and better management of the condition. This approach has the potential to transform diabetic foot care, significantly enhancing early detection and treatment of DFUs.

1.5 Problem Description

Medical image analysis plays a critical role in diagnosing and treating various conditions, where accurate classification is essential for effective clinical decision-making. However, the scarcity of annotated medical images and imbalanced class distribution in available datasets present significant challenges to the performance and reliability of deep learning models in this domain. This research explores the use of GAN-powered data augmentation to address these issues. By generating high-quality and diverse synthetic medical data, GANs enhance training datasets, improving the robustness and accuracy of deep learning models. For instance, in the context of diabetic foot ulcers (DFUs), GANs can mitigate data insufficiency and imbalance, facilitating early diagnosis and effective treatment. The proposed approach integrates GAN-based data augmentation with deep learning models, aiming to develop reliable and effective solutions for medical image classification. This research seeks to advance the field of medical image analysis by leveraging GANs to build robust models better equipped to handle real-world medical datasets.

1.6 Objectives

In this research work, we address critical challenges in medical image processing by leveraging Generative Adversarial Networks (GANs) for data augmentation, specifically tailored for Diabetic Foot Ulcer (DFU) classification. GANs are employed to generate high-quality and diverse synthetic images, tackling issues of data insufficiency and imbalanced class distributions that hinder the performance of deep learning (DL) models. Our approach integrates the GAN-augmented datasets with deep neural networks, resulting in significant improvements in model performance metrics. By automating the generation of realistic and diverse DFU images, we reduce the reliance on manual annotation, which is time-consuming and inconsistent across imaging modalities. This research will contribute to bridging the gap between limited real-world data availability and the need for high-performing DL models, addressing broader challenges such as the lack of labeled data, the rarity of certain disorders, and the need for secure data sharing while preserving patient confidentiality. Through extensive evaluation, we demonstrate the potential of GAN-powered augmentation to enhance dataset diversity and generalizability, providing a robust solution for improving the automation and accuracy of DFU classification systems.

1.7 Report Outline

The structure of this report is as follow. The second chapter provides the theoretical background of the previous approaches that were trying to solve a similar problem. The third chapter describes the our approach. We conclude with future works in fourth chapter.

Chapter 2

Theoretical Background & Literature Survey

2.1 Theoretical Background

Traditional medical image analysis for classification relied on manual processing, which is time-consuming and inconsistent. With advancements in deep learning, automated methods like Convolutional Neural Networks (CNNs) have gained popularity for their ability to learn hierarchical features efficiently. Advanced architectures such as DFUNet, DFU_QUTNet, and hybrid models combining CNNs and Vision Transformers (ViTs) have improved diagnostic accuracy by enhancing feature extraction and focusing on critical regions. However, challenges remain, including small, imbalanced datasets that limit model generalization and a focus on binary tasks over complex multi-class classification. Data augmentation techniques, though helpful, often fall short of addressing these limitations, emphasizing the need for more robust solutions.

2.2 Literature Survey

2.2.1 Literature Review

A detailed literature survey is performed to study various approaches proposed by different researchers.

In [1], a deep neural architecture is proposed for the classification of diabetic foot ulcers (DFUs), addressing the unique challenges posed by DFU imaging data. The architecture integrates convolutional neural networks (CNNs) with residual blocks and feature fusion layers to enhance feature extraction and representation. The methodology includes data augmentation techniques such as rotation, flipping, Gaussian noise, shearing, and translation to improve model robustness. The model achieves remarkable results, with an accuracy of 98.87%, precision of 99.01%, recall of 98.73%, F1-score of 98.86%, and an AUC-ROC of 98.13% on benchmark datasets, including DFU2020 and MICCAI DFUs. However, limitations include the need for more diverse datasets, hyperparameter fine-tuning, and evaluation across different clinical scenarios to improve real-world applicability.

In [2], the paper introduces two novel frameworks based on conditional Generative Adversarial Networks (cGANs) to enhance the segmentation accuracy of brain tumors in magnetic resonance imaging (MRI). Enhancement and Segmentation GAN (ESGAN), Enhancement GAN (EnhGAN) are used here which employ Probability Density Function (PDF) transformation blocks to modify intensity histograms and ensure separability between classes. They used BraTS'13 and BraTS'18 datasets. BraTS'13 dataset contains 30 training cases and 10 testing cases. BraTS'18 dataset includes 285 training cases and 66 leaderboard testing cases from 19 institutions. ESGAN achieves competitive performance on small datasets, EnhGAN improves segmentation results for complex tumor regions. Despite certain limitations, both models effectively demonstrate the utility of cGANs in enhancing tumor segmentation and provide a strong foundation for further improvements in MRI-based medical image analysis.

The study in [3] emphasizes the significance of high-quality and diverse training data for deep learning applications, while also addressing the difficulties in acquiring enough realistic data. It introduces CLS-R GAN, a Classification-Reinforced GAN, designed to improve both the quality and diversity of augmented data using a novel self-training framework. In CLS-R GAN, an independent classifier helps guide the generator to self-train by classifying fake data and enhancing real data in an unsupervised way. Experiments, including liver ultrasound image augmentation, demonstrated the effectiveness of CLS-R GAN in enhancing data quality and diversity. Future research plans to apply this method to datasets like CIFAR10, STL10, and ImageNet, with the goal of refining system parameters and improving the selection of fake data for optimization.

In [4], an ensemble of Convolutional Neural Networks (CNN) and Vision Transformers (ViT) was proposed for diabetic foot ulcer (DFU) classification. The study also integrated a Siamese Neural Network (SNN) with a k-Nearest Neighbors (kNN) classifier for enhanced performance. K-Fold cross-validation was applied to ensure the robustness of the model. Data augmentation techniques were used to address dataset imbalance, improving generalization. The DFUC2021 Challenge dataset was used, which included classes such as "None" (no visible DFU), "Infection" (presence of bacterial or fungal infection), "Ischemia" (reduced blood flow causing tissue damage), and "Both" (co-occurrence of infection and ischemia). The dataset faced class imbalance, and pre-applied data augmentation was used to mitigate this. Despite these advancements, the study noted that it could not explore ensemble modeling or more computationally intensive variations to further enhance results.

In [5], paper systematically reviews the application of Generative Adversarial Networks (GANs) for medical image augmentation. The primary focus is on how GANs address challenges like insufficient datasets and imbalanced class distributions in medical imaging, which are common barriers to training robust deep learning models. The paper examines 52 peer-reviewed studies (2018–2022) to identify popular GAN architectures, Common Medical Modalities and Target Organs, Downstream Tasks like Classification and segmentation of medical images, evaluation metrics like qualitative methods (visual quality of generated images), quantitative direct methods (FID, SSIM), and quantitative indirect methods (performance improvements in tasks using augmented data).

In [6], the study addresses the growing challenge of Diabetic Foot Ulcers (DFUs), particularly those with ischemia and infection, emphasizing the need for early detection. The authors proposed using pre-trained transformer models, fine-tuned on the DFUC-21 dataset, for multi-class DFU classification.A Multi-Model approach was proposed, where features from parallel-trained transformers were fused from the last layers, achieving a macro-average F1-Score of 0.569. Weighted cross-entropy optimization and pairwise feature fusion addressed class imbalance. The results highlight the potential of combining CNNs with transformer architectures for future improvements in DFU classification.

In [7], various advanced models were explored for DFU detection, including EfficientDet, Cascade R-CNN integrated with DetNet, and Faster R-CNN with deformable convolution layers. The study utilized data augmentation techniques such as random rotation and shear transformations to enhance the robustness of the models. Benchmark datasets, including the MS-COCO dataset and the Diabetic Foot Ulcers Grand Challenge (DFUC2020) dataset with designated training, validation, and testing sets, were employed. Despite promising results, challenges remain, such as optimizing CNNs for remote monitoring applications and reducing false positives caused by difficulties in distinguishing ulcers from other skin conditions.

In [8], various techniques for addressing the challenges of diabetic foot ulcer (DFU) classification using the DFUC2021 dataset were performed. Given the dataset's class imbalance, performance evaluation included per-class F1-Score, micro-average F1-Score, and macro-averages of Precision, Recall, F1-Score, and AUC. Pretrained models from ImageNet and data augmentation strategies were employed to enhance model performance. For example, ischaemia images underwent eight augmentation techniques, while infection and ischaemia classes were augmented with three techniques, significantly increasing data samples. Among tested models, DenseNet121 and EfficientNetB0 emerged as top performers. DenseNet121 achieved the highest macro-average AUC of 0.88, while EfficientNetB0 excelled in macro-average Precision, Recall, and F1-Score, particularly improving the infection F1-Score. Analysis with UMAP highlighted EfficientNetB0's ability to enhance intra-class clustering on testing data, although inter-class separation remained challenging. Despite advancements, accurately detecting infection and co-occurrence of ischaemia and infection continues to be difficult, especially in the "both" category.

In [9], study evaluates the performance of multiple deep learning models for diabetic foot ulcer (DFU) detection, using a dataset of 640x480 pixel images annotated with LabelImg and VGG Image Annotator. The models include EfficientDet, which leverages EfficientNet as a backbone and BiFPN for feature fusion, YOLOv5, known for real-time object detection with single-pass processing, and Faster R-CNN, featuring a three-stage architecture for robust region proposals and detection. Data augmentation techniques such as scaling, color adjustments, and mosaic augmentation were applied to enhance diversity. The models were assessed using F1-score and mean average precision (mAP), highlighting EfficientDet's scalability, YOLOv5's speed, and Faster R-CNN's precision in detecting DFUs.

In [10], the study presents the design of four hybrid CNN models for DFU classification, with an empirical comparison of different architectures featuring varying numbers of branches. The models utilized feature aggregation techniques, including a Global Average Pooling layer and fully connected layers with dropout for enhanced classification performance. The final output was determined through a Softmax layer. The dataset used consisted of 754 images of patients' feet, with two categories: abnormal (DFU) and normal (healthy skin). However, the study faced limitations, such as the inability to improve performance by increasing network width, the small size of the dataset, and the model's current focus on only two classes (normal vs. abnormal). Future work includes the application of the model for transfer learning.

In [11], the study explores the use of deep learning, particularly Convolutional Neural Networks (CNNs), for detecting COVID-19 using chest X-rays (CXR). The researchers employed the VGG16 CNN model, which was limited by the availability of a small dataset. To overcome this challenge, they introduced an Auxiliary Classifier Generative Adversarial Network (ACGAN)-based approach called CovidGAN, designed to generate synthetic CXR images. Incorporating these synthetic images into the training process significantly improved the performance of CNNs for COVID-19 detection, increasing classification accuracy from 85% to

95%.

In [12], an ensemble CNN model was compared against handcrafted machine learning algorithms for diabetic foot ulcer (DFU) classification tasks. The study focused on binary classification tasks, specifically differentiating between Ischaemia vs. Non-Ischaemia and Infection vs. Non-Infection. The dataset consisted of 1459 DFU images collected from the Lancashire Teaching Hospitals, with images captured using various camera models, including Kodak DX4530, Nikon D3300, and Nikon COOLPIX P100. Natural data augmentation techniques were employed to enhance the dataset. The study highlighted challenges such as high visual intra-class dissimilarities and inter-class similarities, imbalanced dataset, and limited data quality. Furthermore, there is potential for improving model performance by optimizing hyperparameters for both traditional machine learning algorithms and CNN models. One key limitation noted was that infection in DFU images may not always present clear visual indicators, which complicates the classification task.

In [13], a novel CNN model named DFU_QUTNet was proposed for diabetic foot ulcer (DFU) classification and compared with top-performing CNN architectures like GoogleNet, AlexNet, and VGG16 after fine-tuning. When paired with an SVM classifier, DFU_QUTNet achieved a higher F1-Score of 94.5% compared to the other models. The features extracted by DFU_QUTNet were used to train both SVM and KNN classifiers, with the SVM classifier showing the highest precision, recall, and F1-Score. The study utilized a dataset of 754 foot images from DFU patients, with each image's region of interest (ROI) resized to 224x224. However, the study faced limitations, such as a small dataset of only 754 images, lack of generalization to other tasks like skin cancer classification, and no clinical validation or comparison with expert clinicians.

In [14], the study utilized both Conventional Machine Learning (CML) methods and Convolutional Neural Networks (CNNs) for classifying ulcer and non-ulcer images. The researchers introduced a novel CNN architecture, DFUNet, specifically designed to process input data more effectively and efficiently than other state-of-the-art CNN architectures. Through 10-fold cross-validation, DFUNet achieved an impressive AUC score of 0.961, outperforming all tested machine learning and deep learning classifiers. CNN architectures such as LeNet, AlexNet, and GoogLeNet were also employed to develop a fully automated method for distinguishing DFU-affected skin from normal skin. However, DFUNet demonstrated superior accuracy and sensitivity compared to GoogLeNet and AlexNet. While data augmentation techniques like rotation, flipping, contrast enhancement, color space variations, and random scaling were applied, they did not significantly improve overall performance.

2.2.2 Research Challenges

- Dataset Size and Diversity: Small and non-diverse datasets restrict model generalization to varied real-world scenarios. Limited datasets hinder the ability to learn comprehensive problem representations, leading to suboptimal performance.

- Class Imbalance: Certain classes are underrepresented, causing biased models that struggle to identify rare or infrequent classes accurately.

- Limitations of Traditional Data Augmentation: Techniques like rotation, flipping, and noise addition often fail to significantly improve performance, especially in complex classification tasks. These methods may not adequately capture the diversity required for robust model training.

- Generalization Issues:

  Models perform well on specific datasets but often fail to generalize to unseen or varied clinical and real-world settings.

- Focus on Binary Classification:

  Many studies address binary classification tasks, neglecting the complexities of multi-class or hierarchical classification. This limits the ability to handle nuanced real-world scenarios.

Thus, our proposed solution aims to address data imbalance and exploring multi-class classification to improve model robustness and real-world applicability.

Chapter 3

Implementation Methodology

3.1 Proposed Methodology

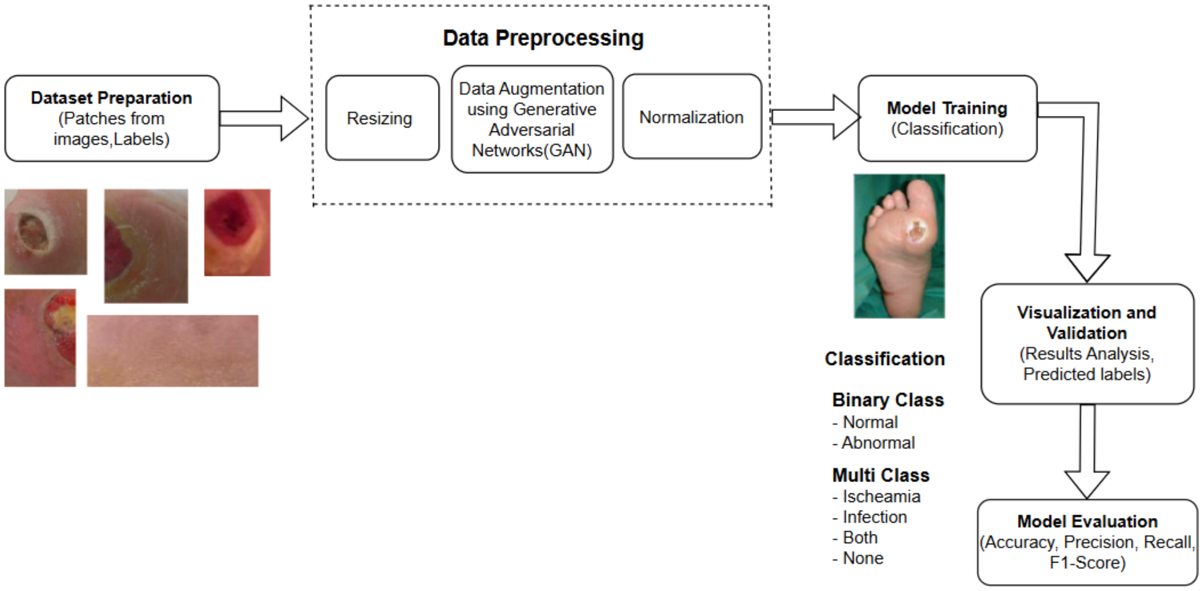


Figure 3.1: Proposed Methodology

This fig 3.1 section consists of five parts: (i) Dataset with samples of diabetic foot ulcers from diverse patients (ii) Labeling process of our dataset into normal skin and abnormal patched skin categories and others (iii) preprocessing of training patches through resizing, normalization and data augmentation using GANs to improve model accuracy (iv) fine-tuning CNN architectures as the base architecture integrated with GAN principles to enhance DFU classification. Our proposed model employs GAN-based data augmentation to address class imbalance and increase data diversity while using its CNN module to extract robust features for accurate classification, achieving improved performance compared to pre-trained models.

- Dataset Preparation: The dataset utilized in this research was obtained (reference number 332) comprises diabetic foot ulcer (DFU) images acquired from Lancashire Teaching Hospitals. To maintain imaging consistency, photographs were taken as close-ups of the entire foot, positioned approximately 30–40 cm away and parallel to the ulcer's plane. Flash photography was avoided to minimize inconsistencies from reflections, and adequate room lighting was employed to ensure uniform image color tones. A podiatrist and a consultant physician specializing in diabetic foot care assisted in retrieving these images from hospital archives.

Table 3.1: DFUC2021 Dataset Description

Data Type	Internal Sub Division	Sample Count	Dimension
DFUC2021_train	Images	5955	224x224
DFUC2021_test	Images	5734	224x224
PartA_DFU_Dataset	Abnormal	1038	Variable
	Normal	641	Variable
PartB_DFU_Dataset	Infection	5890	256x256
	Ischaemia	9870	256x256

- Ground Truth Annotation: The annotation process involved marking the ulcer's location using bounding boxes and labeling each ulcer based on its condition as ischaemia, infection, or none. The software LabelImg was employed for labeling purposes. Initially, the Region of Interest (ROI) was cropped to a standard size of 224×224 pixels, focusing on the significant area surrounding the ulcer, including both normal and abnormal skin tissues. Medical specialists labeled the cropped regions into two categories: normal and abnormal skin patches. The dataset consisted of 1,679 skin patches, with 641 labeled as normal and 1,038 as abnormal (DFU). To ensure annotation accuracy, multiple labels were reviewed by both a podiatrist and a consultant physician. For bounding boxes, an averaging approach was used to create a final, consensus bounding box from multiple annotations. Pathology labels are cross-validated with medical records to ensure consistency and reliability. The dataset was divided into training and testing subsets, with 80% of the skin patches allocated for training and the remaining 20% used for testing. This careful preparation ensures a robust foundation for the development and evaluation of the proposed models.

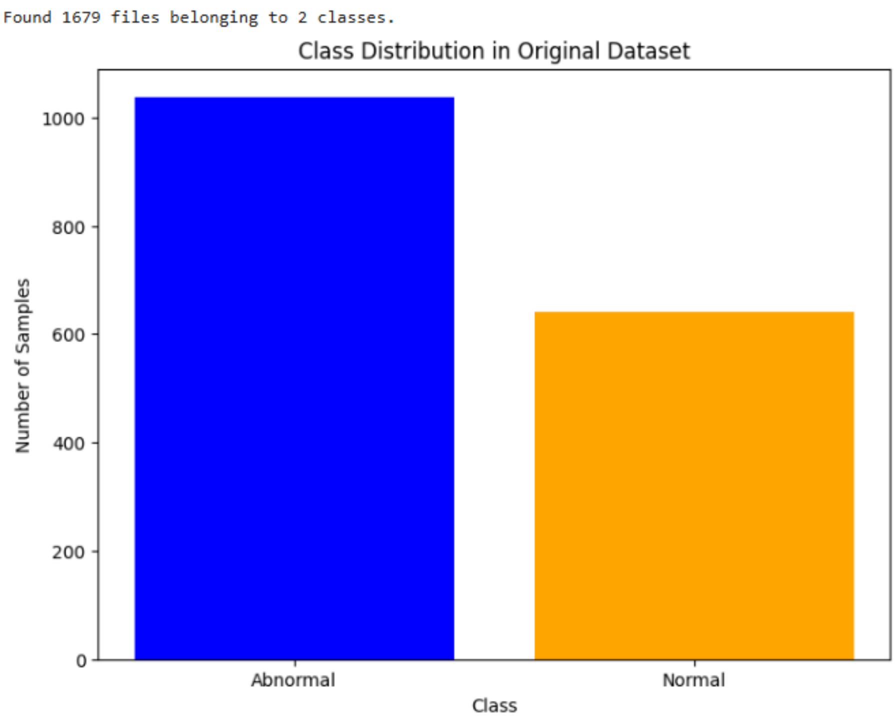
- Data Preprocessing: The images in dataset are DFU patches of normal and abnormal skin which are of varying sizes and need to be converted into same size samples for training data. Thus, it is very important to perform preprocessing on these patches.

Resizing: The first step in the preprocessing pipeline involves resizing all input images to a fixed dimension of 224×224 pixels. This ensures uniformity in the dataset and compatibility with the input requirements of the deep learning models. Additionally, images are center-cropped to retain the most relevant portions of the foot ulcer, focusing on the ulcer and surrounding skin tissue.
Normalization: To prepare the images for efficient model training, pixel values are converted into tensors. These tensor values are normalized to fall within the range $[-1, 1]$, which accelerates model convergence and helps in stabilizing the training process. This normalization step ensures that all input features have a similar scale, reducing the risk of gradient instability during backpropagation.
Data Augmentation: Most studies have employed conventional augmentation techniques such as rotation, flipping, translation, Gaussian noise, shearing, and color jitter. However, these traditional methods are constrained by the quantity of data they can produce and are entirely reliant on the original dataset. In contrast, deep learning approaches, particularly GANs, can generate a diverse range of synthetic data independent of the original dataset. To address these limitations, we utilize Generative Adversarial Networks (GANs) to create synthetic training patches, especially for rare or underrepresented classes in the dataset, which can significantly enhance classification accuracy.

- Model Training :For DFU classification, we employed an 80-20 split for the training and testing datasets. Additionally, 20% of the training set was set aside as a validation set. The DFUC2021 dataset comprises a total of 15,683 images, with 5,955 images allocated for training and 5,734 images reserved for testing. Part A includes 1,038 abnormal images and 641 normal images, while Part B contains 5,890 images of the infection class and 9,870 images of the ischemia class.

- Visualization and Validation: Visualization techniques play a crucial role in understanding model predictions, feature extraction, and dataset properties.Feature Map Visualizations can be used to understand which features the CNN is extracting from input images. Loss Curves from generator and discriminator can be plotted to track adversarial training progress. Generated Samples from the generator can be monitored regularly so as to check for improvements in realism and diversity.

- Model Evaluation :The DFUC2021 dataset exhibits an imbalance in class distribution. To visualize this imbalance, a bar graph was plotted. To address these class imbalances effectively, performance metrics such as per-class F1-Score, micro-average F1-Score, and the area under the Receiver Operating Characteristic (ROC) curve will be reported.



```
Found 1679 files belonging to 2 classes.
```

Figure 3.2: Data Imbalance between Abnormal and Normal Class

3.2 Description of Workflow Design Architecture

3.2.1 Role of GAN in the architecture:

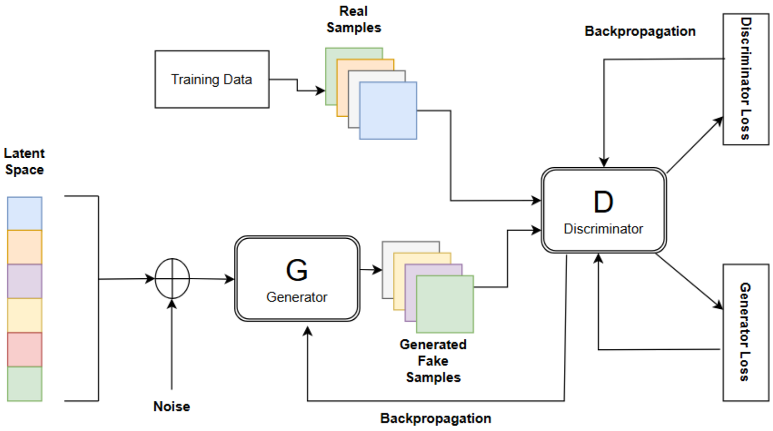


Figure 3.3: Generative Adversarial Network(GAN)

A Generative Adversarial Network (GAN) is a deep learning model comprising two neural networks: the Generator and the Discriminator, trained in a competitive framework. The Generator creates synthetic samples, often referred to as fake data, starting from random noise, while the Discriminator evaluates and differentiates between real and generated data. This iterative process enhances the Generator's ability to produce realistic outputs and improves the Discriminator's capacity to identify fake samples, eventually achieving a balanced state.

- Latent Space: GANs operate using a latent space, which is a representation of random noise serving as the input for the Generator. This noise acts as a foundation for producing synthetic data samples.

- Generator (G): The Generator transforms the noise into realistic outputs, such as images, designed to closely resemble the real data. Its objective is to deceive the Discriminator by generating high-quality samples that are hard to distinguish from actual data. For example, it can generate synthetic DFU images to address imbalances in datasets, particularly in underrepresented categories like infection, ischemia, normal, or abnormal cases.

- Real Samples: Real data samples from the training dataset are provided to the Discriminator as ground truth, helping it compare and evaluate the authenticity of the generated samples.

- Discriminator (D): The Discriminator functions as a classifier, receiving both real samples from the dataset and synthetic samples produced by the Generator. Its role is to distinguish between genuine and fake data by assigning a probability score that indicates the likelihood of a sample being real. It plays a critical role in verifying the authenticity of generated images, ensuring that the augmented dataset closely mirrors the characteristics of the original DFU data.

- Feedback Loop (Loss and Backpropagation):

  1. Generator Loss: When the Discriminator classifies the generated samples as fake, it provides feedback (loss) to the Generator via backpropagation. This feedback updates the Generator's parameters, enhancing its ability to produce more realistic samples. The Generator's loss is determined by the Discriminator's evaluation, rewarding the Generator for successfully deceiving the Discriminator and penalizing it when it fails. The following equation is minimized to training the generator:

$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^{m} \log\left(1 - D\left(G\left(z^{(i)}\right)\right)\right) \tag{3.1}$$

  2. Discriminator Loss: when the Discriminator incorrectly identifies a real sample as fake or a generated sample as real, it receives feedback (loss) to enhance its ability to distinguish between genuine and synthetic data. It adjusts its parameters

by penalizing itself for such misclassifications, striving to maximize the following objective function.

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^{m} \left[\log D\left(x^{(i)}\right) + \log\left(1 - D\left(G\left(z^{(i)}\right)\right)\right)\right] \tag{3.2}$$

where log(D(x)) shows the probability that the generator is properly classifying the real image, maximizing log(1-D(G(z))) will help it to correctly label the fake image that is recieved from the generator.

3. Objective Function of GAN: The GAN framework operates as a min-max optimization game, where the Discriminator aims to maximize its accuracy in distinguishing between real and synthetic data, while the Generator strives to minimize the Discriminator's effectiveness in making this distinction.

   $$\min_G \max_D V(G, D) = \mathbb{E}_x[\log(D(x))] + \mathbb{E}_z[\log(1 - D(G(z)))] \tag{3.3}$$

   The generator task is to minimize this function while the discriminator task is to maximize it.

Both components are trained iteratively and simultaneously, improving each other's performance in a competitive manner.

Training Workflow:

Initially, the Discriminator is trained to differentiate real data from poorly generated fake data. As the Generator improves, it begins creating samples that closely resemble real data, pushing the Discriminator to enhance its ability to distinguish between the two. This iterative process continues until the Generator produces samples that the Discriminator cannot differentiate from real data, achieving a state of equilibrium.

Over time, various GAN architectures have been developed and widely adopted for tasks like medical image synthesis and augmentation. Popular GAN variants include Conditional GAN (cGAN), Deep Convolutional GAN (DCGAN), Cycle-Consistent GAN (CycleGAN), Auxiliary Classifier GAN (ACGAN), Pix2Pix and many more. Among these, studies have identified cGAN as one of the most frequently used architectures for basic augmentation tasks [5]. Although I have not yet experimented with these architectures for my specific use case, the proposed methodology will incorporate one of these well-established architectures for data augmentation. The choice will be based on its ability to generate high-quality synthetic images and address class imbalance challenges in diabetic foot ulcer (DFU) datasets, ultimately improving classification accuracy.

In the below figure, we have displayed the output from generated images using basic GAN architecture when run for few epochs. This represents an instance of output generated after applying GAN on DFU patch from the dataset. Initially, noise is added with the probablity of latent space from real samples and then it gets improved on increasing the number of epochs during training. We still need to process the images for higher epochs to get better quality of image that resembles original patch.

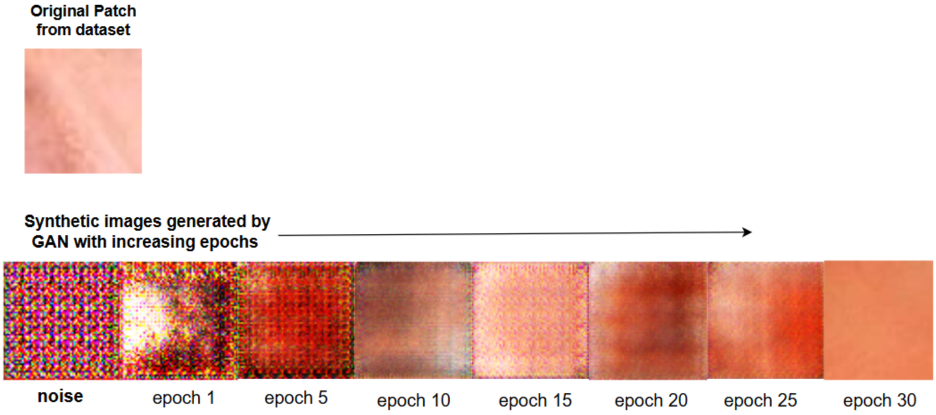

Original Patch from dataset
Synthetic images generated by GAN with increasing epochs
noise
epoch 1
epoch 5
epoch 10
epoch 15
epoch 20
epoch 25
epoch 30

Figure 3.4: Generated Images from GAN

3.2.2 Role of CNN in the architecture:

Convolutional Neural Networks (CNNs) has achieved state-of-the-art performance in medical image generation when trained on sufficient labeled data. However, due to their large number of parameters, CNNs are prone to overfitting on small datasets, making generalization highly dependent on the dataset's size and diversity. This poses a significant challenge in the medical imaging domain, where labeled data is often limited. The Convolutional Neural Network (CNN) acts as the backbone for the classification task, processing the DFU images to extract meaningful features like texture, color variations, shape irregularities with computational efficiency. By incorporating a dense architecture with deep convolutional layers, skip connections, and bottleneck layers, the CNN captures both low-level and high-level features, crucial for distinguishing between normal and abnormal skin patches, infection and ischaemia classes.
The architectural design will incorporate the following:

- Low-Level Feature Extraction: Initial convolutional layers will focus on capturing fundamental features like edges, gradients, and textures. These features are crucial for differentiating healthy skin from ulcer-affected areas based on subtle textural differences.

- High-Level Feature Extraction: Deeper layers in the CNN will extract more complex patterns, such as irregular ulcer boundaries, tissue damage, or signs of infection. This hierarchical feature learning ensures that the network understands both the global structure and local abnormalities present in DFU images.

- Layer Design: Incorporating larger receptive fields in deeper layers and smaller kernels in earlier layers will ensure fine-grained feature extraction.

- Pooling layers: Max-pooling will down-sample spatial dimensions while retaining significant features, reducing computational complexity without losing important information.

- Transfer Learning: Transfer learning will be a pivotal strategy to overcome the challenges of limited training data and accelerate model convergence. State-of-the-art CNN architectures like ResNet, VGG, or EfficientNet, pre-trained on large-scale datasets, will serve as the backbone for feature extraction. These models have already learned generalized image features, which can be fine-tuned to specialize in DFU classification.

- Fine-Tuning: The pre-trained network's initial layers will remain frozen during early training to retain generalized features, while deeper layers will be fine-tuned to adapt to the DFU dataset.This approach reduces overfitting and optimizes the network for domain-specific DFU dataset.

- Evaluation metrics: To evaluate the effectiveness of the CNN architecture, comprehensive performance metrics such as accuracy, precision, recall, ROC-AUC curve.

3.2.3 Integration of GAN and CNN:

The synergy between GAN and CNN is a key aspect of the proposed architecture. While the GAN focuses on improving the quality and quantity of training data, the CNN leverages this augmented dataset to enhance classification performance. This integration ensures:

- Robust feature extraction from diverse and balanced datasets.

- Improved model accuracy and generalization.

- Reduction of overfitting due to enriched training data.

Therefore, we first used GAN architectures for producing high quality Diabetic foot ulcers. Then we present a novel architecture for Diabetic foot ulcers classification using CNN. To enhance classification performance, the GAN is integrated with a Convolutional Neural Network (CNN). We believe that our hybrid architecture proves to be efficient solution for DFU classification and ensures a comprehensive approach to handle challenges like limited data, class imbalance, and complex feature variability.

Chapter 4

Conclusion and Future Scope

4.1 Conclusion

Diabetic foot ulcers (DFUs) remain a critical challenge in diabetes management, requiring timely diagnosis and effective treatment strategies. This study presented a robust approach to DFU classification, integrating advanced deep learning techniques and GAN-based data augmentation to address challenges like class imbalance and limited data diversity. The proposed deep neural network aims to demonstrate improved classification performance by leveraging synthetic images, enabling the extraction of meaningful features for accurate DFU classification. By employing GANs for high-quality data augmentation and leveraging pre-trained CNNs for robust feature extraction, the approach ensures enhanced classification accuracy and generalization. The synergy between GANs and CNNs not only enriches the training dataset but also reduces overfitting, enabling the model to perform effectively on diverse and unseen data.

4.2 Future Scope

Future objectives include the development of a mobile application designed to help in the detection of diabetic foot ulcers (DFUs) from the comfort of a patient's home. This application aims to support patients and their families in identifying and tracking ulcers, thereby minimizing the need for frequent hospital visits. Additionally, the app has the potential to alleviate the burden on healthcare systems by facilitating accessible care for patients in remote areas, ensuring safer and more efficient DFU management. Other goals include expanding the dataset, validating the model in real-world clinical environments, and incorporating it into diagnostic tools for broader use in managing diabetic foot conditions, ultimately improving patient outcomes and supporting clinicians in making more informed decisions.Integrating explainable AI techniques will provide clinicians with insights into the prediction of model, fostering trust and facilitating adoption in clinical practice.

Bibliography

[1] A. V. Nishu Bansal, "Dfootnet: A domain adaptive classification framework for diabetic foot ulcers using dense neural network architecture," *Springer Science+Business Media, LLC, part of Springer Nature*, vol. 16, 2024.

[2] A. L. S. MohammadHamghalama, b, "Medical image synthesis via conditional gans:application to segmenting brain tumours." ELSEVIER, 2024.

[3] S. Kim and S. Lee, "Self-supervised augmentation of quality data based on classification-reinforced gan," in *Proceedings of the 17th International Conference on Ubiquitous Information Management and Communication (IMCOM).* IEEE, 2023, corresponding Author: Sukhan Lee (lsh1@skku.edu).

[4] M. S. A. Toofanee, S. Dowlut, M. Hamroun, K. Tamine, V. Petit, A. K. Duong, and D. Sauveron, "Dfu-siam: A novel diabetic foot ulcer classification with deep learning," *IEEE Access*, 2023, author et al.: Preparation of Papers for IEEE TRANSACTIONS and JOURNALS.

[5] A. Makhlouf, M. Maayah, N. Abughanam, and C. Catal, "The use of generative adversarial networks in medical image augmentation," vol. 35. Springer, 2023.

[6] A. Qayyum, A. Benzinou, M. Mazher, and F. M. deau, "Efficient multi-model vision transformer based on feature fusion for classification of dfuc2021 challenge," in *Proceedings of the DFUC2021 Challenge*. Springer, 2022, dOI: 10.1007/978-3-030-94907-5$_5$.

[7] M. H. Yapa, R. Hachiuma, A. Alavi, R. Brüngeld, B. Cassidy, M. Goyal, H. Zhu, J. Rückert, M. Olshansky, X. Huang, H. Saito, S. Hassanpoure, C. M. Friedrich, D. B. Ascher, A. Song, H. Kajita, D. Gillespie, N. D. Reeves, J. M. Pappachani, C. O'Shea, and E. Frank, "Deep learning in diabetic foot ulcers detection: A comprehensive evaluation," *Computers in Biology and Medicine*, vol. 135, 2021.

[8] M. H. Yap, B. Cassidy, J. M. Pappachan, C. O'Shea, D. Gillespie, and N. D. Reeves, "Analysis towards classification of infection and ischaemia of diabetic foot ulcers," *arXiv:2104.03068v2 [cs.CV]*, 2021, published: 21 June 2021.

[9] B. Cassidy, N. D. Reeves, J. M. Pappachan, D. Gillespie, C. O'Shea, S. Rajbhandari, A. G. Maiya, E. Frank, A. J. M. Boulton, D. G. Armstrong, B. Naja, J. Wu, and M. H. Yap, "Dfuc 2020: Analysis towards diabetic foot ulcer detection," *arXiv:2004.11853v3 [cs.CV]*, 2021, arXiv preprint, 24 May 2021.

[10] L. Alzubaidi, A. A. Abbood, M. A. Fadhel, O. Al-Shamma, and J. Zhang, "Comparison of hybrid convolutional neural networks models for diabetic foot ulcer classification," *Journal of Engineering Science and Technology*, vol. 16, 2021.

[11] A. Waheed, M. Goyal, F. Al-Turjman, D. Gupta, A. Khanna, and P. R. Pinheiro, "Covidgan: Data augmentation using auxiliary classifier gan for improved covid-19 detection," in *Special Section on Emerging Deep Learning Theories and Methods for Biomedical Engineering*. IEEE, 2020.

[12] M. Goyala, N. D. Reeves, S. Rajbhandari, N. Ahmad, C. Wang, and M. H. Yap, "Recognition of ischaemia and infection in diabetic foot ulcers: Dataset and techniques," *Computers in Biology and Medicine*, 2020.

[13] L. Alzubaidi, M. A. Fadhel, S. R. Oleiwi, O. Al-Shamma, and J. Zhang, "Dfu_qutnet: Diabetic foot ulcer classification using novel deep convolutional neural network," *Multimedia Tools and Applications*, 2019.

[14] M. Goyal, N. D. Reeves, A. K. Davison, S. Rajbhandari, J. Spragg, and M. H.

Yap, “Dfunet: Convolutional neural networks for diabetic foot ulcer classification,” *arXiv:1711.10448v2 [cs.CV]*, 2017, published: 10 December 2017.

Appendix A

Introduction

Intel Corporation is a leading American multinational technology company that is widely recognized for its innovation in the semiconductor industry. Founded in 1968 by Robert Noyce and Gordon Moore, Intel is headquartered in Santa Clara, California. The company39;s primary focus is on designing and manufacturing semiconductors, particularly microprocessors, which serve as the central processing units (CPUs) in most personal computers and servers. Intel has been a pioneer in the development of computing technology, playing a critical role in the growth of personal computers and the broader technology landscape. Its x86 architecture and processors, such as the Pentium and Core series, have powered personal computers, servers, and data centers. Intel also advanced Moore&39;s Law, enabling the rapid growth of computing power by doubling transistor density roughly every two years. The company has driven innovation in chipsets, integrated graphics, and memory technologies, like 3D XPoint and SSDs. These advancements have had a profound impact on everything from personal devices to large-scale data infrastructure.

A.1 Internship at Intel, Bangalore

At Intel Bangalore, the internship involved working on different projects and tasks alongside experienced professionals, learning about the company's services and cutting-edge technologies. As part of the NPU AI Frameworks team, I focused on gaining hands-on experience with Intel's Neural Processing Units (NPUs) and their integration with AI frameworks. This included working with tools like OpenVINO, ONNX Runtime, and optimizing inferencing workloads. The internship provided an in-depth understanding of how Intel's hardware accelerates AI computations, helping to improve performance and scalability in real-world AI applications. I also had the opportunity to collaborate with experts in the field, further enhancing my skills in deploying efficient AI solutions. Through collaboration and learning, I deepened my understanding of Intel's role in advancing artificial intelligence, enhancing my skills in deploying scalable, efficient AI models.

A.2 Motivation

The purpose of this internship at Intel Bangalore is to contribute to projects that push the boundaries of AI capabilities while learning from the best minds in the industry. This internship is an invaluable opportunity for me to deepen my knowledge, enhance my skills, and contribute to Intel's mission of advancing AI technologies that will shape the future of computing and society. Also, the aim is to provide valuable learning and development experience to help explore potential career paths and gain valuable industry experience.

A.3 Layout

The structure of this report is as follows: The Task assigned and their details are given in Section 2. The Tools and Technologies learned and used given in section 3. The report is concluded in Section 4.

Appendix B

Project Details

I worked in different domains since I joined Intel. I have done opensource contributions in generative AI, code coverage, bug fixing, testing and validation thereby delivering efficient results.

B.1 Generative AI Workloads - Quantization and Compression of GenAI Models

During my internship in the NPU AI Frameworks team at Intel, I worked on optimizing Generative AI (GenAI) models, specifically Llama and Phi3, focusing on quantization techniques and model compression using Intel's OpenVINO toolkit. The goal was to enhance the inference performance of large language models (LLMs) by converting them into more efficient formats that can run on Intel hardware while maintaining accuracy and reducing computation time.
Workflow of the task-

- Investigating Inference with OpenVINO IR for Phi3 and Llama Models Investigated how to run inference on Phi3 and Llama models using OpenVINO IR (Intermediate Representation). The first step involved understanding the flow of model compression and converting these models from their original framework formats (e.g., PyTorch, TensorFlow) to OpenVINO's IR format. This process required exploring OpenVINO notebooks and utilizing tools like the OpenVINO Model Converter CLI and Python conversion APIs to achieve the model conversion. The primary goal here was to ensure smooth execution of inference on OpenVINO IR and identify the relevant set of operators involved in the inference process, both for standard operations and custom/contributed operations specific to the models.
- Converting GenAI Models to OpenVINO IR for Quantization For performance optimization, I focused on converting the Llama and Phi3 models to OpenVINO IR at various precision levels, including FP16, INT4, and INT8. This was a key task in enhancing the

computational efficiency of the models by reducing their size and memory requirements without significantly sacrificing accuracy. Using OpenVINO39;s quantization tools, I experimented with various configurations to obtain the most efficient quantized models suitable for Intel hardware acceleration.

- Exploring Operators in GenAI Models: For performance optimization, I focused on converting the Llama and Phi3 models to OpenVINO IR at various precision levels, including FP16, INT4, and INT8. This was a key task in enhancing the computational efficiency of the models by reducing their size and memory requirements without significantly sacrificing accuracy. Using OpenVINO quantization tools, I experimented with various configurations to obtain the most efficient quantized models suitable for Intel hardware acceleration. For example, during the quantization process, the models typically used standard operations like DequantizeLinear and MatMul, and I explored how these could be replaced or optimized with custom operators to improve performance.

- Generation of 4-bit Phi3 QDQ Model and Quantization Exploration In this part of the project, I focused on generating a 4-bit QDQ (Quantize-Dequantize) version of the Phi3 model. The process involved the following steps: Exporting the Phi3 model with INT4 precision using the Optimum CLI export tool.

  1. Quantizing the exported model using the latest ONNX Runtime (ORT) branch with a Python script.
  2. Analyzing the operators in the model using the python API to check for any custom operators or discrepancies.
  3. Generating a JSON profiling report during inference to track performance metrics.

- Performance Benchmarking of INT4 QDQ Models: The performance impact of QDQ quantization was significant. By converting the Phi3 model to a 4-bit QDQ format, the average inference time was significantly reduced. This reduction in inference time highlights the effectiveness of quantization-aware training and QDQ techniques in improving the execution speed of deep learning models on Intel hardware. 6. Exploring Optimizations during Inference While analyzing the 4-bit Phi3 QDQ model, I observed that, instead of using the custom MatMulNBits operator, the model utilized a combination of DequantizeLinear and MatMul operators. This was an important finding, as it suggested the optimization of the quantization and matrix multiplication steps. These optimizations resulted in reduced latency and improved performance by fusing the quantization and matrix multiplication into a single FusedMatMul operation, which is more efficient than running them as separate operations. This included testing the model on different backends, identifying compatibility issues, performance analysis and reporting leading to a better understanding of the strengths and limitations of each backend for running the Phi3 QDQ model.

B.2 Bug Fixing and Validation

B.2.1 NuGet Package Generation and Validation

Automated the generation and validation of NuGet packages for Windows environments. Description: This task involved learning the fundamentals of the NuGet ecosystem, setting up workflows for package generation, and writing scripts to automate the build and validation of NuGet packages. The goal was to ensure seamless package creation and validation for distribution. You worked on creating automated scripts to streamline the NuGet package process on a Windows machine, gaining an understanding of the package structure, installation, and deployment.

B.2.2 Fixing C# Sample Bounding Box and Labeling Issues for ONNX Object Detection:

In this task, I worked on resolving issues in C# samples for object detection using the YOLOv3 model from the ONNX Model Zoo. The problem was causing errors when adding bounding boxes and labels to detected objects in the image. You updated the Program.cs file, ensuring proper handling of bounding boxes and labels. After making necessary changes to the C# project files, you verified the success of the fix by performing object detection using ONNX Runtime on sample images, confirming accurate bounding box generation.

B.2.3 Fixing and Updating OpenVINO Execution Provider (OVEP) Samples and ORT Build Instructions

Fixed build issues and updated OpenVINO Execution Provider samples and ONNX Runtime (ORT) build instructions. These tasks collectively contributed to enhancing automation, improving sample functionality, and resolving key build issues to ensure a smoother development experience with NuGet, ONNX Runtime, and OpenVINO Execution Providers. 4. Dashboard Generation using PowerBI for Validation Infrastructure I have worked on creating a graphical validation dashboard where I majorly focused on analysing ORT validation data sheets and performed data cleaning using PowerBI tool. Performed Visualization using PowerBI that helps to represent the latency numbers of models and the count of models that are executed on different Device_EPs.

B.3 Development: Writing Test Cases for Features

- External Weights Test Case: I was responsible for creating and integrating test cases to validate specific features, ensuring their robustness and correct functionality. Two key areas you worked on involved validating the external_weights feature and handling the error scenario when an NPU is not available after a driver upgrade.
- Error String Test Case (Model Needs to Be Recompiled when NPU is Not Available): I created a test case to ensure that the system correctly triggers an error message when trying to use an NPU that is no longer available after a driver update.

Appendix C

Technologies

C.1 Technologies Used

Machine Learning / AI Frameworks:

- ONNX (Open Neural Network Exchange): ONNX is used to define and share deep learning models across different platforms. I worked on validating ONNX models and learning their architectures for operator evaluations. NPU (Neural Processing Unit): NPUs are specialized hardware accelerators designed to efficiently execute AI models, particularly deep learning workloads.
- OpenVINO (Open Visual Inference and Neural Network Optimization): A toolkit by Intel used for optimizing deep learning models and deploying them on Intel hardware, including NPUs, CPUs, and GPUs. You worked on OpenVINO Execution Provider (OVEP) samples and issues related to build instructions and operator compatibility.
- ONNX Runtime (ORT): This is a cross-platform, high-performance scoring engine for Open Neural Network Exchange (ONNX) models. You used ONNX Runtime to run and troubleshoot object detection models (YOLOv3) and improve the performance using different backends (like CPU MLAS, OpenVINO).

Package Management and Build Tools

- NuGet: A package management system for .NET, used for managing libraries and tools. You worked on automating NuGet package generation for Windows environments, learning about the NuGet ecosystem, and creating validation test cases for packaging and distribution.
- Python: Python is commonly used for machine learning tasks, and you likely used it to write scripts for model conversion (e.g., converting models to OpenVINO IR) and automating test cases.

Appendix D

Summary

During my 6-month internship at Intel in Bangalore, I had the opportunity to work on several impactful projects and tasks, ranging from development to generative AI models training, inferencing and quantization. My primary focus was on building GENAI pipelines, developing APIs, and integrating LLMs models with required quantizations. I worked with technologies such as OpenVINO, ONNXRuntime, Python, and OVEP to automate data workflows, building pipelines, and provide valuable inferencing on machine learning and deep learning models.

This internship allowed me to enhance my technical skills, data integration into jenkins pipelines with Python and CPP, machine learning model building and training of LLMs. I also gained experience in creating interactive visualizations on dashboards in validation tasks. The hands-on work in real-time model building and conversions using quantization and model optimization greatly improved my problem-solving abilities.

This internship has helped me build a solid foundation in various technologies and practical applications that will support my professional growth in the tech industry.