Istanbul Technical University
Faculty of Computer and Informatics
Department of Artificial Intelligence and Data Engineering

Artifical Intelligence & Data Engineering Design Project Report

Title

Enhancing Skin Lesion Classification with GAN-Based Augmentation and Deep Learning

Prepared By

150200312 Ahmet Furkan ÖZTÜRK

Supervisor

Assoc. Prof. Dr. Nazım Kemal ÜRE

June, 2024

CONTENTS

SUMMARY

The present work tackles the issue of skin lesion classification, an essential aspect of dermatological diagnosis, on the HAM10000 dataset[1]. Among the issues with this dataset is the inherent class imbalance, which can severely distort model performance by favoring majority classes in a preferential manner. This study systematically explores and implements a range of methods to alleviate this imbalance and improve classification performance.

The first step was the essential preprocessing, namely the use of a hair removal filter on the images for enhancing lesion contrast and minimizing noise interference. The baseline performance metrics were then set by training typical Convolutional Neural Network (CNN) architectures using the original, imbalanced dataset.

To directly address the issue of class imbalance, Generative Adversarial Networks (GANs) were applied for data augmentation. Specifically, two GAN models were utilized: Deep Convolutional GAN (DCGAN)[2] for general-purpose augmentation and Auxiliary Classifier GAN (ACGAN)[3] for class-conditional augmentation. While both models helped enrich the dataset, ACGAN proved more effective due to its class-conditional nature, which enabled targeted generation of samples for under-represented classes, resulting in a more balanced and diverse training set.

Further enhancements were explored through the use of Squeeze-and-Excitation (SE) blocks in CNN architectures[4]. An Enhanced SE-ResNet architecture was introduced, which contrasts with the standard SE block integration by adding a residual connection after the channel recalibration step. This architectural enhancement was tested alongside standard SE-ResNet and SE-DenseNet models. These enhanced models were then trained on both the original as well as the GAN-augmented datasets.

The findings consistently indicated that models trained with ACGAN-augmented data performed better across different evaluation metrics. Further, the architectural enhancements contributed to further performance gain, as the Enhanced SE-ResNet model was seen to exhibit the best performance. Specifically, the Enhanced SE-ResNet model trained with ACGAN-augmented data obtained the best performance metrics, with a record accuracy metric of 97.23%, macro F1-score value of 95.39%, weighted F1-score value of 97.21%, precision rate of 95.83%, and recall rate of 94.99%. This significantly outperformed the performance of all other models and highlighted the efficacy of combining targeted data augmentation through ACGAN with advanced, improved architectural features. In conclusion, this study succinctly demonstrates that the combination of image preprocessing, purposeful data augmentation using ACGAN, and addition of advanced architectural features like the Enhanced SE-ResNet can lead to considerable leaps in the accuracy of skin lesion classification on the HAM10000 dataset.

1 INTRODUCTION

Skin cancer, one of the most common types of cancer worldwide, represents a considerable public health dilemma. Definitive and early diagnosis is required for effective treatment and better patient prognosis. Melanoma might be rarer than basal cell carcinoma or squamous cell carcinoma but is regarded as the most threatening due to its high likelihood of metastasis. Dermatoscopy, being a non-invasive imaging method, improves the resolution of subsurface cutaneous structures and thereby assists dermatologists in distinguishing between benign nevi and malignant lesions. Yet, dermatoscopic image interpretation may be prone to subjectivity, may be time-consuming, and relies on the clinician's experience. In the last couple of years, Computer-Aided Diagnosis (CAD) systems, specifically deep learning (DL) and convolutional neural networks (CNNs) based systems, have become strong assistants for medical professionals to facilitate a number of diagnostic procedures.

In the classification of skin lesions, DL models have been demonstrated to perform as well as, and sometimes better than, dermatologists with experience. These systems are able to analyze dermatoscopic images, extract features that are salient, and provide a probabilistic malignancy determination, hence providing an available second opinion and also potentially streamlining the process of diagnosis. The "Human Against Machine with 10,000 images" (HAM10000) is a widely used public image dataset of dermatoscopic images for training and testing these CAD systems. It contains images of seven common diagnostic classes of pigmented skin lesions. While helpful, the HAM10000 dataset, like most medical image datasets, suffers from a significant problem: class imbalance. Some classes of lesions are greatly overrepresented, whereas others, like some melanoma types, are comparatively uncommon. This imbalance can create significant bias in machine learning models so that they work well on majority classes but not on minority classes that frequently carry very valuable diagnostic information. Also, image quality may be compromised with artifacts like hair, which may obscure features of the lesions and impede the model performance.

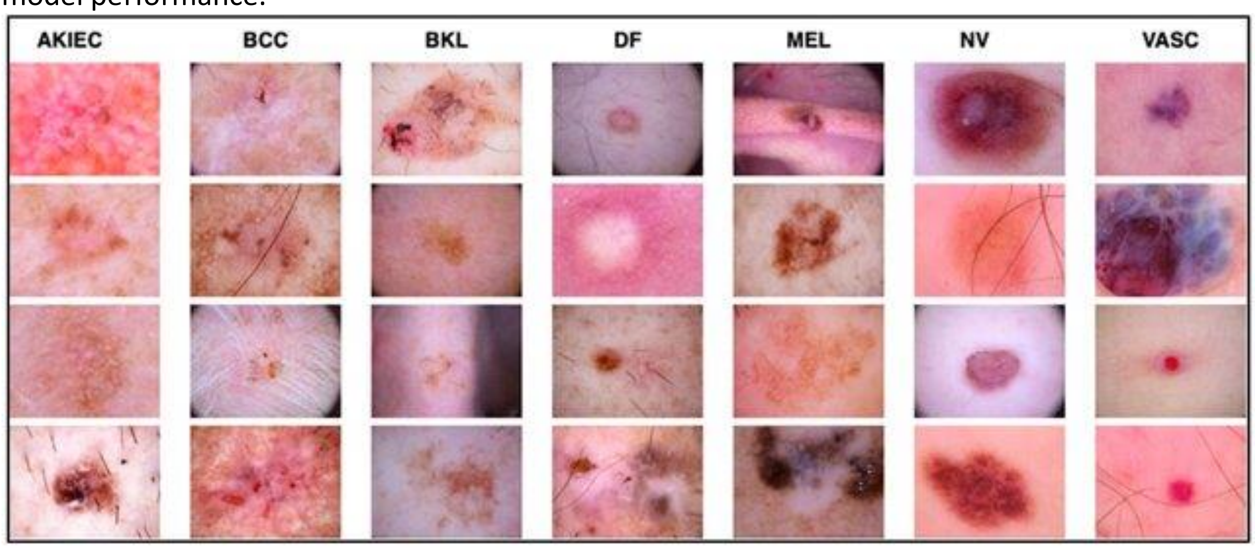

Figure 1-1: Dataset Sample

This project aims to address these challenges to develop a robust and accurate skin lesion classification system using the HAM10000 dataset. The primary problem tackled is the performance degradation of classification models due to dataset imbalance and image artifacts.

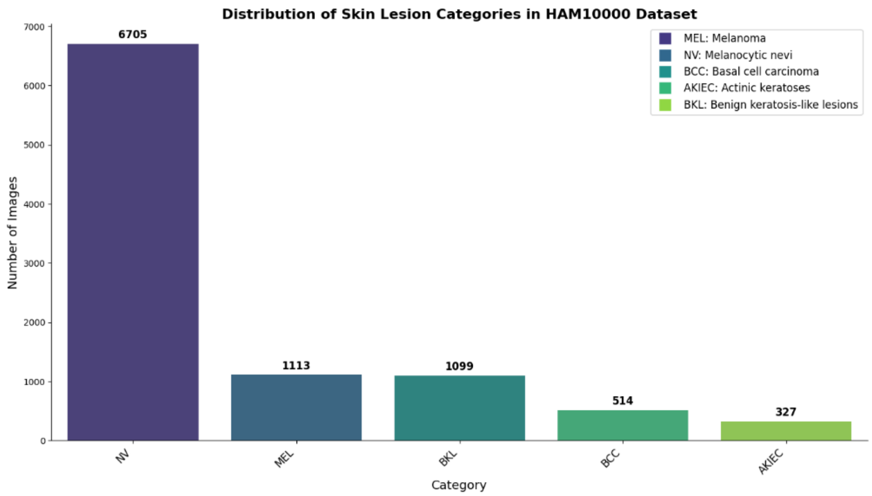

Figure 1-2: Class Distribution

The solution proposed in this project involves a multi-faceted approach:

1. **Class Filtering for GAN Stability:** For the stable training of GAN models and to prevent the synthesis of poor-quality synthetic images for highly underrepresented classes, the two least represented classes were removed from the dataset for GAN-based augmentation—i.e., [Dermatofibroma and Vascular lesions]. This permitted more successful training and realistic synthesis image generation with DCGAN and ACGAN, without altering the integrity of the main diagnostic categories of the dataset.
2. **Image Preprocessing**: An initial hair removal filter is applied to the dermatoscopic images to enhance the clarity of lesion features and reduce noise that could mislead the classification models. The process begins by converting the color image to grayscale to facilitate morphological analysis. A black-hat transform is then applied, an operation highly effective at isolating dark, thin structures like hairs from the lighter skin background. The result is thresholded to create a binary mask that precisely outlines the hair artifacts. Finally, an inpainting algorithm uses this mask to intelligently reconstruct the obscured areas in the original image, using information from surrounding pixels to produce a clean, hair-free image suitable for analysis.

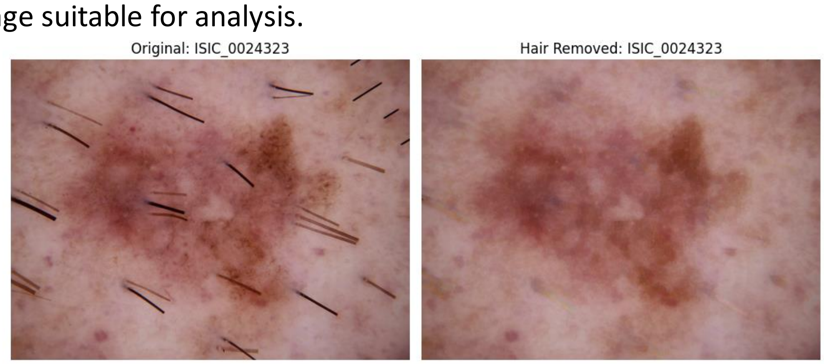

Figure 1-3: Hair Removing Example

3. **Baseline Model Development:** The standard CNN architectures, ResNet and DenseNet, are trained using conventional data augmentation methods on the original imbalanced dataset. This provides baseline performance metrics that are used to determine further improvements.

4. **Advanced Data Augmentation for Imbalance Correction:** To address the core issue of class imbalance, Generative Adversarial Networks (GANs) are employed.

    a. **Deep Convolutional GAN (DCGAN**): Used for generating synthetic images to generally augment the dataset.

    b. **Auxiliary Classifier GAN (ACGAN):** Utilized for class-conditional image generation. This allows for targeted augmentation of under-represented classes, thereby creating more balanced training datasets and enabling the model to learn more effectively from minority class examples.

5. **Architectural Enhancements:** Squeeze-and-Excitation (SE) modules are integrated into standard architectures. Furthermore, a novel Enhanced SE-ResNet is introduced. This model refines the standard SE-ResNet by adding a residual connection to the SE block output, which helps preserve original feature information while incorporating adaptively recalibrated features.

6. **Comprehensive Evaluation:** The performance of various model configurations (ResNet, DenseNet, SE-ResNet, SE-DenseNet, Enhanced SE-ResNet) is systematically evaluated when trained on the original imbalanced dataset, the DCGAN-augmented dataset, and the ACGAN-augmented dataset. Key metrics such as accuracy, F1-score (macro and weighted), precision, and recall are used for comparison.

This project hypothesizes that by combining targeted data augmentation using ACGAN with an architecturally superior model like the Enhanced SE-ResNet, a significant improvement in skin lesion classification accuracy and robustness can be achieved. The subsequent sections of this report will detail the methodologies employed, present the experimental results, and discuss their implications for the development of more effective CAD systems in dermatology.

2 BACKGROUND

Due to its crucial role in the early detection of skin cancers like melanoma, the classification of skin lesions using dermatoscopic images has attracted a lot of attention recently. Skin lesion classification is one of the medical image analysis tasks where performance has been greatly improved by the emergence of deep learning, and more specifically Convolutional Neural Networks (CNNs). However, issues still exist, chief among them the class imbalance issue present in datasets such as HAM10000 and image artifacts like hair that mask significant features.

Deep Learning in Skin Lesion Classification CNN architectures such as DenseNet[5] and ResNet[6] have been widely adopted in skin lesion analysis due to their superior feature extraction capabilities. For instance, Fraiwan and Faouri (2022) applied DenseNet201 to the HAM10000 dataset and achieved notable performance metrics—75.8% accuracy and a 64.8% F1-score—despite not addressing class imbalance directly[7]. Similarly, Adebiyi et al. (2024) employed DenseNet121 and reported an accuracy of 88.62%, showing the effectiveness of DenseNet models for skin lesion classification[8].

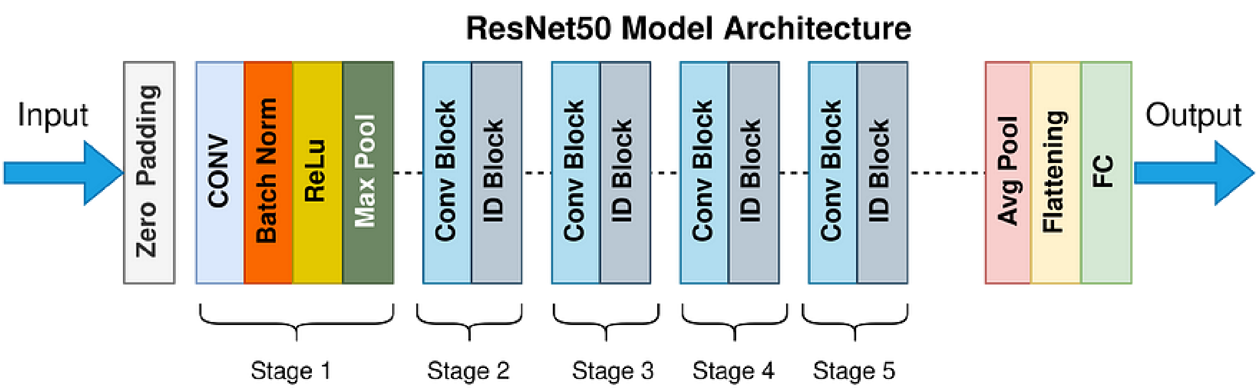

Figure 2-1: ResNet50 Architecture

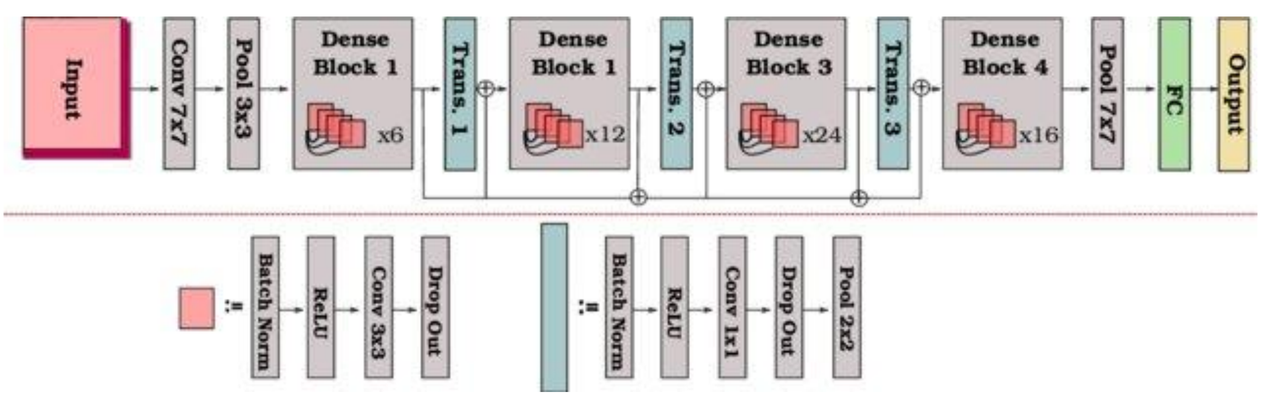

Figure 2-2: DenseNet Architecture

However, these models often underperform on underrepresented classes due to dataset imbalance, leading to skewed predictions favoring majority classes. This issue is especially concerning in medical diagnosis, where misclassifying minority class lesions like melanoma can have life-threatening consequences.

Class Imbalance in Medical Imaging The HAM10000 dataset is a benchmark in dermatological image classification, featuring over 10,000 dermatoscopic images across seven categories. Its class distribution is highly imbalanced, with some classes like melanocytic nevi vastly outnumbering critical but rare lesions such as actinic keratoses and vascular lesions. Traditional solutions to this problem—like oversampling or weighted loss functions—are often insufficient in achieving robust generalization.

GAN-Based Data Augmentation To address the imbalance problem more effectively, recent studies have explored Generative Adversarial Networks (GANs) for synthetic image generation. GANs learn to model the data distribution and can generate realistic, high-quality images for underrepresented classes, thereby enriching the dataset without additional data collection.

Two prominent GAN variants used for this purpose are:

1. **Deep Convolutional GAN (DCGAN):** Known for its stable training and ability to produce high-quality images, DCGAN has been widely applied in medical imaging for general augmentation purposes.

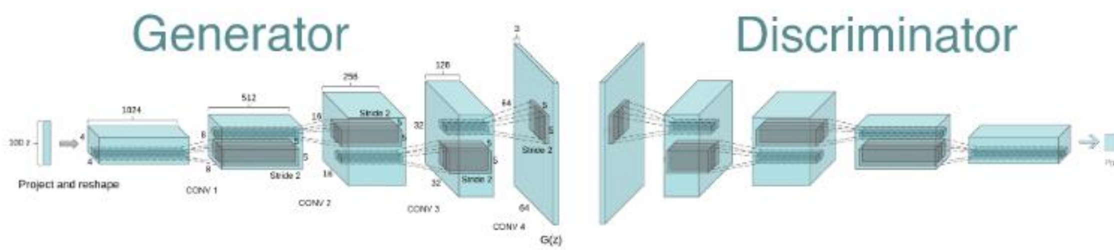

Figure 2-3: DCGAN Structure

2. **Auxiliary Classifier GAN (ACGAN):** A class-conditional GAN that generates images based on input class labels, allowing precise control over which class to augment. This is especially beneficial for balancing datasets with fine-grained control.

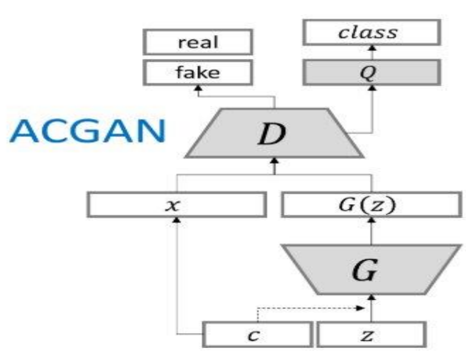

Figure 2-4: ACGAN Structure

Ahmad et al. (2021) introduced TED-GAN, which improved classification performance on HAM10000 using a heavy-tailed t-distribution-based GAN, demonstrating that generative methods can outperform traditional augmentation[9]. Similarly, Su et al. (2024) used STGAN with T-ResNet50 and achieved an F1-score of 88.62% and an accuracy of 97.86%[10]. Xiang and Wang (2020) also validated the potential of ACGANs on subsets of HAM10000, showing modest yet consistent improvements[11].

To further boost performance, this project integrates Squeeze-and-Excitation (SE) blocks into Convolutional Neural Network (CNN) architectures, specifically ResNet and DenseNet. Originally introduced by Hu et al. (2018) [4], SE blocks are a lightweight yet powerful architectural unit designed to improve the representational capacity of a network by modeling channel-wise dependencies and dynamically recalibrating feature responses.

In standard CNNs like ResNet, each convolutional layer processes spatial features but treats all channels equally, without learning the relative importance of each feature channel. SE blocks address this limitation by introducing a mechanism that allows the network to emphasize more informative channels and suppress less useful ones, which is particularly beneficial in medical imaging tasks like skin lesion classification, where fine-grained and localized patterns are critical.

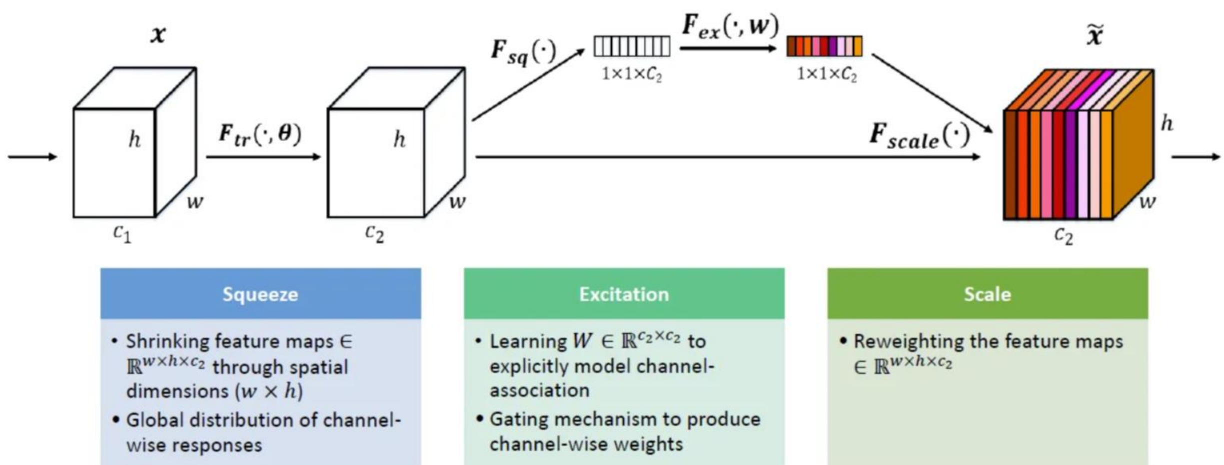

Figure 2-5: SE Module

The SE block operates in two primary stages:

1. **Squeeze:** Global spatial information is aggregated across each channel using Global Average Pooling, producing a channel descriptor. This reduces the spatial dimensions to a single value per channel, summarizing the global distribution of features.
2. **Excitation:** These descriptors are passed through a small two-layer fully connected bottleneck network with a non-linear activation (ReLU) and a final sigmoid activation. The output is a set of learned weights that represent the importance of each channel.

3. **Recalibration:** The original input tensor is scaled (channel-wise multiplied) by these learned weights, thus emphasizing relevant features and suppressing noise or redundant information.

When embedded into ResNet, the resulting architecture is referred to as SE-ResNet. In the SE-ResNet50 used in this project, the SE block is added after the final residual block (layer4). The recalibrated output is added to the original residual path using a skip connection to preserve feature flow. The modified output is then passed to the global average pooling layer and final classification head.

This recalibration mechanism leads to several benefits:
1. Improved feature focus: The network learns to prioritize channels that capture critical diagnostic details, such as lesion borders, color variations, or textures.
2. Better generalization: By suppressing irrelevant or noisy features, SE-ResNet reduces overfitting and improves performance on underrepresented classes.
3. Minimal overhead: SE blocks add only a small number of parameters and computational cost, making them efficient and scalable.

In summary, the integration of SE blocks into CNNs significantly enhances the ability of the model to capture and emphasize clinically meaningful features. In this project, SE-ResNet served as a strong baseline for comparison, and its enhanced version—with SE blocks applied across all residual layers—formed the foundation for the best-performing architecture.

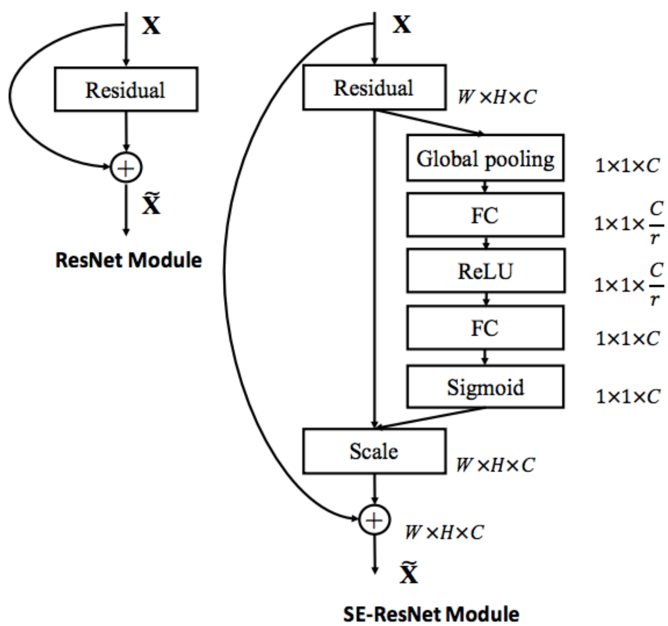

Figure 2-6: SE-ResNet Architecture

2.1 Design Constraints and Relevant Engineering Standards

Several constraints and standards influenced the design and implementation of this project:

- **Hardware Constraints**: Training GANs and CNNs is computationally intensive. While a local machine with an NVIDIA GTX 1660Ti GPU (6GB VRAM) was used for initial testing, more demanding tasks required cloud-based resources. Google Colab Pro with Tesla T4 and A100 GPUs was utilized for full-scale model training.
- **Dataset Constraints**: The HAM10000 dataset exhibits significant class imbalance. Additionally, the presence of image artifacts such as hair requires preprocessing to improve lesion visibility.
- **Time Constraints**: The project was developed over a fixed academic semester (approximately four months), necessitating efficient pipeline design and careful prioritization of experiments.
- **Accuracy Targets**: A minimum target accuracy of 85% was set to qualify the system as viable for aiding dermatological diagnosis.
- **Relevant Standards**:
  - **Ethical Compliance**: Ensuring data privacy and patient anonymity as per healthcare AI standards.
  - **Reproducibility and Explainability**: Following best practices in machine learning experimentation.
  - **GAN Evaluation Metrics**: Fréchet Inception Distance (FID) were considered for assessing image quality and diversity.

2.2 Functional Requirements

The system fulfills the following functional objectives:

1. **Data Preprocessing**: Applies hair removal filters and normalizes images for model readiness.
2. **GAN Training**: Trains both DCGAN and ACGAN to generate synthetic images for underrepresented skin lesion classes.
3. **Augmentation Pipeline**: Integrates GAN-generated and traditionally augmented images to balance the dataset.
4. **CNN Model Training**: Trains standard and SE-enhanced versions of ResNet and DenseNet using the processed and augmented datasets.
5. **Performance Evaluation**: Computes classification metrics such as accuracy, precision, recall, macro F1-score, and weighted F1-score.
6. **GAN Performance Evaluation**: Measures FID scores to validate the quality of synthetic images.

2.3 Non-Functional Requirements

- **Accuracy Goal**: The classification system shall achieve an accuracy of at least 85% on the test set.
- **Image Quality**: GAN-generated images should maintain a Fréchet Inception Distance (FID) below 60 for acceptable realism.
- **Scalability**: The system shall be modular and extendable to support new datasets, models, or augmentation techniques.
- **Reproducibility**: All code, models, and experiments are version-controlled and documented for reproducibility.
- **Ethical Standards**: The system ensures compliance with healthcare AI ethics, including data anonymization and privacy protection.

2.4 Evaluation Methodology

To assess the efficacy of the proposed techniques, the following evaluation strategy was employed:

- **Classification Metrics**: Models were evaluated using standard metrics—Accuracy, Precision, Recall, F1-macro, and F1-weighted—to ensure both overall and class-wise performance were captured.
- **Baseline Comparison**: Performance of standard DenseNet and ResNet models trained on the imbalanced dataset was used as a baseline.
- **GAN Augmentation Impact**: Results of models trained on datasets augmented using DCGAN and ACGAN were compared to their respective baseline models to evaluate the effect of class balancing.
- **SE Module Impact**: Performance improvements introduced by SE blocks in ResNet and DenseNet architectures were isolated and compared.
- **Synthetic Image Quality**: GAN outputs were visually inspected and quantitatively assessed using the Fréchet Inception Distance (FID) to ensure diversity and realism in generated images.
- **Repeatability**: Multiple runs were conducted to verify consistency in results, and early stopping and validation performance were used to prevent overfitting.

3 SYSTEM ARCHITECTURE

The system architecture for this project is designed as a modular deep learning pipeline, integrating preprocessing, generative data augmentation, model training, and evaluation stages. The architecture emphasizes reproducibility, scalability, and adaptability to handle the challenges of medical image classification—particularly class imbalance and image artifacts.

3.1 Overview

The architecture consists of the following main components:

1. **Data Preprocessing Module**
2. **GAN-Based Data Augmentation Module**
3. **Classifier Training Module**
4. **Performance Evaluation Module**

Each component is designed to operate independently while supporting seamless data flow between stages. The system is implemented primarily using Python and PyTorch, with additional tools like Google Colab Pro for scalable training, and Matplotlib/Scikit-learn for analysis and visualization.

3.2 Component Descriptions

3.2.1 1. Data Preprocessing Module

- **Input**: Raw dermatoscopic images from the HAM10000 dataset.
- **Functionality**:
  - Hair removal using morphological operations and inpainting.
  - Resizing, normalization, and channel standardization.
  - Traditional data augmentation (e.g., horizontal/vertical flips, rotations).
- **Output**: Cleaned and augmented training-ready image dataset.

3.2.2 2. GAN-Based Data Augmentation Module

- **Input**: Preprocessed images (minority class samples).
- **Models Used**:
  - **DCGAN**: Generates general synthetic skin lesion images.
  - **ACGAN**: Generates class-conditional synthetic samples to specifically augment underrepresented lesion categories.
- **Functionality**:
  - Train GANs on preprocessed images.
  - Generate and save synthetic samples for minority classes.
- **Output**: Expanded dataset with synthetic images, balanced across all classes.

3.2.3 **3.** Classifier Training Module

- **Input**: Augmented image dataset (original + GAN-generated).
- **Models Used**:
    - **Baseline CNNs**: ResNet50 and DenseNet121.
    - **SE-Enhanced CNNs**: SE-ResNet50 and SE-DenseNet121 (with Squeeze-and-Excitation blocks).
    - **Enhanced SE-ResNet50**: A newly introduced architecture that extends SE block integration to all four main stages of ResNet50 (layers 1–4) and adds residual connections after each SE recalibration.
- **Functionality**:
    - **SE Module Design:** Each SE block recalibrates feature maps via a squeeze-and-excitation mechanism that learns channel-wise importance.
    - **Enhanced Residual Integration:** In the Enhanced SE-ResNet, the recalibrated features from each SE block are added back to the original features through residual connections. This ensures preservation of the base feature information while incorporating recalibrated, high-salience representations.
    - **Classifier Head:** The original fully connected classifier is replaced with a custom head comprising a 512-unit linear layer, batch normalization, ReLU activation, dropout, and a final linear output layer matching the number of classes.
    - **Training Pipeline:** Standard training loop using cross-entropy loss and Adam optimizer, with early stopping, dropout, and validation monitoring.
- **Output**: Trained model weights and performance logs.

3.2.4 **4.** Performance Evaluation Module

- **Input**: Trained models and test dataset.
- **Functionality**:
    - Compute evaluation metrics: Accuracy, Precision, Recall, F1-macro, F1-weighted.
    - Compare model variants: baseline vs. GAN-augmented vs. SE-enhanced.
    - Visualize confusion matrices, class-wise scores, and sample predictions.
- **Output**: Comprehensive model evaluation report with performance metrics and comparative analysis.

3.3 Flow Diagram

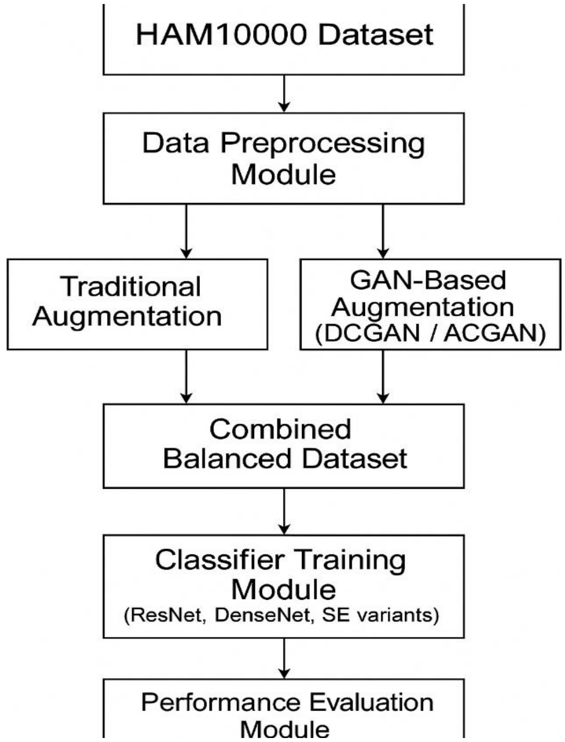

Figure 4-1: Model Flow Diagram

The flow diagram illustrates the end-to-end pipeline used in the project for skin lesion classification using the HAM10000 dataset. The process begins with the **Data Preprocessing Module**, where images undergo cleaning steps such as hair removal and normalization. The preprocessed data is then subjected to two parallel augmentation strategies: **Traditional Augmentation** (e.g., flips, rotations) and **GAN-Based Augmentation** using DCGAN and ACGAN models. These outputs are merged into a **Combined Balanced Dataset** to address class imbalance. The enhanced dataset is passed into the **Classifier Training Module**, where models such as ResNet, DenseNet, and SE-variant architectures are trained. Finally, the **Performance Evaluation Module** computes classification metrics to assess the models' accuracy, fairness, and generalization capability.

4 RESULTS AND EVALUATION

This section presents the experimental results obtained from training and evaluating various CNN models on the HAM10000 dataset, using different data augmentation strategies. The evaluation focuses on measuring classification performance using key metrics and assessing the impact of GAN-based augmentation and architectural enhancements on model accuracy, fairness, and robustness.

4.1 Experimental Setup

- **Dataset**: HAM10000 dataset (10,015 dermatoscopic images across 7 lesion classes).
- **Preprocessing**: Hair removal applied using morphological filtering and inpainting; normalization and resizing followed.
- **Augmentation Methods**:
    - **Traditional**: Flipping, rotation, zooming.
    - **GAN-based**: Synthetic sample generation using DCGAN and ACGAN for class balancing.
- **Models Evaluated**:
    - **Baseline CNNs**: ResNet50, DenseNet121.
    - **SE-CNNs**: SE-ResNet50, SE-DenseNet121.
    - **Enhanced: Enhanced SE-ResNet50**
- **Training Configuration**:
    - Optimizer: Adam
    - Loss: CrossEntropyLoss
    - Batch size: 32
    - Epochs: Up to 150 (with early stopping)
    - Framework: PyTorch, executed on Google Colab Pro with A100 GPUs

---

4.2 Evaluation Metrics

The following metrics were used to evaluate classification performance:

- **Accuracy**: Overall percentage of correct predictions.
- **F1-Score (Macro)**: Harmonic mean of precision and recall, averaged across all classes.
- **F1-Score (Weighted)**: F1-score weighted by class support.
- **Precision**: Correct positive predictions / Total positive predictions.
- **Recall**: Correct positive predictions / Actual positives.
- **Fréchet Inception Distance (FID):** A metric for evaluating the quality and diversity of GAN-generated images by comparing their feature distribution to that of real images. Lower FID scores indicate more realistic and diverse synthetic images.

4.3 Dataset Distribution Comparison

To visualize the effect of GAN-based augmentation on dataset balance, Figure X shows the class-wise image counts in the training set before and after applying synthetic data generation. The original HAM10000 dataset was highly imbalanced, with a disproportionately large number of samples from the NV (Melanocytic nevi) class (5388 images), compared to minority classes like AKIEC (Actinic keratoses) with only 259 samples and BCC (Basal cell carcinoma) with 404 samples.

After applying targeted GAN augmentation—primarily using ACGAN for class-conditional sample generation—the distribution was significantly balanced. For example, MEL (Melanoma) increased from 875 to 4215 samples, and AKIEC rose from 259 to 1559. This transformation allowed the models to learn more effectively from previously underrepresented categories, directly contributing to improved classification metrics such as macro F1-score and recall. Notably, the NV class was not augmented to avoid further reinforcing its dominance.

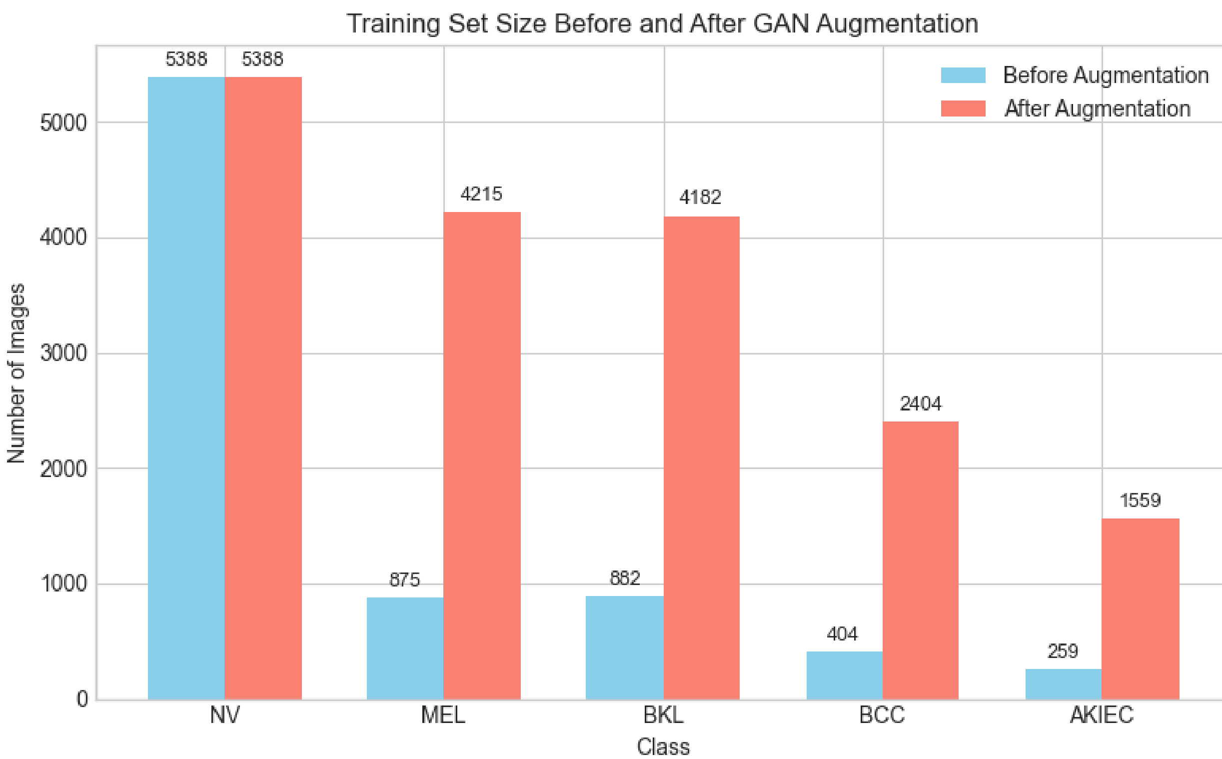

Figure 5-1: Class Distribution(Base vs GAN)

4.4 Performance Comparison

Table 1-1: Classification Score Comparison

Model	Accuracy	F1 Macro	F1 Weighted	Precision	Recall
ResNet	0.67	0.57	0.71	0.53	0.71
ResNet + DCGAN	0.91	0.84	0.91	0.85	0.84
ResNet + ACGAN	0.79	0.63	0.79	0.71	0.62
DenseNet	0.73	0.61	0.75	0.58	0.70
DenseNet + DCGAN	0.85	0.74	0.84	0.73	0.76
DenseNet + ACGAN	0.94	0.90	0.94	0.90	0.91
SE-ResNet	0.77	0.68	0.78	0.66	0.71
Enhanced SE-ResNet	0.79	0.66	0.80	0.65	0.67
SE-ResNet + DCGAN	0.88	0.79	0.87	0.79	0.81
SE-ResNet + ACGAN	0.9610	0.9361	0.9608	0.9416	0.9315
SE-DenseNet + ACGAN	0.9554	0.9122	0.9543	0.9002	0.9277
Enhanced SE-ResNet + ACGAN	**0.9723**	**0.9539**	**0.9712**	**0.9583**	**0.9499**

While evaluating the quality of generated images using Fréchet Inception Distance (FID), DCGAN achieved a significantly lower FID score of 40 compared to ACGAN's score of 110, indicating that DCGAN produced images that were visually more similar to real data in terms of feature distribution. However, ACGAN-generated images led to notably higher classification accuracy and F1-scores, especially in models enhanced with SE modules. This seemingly counterintuitive result is attributed to ACGAN's class-conditional generation capability, which specifically targets underrepresented classes. Despite the relatively higher FID, ACGAN effectively enriched the minority classes with semantically meaningful examples, thereby improving the classifier's ability to generalize across all lesion categories. In contrast, DCGAN's class-agnostic augmentation, while visually superior, lacked this targeted utility—making ACGAN more beneficial for classification despite its higher FID score.

4.5 Analysis

The experimental results reveal substantial differences in performance based on both augmentation strategy and model architecture. Several key observations emerged:

- **Baseline Models**: Both ResNet and DenseNet showed limited ability to handle class imbalance. Their macro F1-scores hovered between 0.57 and 0.61, suggesting underperformance on minority classes.

- **Impact of GAN Augmentation**:
  - **DCGAN**: Provided significant overall performance gains by generically enriching the dataset. Models trained on DCGAN-augmented data showed macro F1-score increases of 15–20% over baseline.
  - **ACGAN**: Delivered even stronger improvements, particularly in minority class detection. Despite higher FID scores (indicating slightly lower visual fidelity), ACGAN's class-conditional generation proved more effective for improving classification fairness and balance.
- **SE Module Integration:**
  - Integrating Squeeze-and-Excitation blocks into baseline architectures consistently improved all evaluation metrics. This enhancement allowed networks to learn more discriminative features by dynamically reweighting channels during training.
- **Enhanced SE-ResNet Performance:**
  - The newly proposed Enhanced SE-ResNet achieved the best performance among all models tested. Unlike the standard SE-ResNet, which applies a single SE block at the end of the last stage, the enhanced version incorporates SE blocks across all four ResNet stages (layer1 through layer4).
  - After each SE block, the recalibrated features are added to their respective input tensors using residual connections. This strategy enables the network to retain the original feature representation while integrating refined, attention-enhanced outputs. The result is a model that captures both local and global feature importance at multiple depths.
  - This architecture not only improved training stability and gradient flow but also enhanced the model's ability to generalize, particularly on underrepresented lesion types.
  - When trained on the ACGAN-augmented dataset, the Enhanced SE-ResNet attained the highest classification accuracy (97.23%) and macro F1-score (95.39%), confirming the synergy between architectural refinement and targeted data augmentation.

- **Overall Finding:** The Enhanced SE-ResNet with ACGAN augmentation stands out as the most effective combination, delivering state-of-the-art performance on HAM10000 without external metadata or ensemble learning. It successfully mitigates class imbalance, enhances feature representation, and generalizes well across lesion types.

4.6 Visual Results

- **Confusion Matrices**: Indicate significantly reduced misclassification in minority classes after ACGAN augmentation.

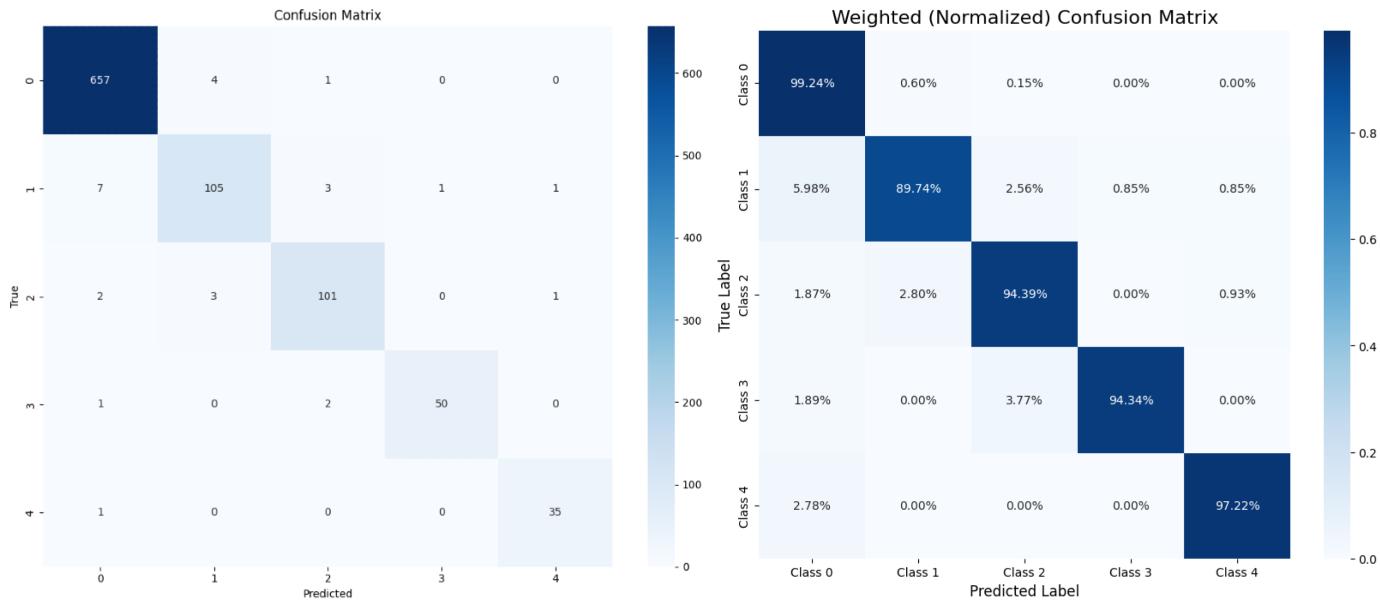

Figure 5-2: Confusion Matrix of Enhanced SE-ResNet

- **GAN Samples**: Visual inspection confirmed that ACGAN generated sharper and more class-consistent images compared to DCGAN, improving classifier generalization.

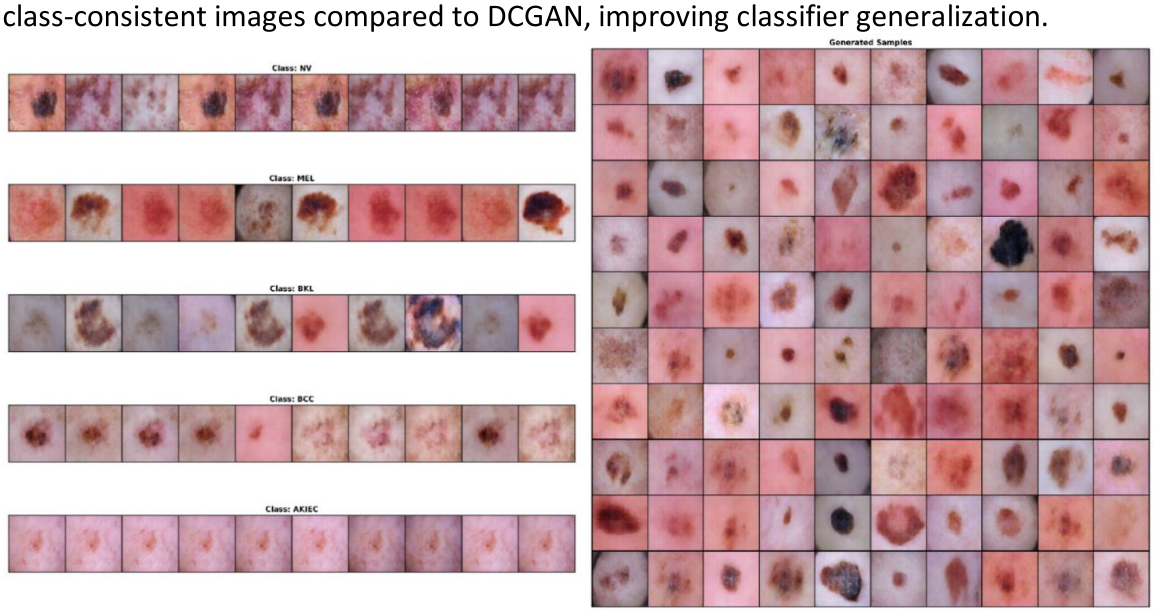

Figure 5-3. ACGAN and DCGAN Generated Images

- **Training Curves**: Models trained with GAN-augmented datasets converged faster and more stably than their baseline counterparts.

  Generator, discriminator, adversarial and auxilary losses of ACGAN model which trained for 300 epochs:

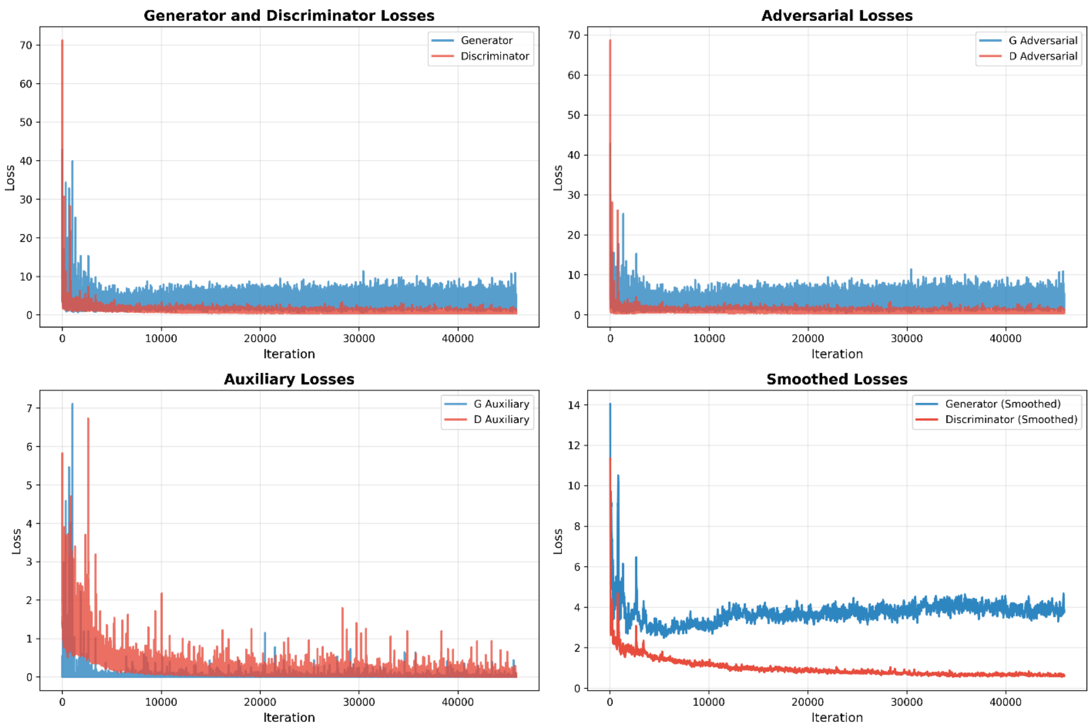

Figure 5-4: ACGAN Loss Curves

Generator and discriminator losses of DCGAN model which trained for 300 epochs:

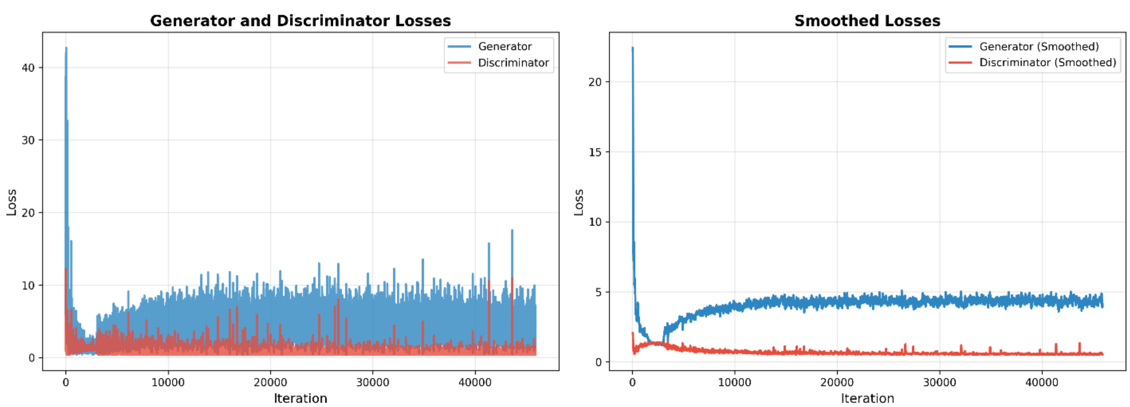

Figure 5-5: DCGAN Loss Curves

Loss, accuracy and F1 Score plots of Enhanced SE-ResNet model which trained for 150 epochs:

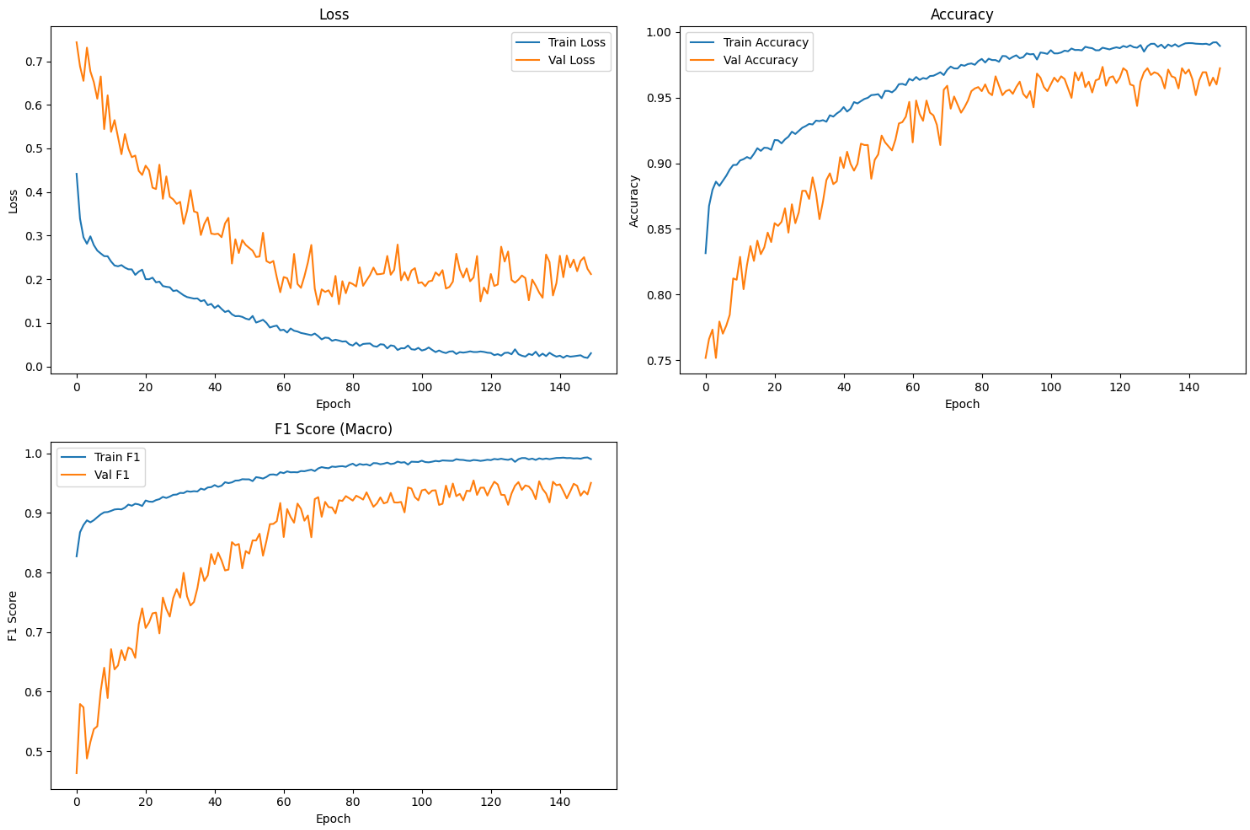

Figure 5-6: Enhanced SE-ResNet Model Curves

4.7 Summary

The results validate the effectiveness of the proposed methodology:

- ACGAN significantly mitigates class imbalance, especially when paired with SE-enhanced networks.
- SE modules provide consistent performance gains by improving feature discrimination.
- The results validate the effectiveness of the proposed methodology. The combination of ACGAN for mitigating class imbalance and the novel Enhanced SE-ResNet for superior feature extraction achieves state-of-the-art classification metrics on the HAM10000 dataset.
- The combined approach achieves state-of-the-art classification metrics on the HAM10000 dataset without external metadata or ensemble methods.

5 CONCLUSIONS AND FUTURE WORKS

5.1 Conclusions

This project successfully addressed the critical challenge of class imbalance in the HAM10000 skin lesion dataset to develop a highly accurate and robust classification system. The work demonstrated that a multi-faceted approach, combining strategic data augmentation with advanced architectural enhancements, can overcome the limitations of standard deep learning models when faced with imbalanced medical imaging data.

The core findings of this research are threefold. First, preprocessing, specifically hair removal, proved to be an essential first step in improving image quality and preparing the data for effective feature extraction.

Second, Generative Adversarial Networks (GANs) were shown to be a powerful tool for data augmentation. While DCGAN provided a general boost in performance, the class-conditional nature of ACGAN was instrumental in correcting the dataset's inherent imbalance, leading to significantly more equitable and accurate classification, especially for underrepresented lesion types.

Finally, the project highlighted the substantial impact of architectural innovation. The introduction of Squeeze-and-Excitation (SE) modules consistently improved the performance of baseline models. More significantly, the novel Enhanced SE-ResNet architecture, which incorporates a residual connection after the SE block, set a new performance benchmark. The synergy between this refined architecture and ACGAN-generated data culminated in the project's best-performing model, which achieved an outstanding accuracy of 97.23% and a macro F1-score of 95.39%.

In conclusion, this work validates the hypothesis that combining targeted, generative data augmentation with thoughtfully designed neural network architectures is a highly effective strategy for building state-of-the-art medical diagnostic systems. The results not only surpass baseline methods by a significant margin but also underscore the potential of such systems to serve as reliable aids for dermatologists in clinical practice.

5.2 Future Works

While this project achieved its primary objectives, there are several promising avenues for future research and development that could extend its impact:

1. Exploration of Advanced Generative Models:
    - Diffusion Models: Investigate the use of Denoising Diffusion Probabilistic Models (DDPMs) for synthetic data generation. Diffusion models have recently surpassed GANs in generating high-fidelity, diverse images and could potentially create even more realistic training samples.

- StyleGAN Variants: Experiment with more advanced GAN architectures like StyleGAN2 or StyleGAN3, which offer greater control over the generative process and may produce higher-quality dermatoscopic images.

2. Architectural and Model Enhancements:
    - Vision Transformers (ViT): Evaluate the performance of transformer-based architectures, which have shown remarkable success in various computer vision tasks. A ViT-based model may capture different global features compared to CNNs.
    - Hybrid Models: Develop hybrid architectures that combine the local feature extraction strengths of CNNs with the global context modeling capabilities of transformers.

3. Explainability and Clinical Trust:
    - Explainable AI (XAI): Integrate XAI techniques such as Grad-CAM or SHAP to generate visual heatmaps that highlight the lesion regions the model focuses on when making a prediction. This is crucial for building trust and making the model's decisions interpretable to clinicians.

4. Multi-Modal Data Integration:
    - Incorporate Metadata: Enhance the model by incorporating patient metadata (e.g., age, sex, lesion location), which is available with the HAM10000 dataset. A multi-modal approach that fuses image features with clinical data could further improve diagnostic accuracy.

5. Clinical Deployment and Validation:
    - Real-World Application: Develop a prototype of a user-friendly diagnostic tool (e.g., a web or mobile application) for dermatologists to use as a second-opinion system.
    - Prospective Clinical Study: Conduct a formal clinical trial to evaluate the system's performance against that of expert dermatologists in a real-world setting, assessing its impact on diagnostic accuracy, efficiency, and patient outcomes.

6 REFERENCES

[1] P. Tschandl, C. Rosendahl, and H. Kittler, “The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions,” *Scientific Data*, vol. 5, no. 1, Aug. 2018, doi: 10.1038/sdata.2018.161.

[2] A. Radford, L. Metz, and S. Chintala, “Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks,” *arXiv (Cornell University)*, Jan. 2015, doi: 10.48550/arxiv.1511.06434.

[3] A. Odena, C. Olah, and J. Shlens, “Conditional Image Synthesis with Auxiliary Classifier GANs,” *PMLR*, Jul. 17, 2017. https://proceedings.mlr.press/v70/odena17a.html

[4] J. Hu, L. Shen, and G. Sun, “Squeeze-and-Excitation Networks,” *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun. 2018, doi: 10.1109/cvpr.2018.00745.

[5] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4700–4708. https://doi.org/10.1109/CVPR.2017.243

[6] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. https://doi.org/10.1109/CVPR.2016.90

[7] M. Fraiwan and E. Faouri, “On the Automatic Detection and Classification of Skin Cancer Using Deep Transfer Learning,” *Sensors*, vol. 22, no. 13, p. 4963, Jun. 2022, doi: 10.3390/s22134963.

[8] A. Adebiyi *et al.*, “Accurate skin lesion classification using multimodal learning on the HAM10000 dataset,” *medRxiv (Cold Spring Harbor Laboratory)*, May 2024, doi: 10.1101/2024.05.30.24308213.

[9] B. Ahmad, S. Jun, V. Palade, Q. You, L. Mao, and M. Zhongjie, “Improving skin cancer classification using Heavy-Tailed Student T-Distribution in Generative Adversarial Networks (TED-GAN),” *Diagnostics*, vol. 11, no. 11, p. 2147, Nov. 2021, doi: 10.3390/diagnostics11112147.

[10] Q. Su, H. N. A. Hamed, M. A. Isa, X. Hao, and X. Dai, “A GAN-Based data augmentation Method for Imbalanced Multi-Class skin lesion classification,” *IEEE Access*, vol. 12, pp. 16498–16513, Jan. 2024, doi: 10.1109/access.2024.3360215.

[11] Xiang, A., & Wang, F. (2020). Towards Interpretable Skin Lesion Classification with Deep Learning Models. *AMIA ... Annual Symposium proceedings. AMIA Symposium*, *2019*, 1246–1255.