Project Proposal: GANs for Data Augmentation and Domain Adaptation in Medical Imaging

Contents

1 Dataset

1.1 Source and Description

Dataset Name: ISIC (International Skin Imaging Collaboration) Dataset
Domain: Medical Imaging - Dermatology / Skin Lesion Classification
Source:

- Hospital Clínic de Barcelona
- ViDIR Group, Department of Dermatology, Medical University of Vienna
- Anonymous contributors

License: Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)

1.2 Dataset Characteristics

Problem Type: Binary classification of skin lesions

- **Class 0:** Benign lesions
- **Class 1:** Malignant lesions (target for augmentation)

Dataset Statistics:

- **Baseline Dataset:**
  - Training set: ∼17,000 images total
    - Benign: ∼15,000 images
    - Malignant: ∼2,000 images
  - **Imbalance Ratio:** 7.5:1 (benign:malignant)
  - Validation and test sets maintained with similar distributions

Image Specifications:

- **Resolution:** 128×128 pixels (resized)
- **Channels:** RGB (3 channels)
- **Format:** PNG

Domain Challenges:

- **Class Imbalance:** Severe imbalance with malignant cases being $10\times$ less frequent
- **Data Scarcity:** Limited malignant samples for training robust classifiers
- **Intra-class Variability:** High diversity in lesion appearance, color, shape, and texture
- **Medical Relevance:** Critical need for accurate malignant lesion detection

1.3 Dataset Structure

```
data/
|-- raw/
|   |-- images/                    # Original ISIC images
|   +-- metadata.csv               # Image metadata and labels
|-- processed/
|   |-- baseline/                  # Initial imbalanced dataset
|   |   |-- train/
|   |   |   |-- benign/           (~15,000 images)
|   |   |   +-- malignant/        (~2,000 images)
|   |   |-- val/
|   |   +-- test/
|   |-- augmented/                 # Baseline + GAN-generated samples
|   |   +-- train/
|   |       +-- malignant/        (baseline + ~3,000 synthetic)
|   +-- domain_adaptation/         # Domain shift evaluation
|       |-- source_synthetic/      # Training: synthetic malignant + real benign
|       +-- target_real/           # Testing: real malignant + real benign
+-- synthetic/                     # GAN-generated samples by version
    |-- dcgan_hinge/
    |-- dcgan_mse/
    |-- dcgan_wasserstein/
    +-- cdcgan_*/
```

1.4 Augmentation Objectives

1. **Class Balancing:** Generate synthetic malignant samples to reduce class imbalance
2. **Diversity Enhancement:** Increase intra-class diversity for improved generalization
3. **Performance Improvement:** Enhance downstream classifier performance on minority class
4. **Domain Adaptation:** Evaluate classifier robustness across real-synthetic domain shifts

1.5 Domain Shift Specification

Objective: Investigate the domain gap between real and synthetic images and its impact on classifier generalization.

Domain Configuration:

- **Source Domain:** Synthetic malignant images (GAN-generated) + Real benign images
- **Target Domain:** Real malignant images + Real benign images

Domain Shift Characteristics:

1. **Class-Specific Shift:** Only malignant class experiences domain shift (synthetic → real)
2. **Benign Class:** Remains real in both domains (no shift)
3. **Asymmetric Challenge:** Tests if synthetic malignant samples can substitute real ones

Dataset Split:

```
Source Domain (Training):
  - Benign: 15,000 real images
  - Malignant: 3,000 synthetic (GAN-generated) images

Target Domain (Testing):
  - Benign: Real test set (~same distribution as source benign)
  - Malignant: Real test set (domain-shifted from synthetic)
```

Expected Challenges:

- **Distribution Mismatch:** Synthetic images may not capture all real-world variations
- **Fine-grained Details:** GANs may miss subtle diagnostic features
- **Generalization Gap:** Classifier may overfit to synthetic artifacts
- **Performance Drop:** Expected accuracy reduction on real malignant samples

Evaluation Metrics:

- Source domain performance (synthetic malignant)
- Target domain performance (real malignant)
- Domain gap quantification (performance difference)
- Per-class accuracy analysis
- Confusion matrix comparison across domains

Domain Adaptation Approach:

- **Primary Method:** Domain gap evaluation (train on source, test on target)
- **Classifier Loss:** Standard Cross-Entropy Loss
- **Optional Extension:** Domain-Adversarial Neural Network (DANN) could be considered to actively reduce domain gap by:
  - Adding a domain discriminator to distinguish source vs target features
  - Training classifier features to be domain-invariant
  - Loss formulation: $\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{classification}} + \lambda \times \mathcal{L}_{\text{domain\_adversarial}}$

2 Architecture

2.1 Generator Architecture

2.1.1 DCGAN Generator (Unconditional)

- **Architecture Type:** Deep Convolutional GAN (DCGAN)
- **Input:** Random latent vector $\mathbf{z} \in \mathbb{R}^{100}$
- **Output:** $128 \times 128 \times 3$ RGB image

Parameters:

- Latent dimension: 100
- Base filters (n_1): 512
- Total parameters: $\sim$11.7M

2.1.2 Conditional DCGAN Generator (cDCGAN)

- **Extension:** Adds class-conditional generation
- **Input:** $\mathbf{z} \in \mathbb{R}^{100}$ + class label embedding
- **Class Embedding:** 2 classes $\rightarrow$ 50-dimensional embedding
- **Concatenated Input:** 150-dimensional vector

2.2 Discriminator Architecture

2.2.1 PatchGAN Discriminator (Multiple Loss Variants)

Design Philosophy:

- **PatchGAN:** Classifies $N \times N$ patches instead of entire image
- **Advantages:** Better captures local texture details critical for medical images

Architectural Variants:

1. **PatchGAN with BatchNorm** (for BCE, MSE, Wasserstein losses)
    - Batch Normalization after each conv layer (except first)
    - Dropout: 0.1–0.3
2. **PatchGAN with Spectral Normalization** (for Hinge loss)
    - Spectral normalization on all conv layers
    - No Batch Normalization (incompatible with SN)
    - Dropout: 0.3

Parameters:

- Base filters (n_{df}): 64
- Number of downsampling layers (n_{layers}): 3
- Dropout probability: 0.1–0.3 (tuned per loss function)
- Output: 7×7 spatial predictions

2.2.2 Conditional PatchGAN Discriminator

- **Projection Discriminator:** Class conditioning via projection
- **Class Embedding:** Projects to feature dimension (512)
- **Integration:** Inner product of features with class embedding
- **Output:** Class-conditional patch predictions

3 Training Setup

3.1 Loss Functions

Multiple loss functions are explored and compared:

3.1.1 Hinge Loss (with Spectral Normalization)

Discriminator Loss:

$$\mathcal{L}_D = \mathbb{E}[\max(0, 1 - D(\mathbf{x}_{\text{real}}))] + \mathbb{E}[\max(0, 1 + D(\mathbf{x}_{\text{fake}}))] \tag{1}$$

Generator Loss:

$$\mathcal{L}_G = -\mathbb{E}[D(G(\mathbf{z}))] \tag{2}$$

Characteristics:

- Margin-based loss with stable gradients
- No saturation issues
- Used with Spectral Normalization
- Recommended by SN-GAN (Miyato et al., 2018)

3.1.2 Wasserstein Loss with Gradient Penalty

Discriminator Loss:

$$\mathcal{L}_D = \mathbb{E}[D(\mathbf{x}_{\text{fake}})] - \mathbb{E}[D(\mathbf{x}_{\text{real}})] + \lambda_{\text{GP}} \cdot \text{GP} \tag{3}$$

where GP is gradient penalty term

Generator Loss:

$$\mathcal{L}_G = -\mathbb{E}[D(G(\mathbf{z}))] \tag{4}$$

Hyperparameters:

- λ_{GP}: 10 (gradient penalty coefficient)
- n_{critic}: 1–2 (discriminator updates per generator update)

3.1.3 Binary Cross-Entropy (BCE) Loss

Discriminator Loss:

$$\mathcal{L}_D = -\mathbb{E}[\log D(\mathbf{x}_{\text{real}})] - \mathbb{E}[\log(1 - D(\mathbf{x}_{\text{fake}}))] \tag{5}$$

Generator Loss:

$$\mathcal{L}_G = -\mathbb{E}[\log D(G(\mathbf{z}))] \tag{6}$$

3.1.4 Mean Squared Error (MSE) Loss

Discriminator Loss:

$$\mathcal{L}_D = \mathbb{E}[(D(\mathbf{x}_{\text{real}}) - 1)^2] + \mathbb{E}[D(\mathbf{x}_{\text{fake}})^2] \tag{7}$$

Generator Loss:

$$\mathcal{L}_G = \mathbb{E}[(D(G(\mathbf{z})) - 1)^2] \tag{8}$$

3.2 Optimization

Optimizer: Adam

Learning Rates (tuned per loss function):

- Generator (g_{lr}): 1×10^{-4} to 2×10^{-4}
- Discriminator (d_{lr}): 1×10^{-4} to 2×10^{-4}
- Learning rate combinations tested:
  - $g_{\text{lr}} = 2 \times 10^{-4}$, $d_{\text{lr}} = 2 \times 10^{-4}$ (balanced)
  - $g_{\text{lr}} = 1 \times 10^{-4}$, $d_{\text{lr}} = 2 \times 10^{-4}$ (slower generator)
  - $g_{\text{lr}} = 2 \times 10^{-4}$, $d_{\text{lr}} = 1 \times 10^{-4}$ (faster generator)
  - $g_{\text{lr}} = 1 \times 10^{-4}$, $d_{\text{lr}} = 1 \times 10^{-4}$ (conservative)
  - $g_{\text{lr}} = 3 \times 10^{-4}$, $d_{\text{lr}} = 3 \times 10^{-4}$ (aggressive)

Adam Hyperparameters:

- β_1: 0.5
- β_2: 0.999

Batch Size: 32 or 64 (hyperparameter tuned)
Training Duration: 300 epochs (full training) or 25 epochs (hyperparameter tuning)
Gradient Clipping: Applied to both networks (max_norm=1.0)

3.3 Hyperparameter Tuning Strategy

Two-Stage Approach:

Stage 1: Learning Rate Tuning

- Grid search over 5 learning rate combinations
- 25 epochs per configuration
- Metric: FID score (Fréchet Inception Distance)
- Select best learning rates for Stage 2

Stage 2: Architecture Hyperparameter Tuning

- Random search ($N = 10$ iterations)
- Parameters tuned:
  - Latent dimension: [100, 128, 256]
  - Number of layers (n_{layers}): [2, 3, 4]
  - Dropout: [0.1, 0.3, 0.5]
  - Batch size: [32, 64]
  - n_{critic}: [1, 2]
- 25 epochs per configuration
- Metric: Minimum FID score achieved
- Best configuration selected for final training

Final Training:

- 300 epochs with best hyperparameters
- Generate 3,000 synthetic malignant samples

3.4 Training Stability Techniques

1. **Spectral Normalization:** Constrains Lipschitz constant (for Hinge loss)
2. **Gradient Clipping:** Prevents exploding gradients
3. **Gradient Penalty:** Enforces 1-Lipschitz constraint (for Wasserstein)
4. **Dropout:** Regularization in discriminator (0.1–0.3)
5. **Progressive Monitoring:** Track mode collapse and vanishing gradients

4 Evaluation Metrics

4.1 GAN Quality Metrics

4.1.1 Fréchet Inception Distance (FID)

Purpose: Measures distributional similarity between real and generated images
Computation:

$$\text{FID} = \|\boldsymbol{\mu}_r - \boldsymbol{\mu}_g\|^2 + \text{Tr}(\boldsymbol{\Sigma}_r + \boldsymbol{\Sigma}_g - 2\sqrt{\boldsymbol{\Sigma}_r \cdot \boldsymbol{\Sigma}_g}) \tag{9}$$

where:

- $\boldsymbol{\mu}_r$, $\boldsymbol{\Sigma}_r$: Mean and covariance of real image features
- $\boldsymbol{\mu}_g$, $\boldsymbol{\Sigma}_g$: Mean and covariance of generated image features
- Features extracted from Inception V3 network

Interpretation: Lower is better (closer distributions)
Evaluation Frequency: Every 50 epochs
Sample Size: 256 images (real and fake)

4.1.2 Inception Score (IS)

Purpose: Measures quality and diversity of generated images
Computation:

$$\text{IS} = \exp\left(\mathbb{E}_{\mathbf{x}}\left[\text{KL}(p(y|\mathbf{x})\|p(y))\right]\right) \tag{10}$$

where:

- $p(y|\mathbf{x})$: Class probabilities from Inception network
- $p(y)$: Marginal class distribution

Interpretation: Higher is better (more diverse and confident)
Output: Mean ± standard deviation over 10 splits

4.1.3 Mode Collapse Detection

Method: Cosine similarity analysis of Inception features
Metrics:

- Mean similarity score
- Diversity score = 1 - mean_similarity
- Collapse threshold: 0.7

Interpretation: High similarity indicates mode collapse

4.1.4 Vanishing Gradient Detection

Method: Analyze loss trajectory changes
Window Size: 10 epochs
Threshold: 0.001
Metrics:

- Generator gradient magnitude
- Discriminator gradient magnitude

4.2 Classifier Performance Metrics

4.2.1 Primary Metrics

1. **Accuracy:** Overall classification accuracy
2. **F1-Score:** Harmonic mean of precision and recall (critical for imbalanced data)
3. **Recall (Sensitivity):** True positive rate for malignant detection
4. **Precision:** Positive predictive value
5. **ROC-AUC:** Area under ROC curve
6. **Confusion Matrix:** Detailed breakdown of predictions

4.2.2 Evaluation Scenarios

Baseline Classifier:

- Trained on original imbalanced dataset (10k benign, 1k malignant)
- Establishes performance ceiling with limited data

Augmented Classifier:

- Trained on baseline + 3,000 GAN-generated malignant samples
- Expected improvements:
  - Higher recall on malignant class
  - Better F1-score
  - Reduced overfitting

Comparison:

- Improvement in F1-score
- Change in recall/precision trade-off
- Reduction in false negatives (critical in medical domain)

4.3 Training Monitoring

Tracked Metrics per Epoch:

- Generator loss
- Discriminator loss
- D(real): Discriminator confidence on real images
- D(fake): Discriminator confidence on fake images

Visualization:

- Loss curves (generator vs discriminator)
- Discriminator confidence over time
- FID/IS score progression
- Sample image grids at checkpoints

Selection Criteria:

1. Minimum FID score (best match to real distribution)
2. Highest classifier F1-score improvement
3. Training stability (no mode collapse)
4. Visual quality assessment

5 Experimental Pipeline

5.1 Phase 1: Baseline Establishment

1. Train classifier on imbalanced baseline dataset
2. Evaluate baseline performance metrics
3. Document limitations (overfitting, poor minority class performance)

5.2 Phase 2: GAN Training

1. Implement DCGAN and cDCGAN architectures
2. Test multiple loss functions (Hinge, Wasserstein, BCE, MSE)
3. Perform hyperparameter tuning:
    - Learning rate optimization
    - Architecture parameter search
4. Train final models with best configurations
5. Generate 3,000 synthetic malignant samples per configuration

5.3 Phase 3: Augmented Training

1. Combine baseline + synthetic malignant samples
2. Train classifiers on augmented datasets
3. Evaluate performance improvements
4. Compare across different GAN variants

5.4 Phase 4: Domain Adaptation Evaluation

1. **Dataset Preparation:**
   - Source domain: Synthetic malignant + real benign (training)
   - Target domain: Real malignant + real benign (testing)
2. **Training Strategy:**
   - Train classifier exclusively on source domain
   - No access to real malignant samples during training
3. **Evaluation:**
   - Test on target domain (real malignant samples)
   - Measure domain gap: performance drop from source to target
   - Analyze failure modes and misclassifications
4. **Domain Gap Analysis:**
   - Compare source vs target accuracy
   - Per-class performance breakdown
   - Confusion matrices for both domains
   - Feature space visualization (t-SNE/UMAP)
   - Error analysis on domain-shifted samples
5. **Comparison Across GAN Variants:**
   - Which GAN loss produces most domain-robust synthetic data?
   - Correlation between FID score and domain gap
   - Trade-off between augmentation benefit and domain transfer

5.5 Phase 5: Analysis and Reporting

1. Statistical comparison of metrics across all scenarios:
    - Baseline (real imbalanced data)
    - Augmented (real + synthetic)
    - Domain shift (synthetic malignant training $\rightarrow$ real malignant testing)
2. Visual quality assessment
3. Domain adaptation analysis and insights
4. Final recommendations and conclusions