MSc Research Practicum

Enhancing Leukemia Diagnosis with Synthetic Data and Explainable Deep Learning Architectures

By *Ibrahim Malik (x23373385)*

National College of Ireland

MSc in Data Analytics

Research Context & Motivation

Critical Healthcare Challenge

- **Leukemia:** Blood cancer affecting 400,000+ annually
- **Current diagnosis**: Invasive bone marrow biopsies (painful, risky)
- **Expert dependency**: Requires specialized hematopathologists
- **Class imbalance**: 2.12:1 cancer-to-normal cell ratio in datasets
- **Clinical risk**: False negatives can be life-threatening

Our Solution

- **Non-invasive**: Microscopic blood smear analysis
- **Balanced datasets**: GAN-generated synthetic images
- **Explainable**: Transparent AI decision-making
- **Improved sensitivity**: 167% increase (0.187->0.500)

Literature Review & Research Gap

Study	Method	Best Accuracy	XAI Technique	Key Limitations
Giammarco et al. (2024)	CNN Ensemble + CAM	94%	Grad-CAM + CAM	Partial transparency, limited clinical testing
Nunna et al. (2024)	VISTA (ViT + CNN)	99.96%	Not Reported	Generalizability issues, no XAI
Raghaw et al. (2024)	CoTCoNet (CNN + ViT + Graph)	98.94%	Grad-CAM only	Limited clinical validation, single XAI method
Inturu et al. (2024)	Hybrid CNN-ViT	N/A	SHAP	Computational limitations, inconsistent data

Existing Approached - Limitations

- **CNN-only models**: Limited global context, ~70-94% accuracy
- **Pure ViTs**: Underperform on small medical datasets
- **Basic GANs**: Mode collapse, poor quality synthetic data
- **Black-box predictions**: No clinical interpretability
- **Single architecture focus**: Missing comparative analysis

Literature Review & Research Gap

Our Novel Contribution

- **First comprehensive framework** combining GANs + CNNs + ViTs + XAI
- **Architecture-specific GAN optimization**: Different GANs for different models
- **Rigorous validation**: FID (1.1897), SSIM (0.8869), IS metrics
- **Multi-level explainability**: Grad-CAM + SHAP + LIME integration
- **Clinical focus**: Prioritizing sensitivity over accuracy

Key Innovation: *Discovery that optimal GAN variants differ by architecture (CNNs prefer cGAN, ViTs prefer WGAN)*

Research Objectives & Methodology

Primary Research Question

> ***"To what extent can the combination of GANs, CNNs, Vision Transformers, and Explainable AI improve accuracy and interpretability of leukemia classification on microscopic blood smear images?"***

Research Design

- **Dataset**: C-NMC (13,000+ pediatric leukemia images)
- **Experimental approach**: Controlled comparison across architectures
- **Validation**: 3-fold cross-validation + separate test set

Research Objectives & Methodology

Quantitative Scope

Component	Variants Tested	Key Metrics
GANs	3 (DCGAN, WGAN, cGAN)	FID, SSIM, IS
Classifiers	4 (CNN, ViT, Hybrid, Baseline)	Accuracy, Sensitivity, AUC
XAI Methods	3 (Grad-CAM, SHAP, LIME)	Clinical relevance
Synthetic Images	1,267 generated	Quality validation

Technical Architecture - GAN Implementations

DCGAN Configuration

- **Architecture**: Generator (7 Conv2DTranspose layers), Discriminator (2 conv layers)
- **Training**: 225 epochs, batch_size=32, lr_disc=0.0002, lr_gen=0.0001
- **Results**: FID=1.2745, SSIM=0.8604, stable training

WGAN Configuration

- **Innovation**: Wasserstein loss + gradient penalty (λ=10)
- **Training**: 200 epochs, 5 critic iterations per generator update
- **Results**: FID=2.9208, SSIM=0.8685, improved diversity

cGAN Configuration

- **Key Features**: Class-conditional generation (0=healthy, 1=leukemic)
- **Training**: 250 epochs, label embedding, different learning rates
- **Results**: *Best quality* - FID=1.1897, SSIM=0.8869

Critical Finding: *cGAN produced highest quality images but CNNs performed best with cGAN while ViTs performed best with WGAN.*

CNN vs ViT vs Hybrid Analysis

CNN Architecture Performance

- **Architecture:** Conv2D(32->64->128) + Dense(128) + Dropout(0.5)
- **Best with:** cGAN augmentation
- **Results:** 71% accuracy, 26.7% sensitivity, 96.8% specificity

Vision Transformer Performance

- **Architecture:** 16×16 patches, 4 attention heads, 4 transformer blocks
- **Best with:** WGAN augmentation (counterintuitive!)
- **Results:** 74% accuracy, 50.0% sensitivity, 86.3% specificity

Hybrid CNN-ViT Results

- **Architecture:** CNN features -> ViT processing -> Element-wise fusion
- **Performance:** Consistently underperformed individual models
- **Best:** 68% accuracy with DCGAN (architectural mismatch)

Key Insight: *Hybrid architectures showed competing optimization objectives between convolutional and attention mechanisms*

Breakthrough Results - Sensitivity Improvements

Critical Clinical Metric: Sensitivity (True Positive Rate)

Model	Original Dataset	Best GAN Augmentation	Improvement
CNN	19.4%	26.7% (cGAN)	+37.6%
ViT	18.7%	**50.0%** (WGAN)	**+167%**
Hybrid	18.7%	48.2% (DCGAN)	+158%

Clinical Significance

- **Before:** Missing 4 out of 5 leukemia cases
- **After:** Detecting 1 in 2 leukemia cases
- **Impact:** Dramatically reduces life-threatening false negatives
- **Trade-off:** Modest specificity decrease (acceptable for screening)

Why This Matters: *In cancer diagnosis, missing a positive case (false negative) is far more dangerous than a false positive*

Novel Discovery - Architecture-GAN Interactions

Unexpected Finding: GAN Effectiveness Varies by Architecture

Architecture	Best GAN	Why?
CNN	cGAN	High-quality structured images match CNN's spatial hierarchies
ViT	WGAN	Global attention mechanisms robust to WGAN's morphological anomalies
Hybrid	DCGAN	Requires more structured/stable synthetic data for convergence

Contradicts Conventional Wisdom

- **Expected:** Higher quality synthetic data -> better performance
- **Reality:** Architecture compatibility matters more than image quality
- **Implication:** Must match GAN characteristics to model architecture

Research Impact: *First study to systematically demonstrate architecture-specific synthetic data preferences*

Explainable AI - Making Black Boxes Transparent

Grad-CAM Results

- **Focus areas:** Cell boundaries, nuclear irregularities
- **Clinical relevance:** Matches pathologist attention patterns
- **Quality metric:** Higher confidence -> more focused heatmaps

SHAP Analysis

- **Healthy cells:** Concentrated importance (1e-5 to 8e-5 range)
- **Cancer cells:** Distributed patterns (0 to 0.00012 range)
- **Insight:** Normal cells have consistent features, cancer cells show variability

LIME Explanations

- **Superpixel analysis:** 64 segments for healthy, 58 for cancer
- **Decision pattern:** Distributed feature importance (-0.1 to +0.1)
- **Clinical alignment:** Multiple morphological factors, not single markers

Clinical Validation: *XAI reveals models learning biologically relevant features, not artifacts*

Performance Comparison - Complete Results Table

Comprehensive Model Evaluation

Model	Dataset	Accuracy	Sensitivity	Specificity	F1-Score	AUC
CNN	Original	70%	19.4%	96.1%	60%	0.594
CNN	cGAN	**71%**	**26.7%**	**96.8%**	**65%**	**0.602**
ViT	Original	70%	18.7%	96.7%	55%	0.577
ViT	WGAN	**74%**	**50.0%**	**86.3%**	**73%**	**0.681**
Hybrid	Original	46%	18.7%	96.7%	43%	0.658
Hybrid	DCGAN	**68%**	**48.2%**	**77.7%**	**66%**	**0.661**

Key Takeaway: *GAN augmentation consistently improves sensitivity while maintaining clinical viability*

Clinical Impact & Translation Potential

Clinical Benefits	Implementation Pathway
● **Non-invasive:** Replaces painful bone marrow biopsies ● **Rapid Diagnosis:** 22ms inference time per image vs. hours for lab analysis ● **Standardized:** Consistent results across institutions ● **Accessible:** Reduces need for specialized hematopathologists ● **Scalable:** Can process thousands of images daily	● **Regulatory Approval:** FDA/CE marking for medical devices ● **Clinical Validation:** Multi-center validation studies ● **Integration:** PACS/LIS system connectivity ● **Training:** Healthcare professional education programs ● **Deployment:** Phased rollout in cancer centers

This research demonstrates the potential of AI to revolutionize cancer diagnosis, making it faster, more accurate, and less invasive for patients.

Limitations & Future Research

Current Limitations

- **Single dataset:** CNMC pediatric focus limits generalizability
- **Binary classification:** ALL vs. normal (not multi-subtype)
- **Synthetic artifacts:** Potential GAN-generated anomalies
- **Computational requirements:** GPU-intensive for real-time use
- **Clinical validation:** Requires extensive prospective studies

Future Research Directions - *Medium Term*

- **Multi-class:** ALL subtypes + AML + other leukemias
- **Foundation models:** Incorporate CLIP, SAM for semi-supervised learning
- **Real-world validation:** Multi-center clinical trials

Future Research Directions - *Long Term*

- **Federated learning:** Privacy-preserving multi-institutional training
- **Multimodal integration:** Combine with genomic, flow cytometry data
- **Regulatory approval:** FDA clearance pathway

Conclusion & Key Contributions

Research Question Answered

> ***"GAN-augmented Vision Transformers with explainable AI significantly improve leukemia diagnosis, achieving 167% sensitivity improvement while maintaining interpretability essential for clinical adoption."***

Novel Contributions:

1. **Architecture-GAN matching:** First systematic study showing CNN-cGAN, ViT-WGAN optimization
2. **Comprehensive XAI framework:** Multi-method interpretability (Grad-CAM + SHAP + LIME)
3. **Clinical-focused metrics:** Prioritizing sensitivity over accuracy for cancer screening
4. **Rigorous synthetic validation:** FID, SSIM, IS quality assessment
5. **Hybrid architecture analysis:** Show fusion challenges in CNN-ViT combinations

167%	74%	22ms	1,267
sensitivity improvement	peak accuracy	Inference time	synthetic images

THANK YOU