

GramMorph: Comparing CNN, Vision Transformer and Hybrid Models for Positive Blood Culture Gram-Stain Classification

A comparative study on a new multi-class Gram stain dataset

William Roach^{1,2}, Christopher Lawrence¹, Sokratia Georgaka²,
Magnus Rattray², and Victoria McCune¹

¹University Hospitals Tees, ²University of Manchester



<https://github.com/wroach132/gram-morph>

1. Introduction

Positive blood culture Gram stains provide rapid preliminary information for bloodstream infection management but require expert manual interpretation. Deep learning may support automated morphology classification. CNNs capture local image features, Vision Transformers model global context, and hybrid models may combine both.

This study directly compares CNN (ConvNeXt), ViT (UNI), and hybrid approaches for Gram-stain classification from positive blood cultures.

2. Dataset

Gram stain data set

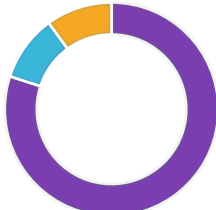


10,982 images
Five classes

Class Distribution (Internal Test Set)

Core class	Count
Gram positive cocci in clusters	5,071
Gram negative rods	3,864
Gram positive cocci in chains	1,499
Gram positive rods	415
Yeast	133
Total	10,982

Data Split



3. Methods

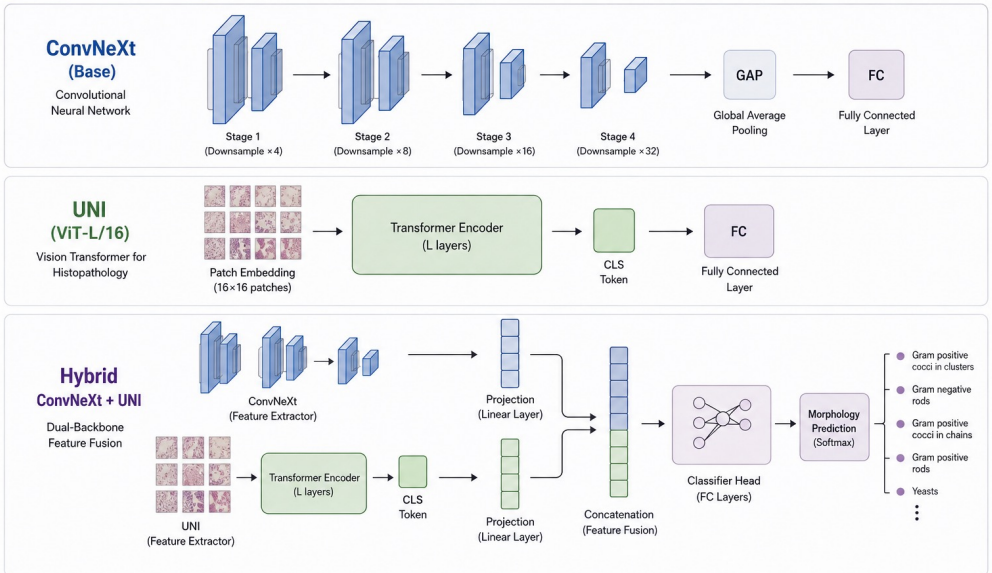
MODELS COMPARED

	ConvNeXt CNN (CNN-only baseline)	CNN-only baseline using ConvNeXt to learn local morphology, texture and staining features from PBC Gram-stain crops.
	UNI ViT (Vision Transformer)	Pretrained Vision Transformer feature extractor evaluating transfer of global contextual representations to Gram-stain morphology classification.
	Hybrid ConvNeXt + UNI	Feature-level fusion model combining ConvNeXt local features with UNI contextual features. Projected features were concatenated before classification.

TRAINING DETAILS

	Input size: 224 × 224 px
	Optimiser: AdamW
	Batch size: 32
	Augmentation: RandAugment, MixUp, CutMix
	Early stopping: validation macro F1
	Hybrid strategy: ConvNeXt partially unfrozen after epoch 5; UNI frozen

Model Architectures



GAP: Global Average Pooling FC: Fully Connected Layer CLS: Classification Token L: Number of Transformer Layers

4. Results/ Discussion

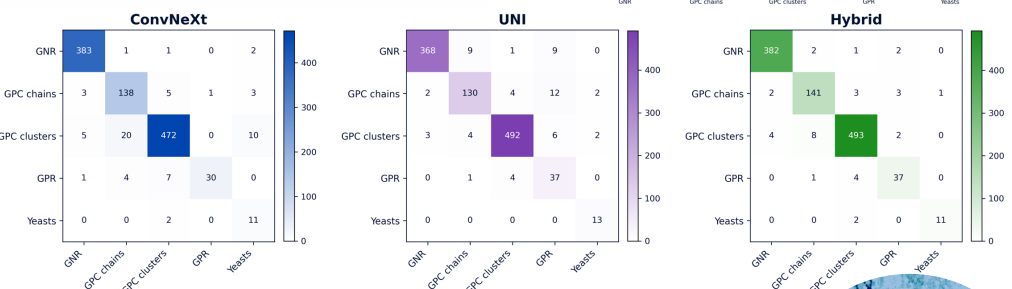
Internal Test Performance Summary

Model	Accuracy (%)	Macro F1-score	Macro AUC (ROC)	MCC
CNN (ConvNeXt)	94.09	0.8402	0.9917	0.9093
ViT (UNI)	94.63	0.8787	0.9934	0.9179
Hybrid (ConvNeXt + UNI)	96.82	0.9273	0.9961	0.9506

Internal Test Per-Class Macro F1 by Model

Model	Gram negative rods	Gram positive cocci in chains	Gram positive cocci in clusters	Gram positive rods	Yeasts
CNN (ConvNeXt)	0.9833	0.8818	0.9497	0.8219	0.5641
ViT (UNI)	0.9684	0.8844	0.9762	0.6981	0.8667
Hybrid (ConvNeXt + UNI)	0.9858	0.9338	0.9762	0.8605	0.8800

AUC = Area Under the ROC Curve MCC = Matthews Correlation Coefficient



Hybrid feature fusion produced the strongest internal classifier, suggesting complementary value from CNN local morphology features and UNI transformer representations. Performance was high for common morphology classes, but minority classes remained more challenging, particularly Gram-positive rods and yeasts.

6. Conclusion

With further external validation, this approach could support diagnostic workflows by reducing repetitive manual review, flagging uncertain cases, and providing an additional check to reduce reporting variation or error.

Future work:

- External validation,
- Improving minority-class performance
- Developing subclassification of Gram-positive cocci in chains (Pneumococcus, Enterococci, viridans group streptococci, and haemolytic streptococci).

