Deep Learning, Small Data Pathology Classifiers

OVERVIEW

While machine learning, particularly deep learning, excels in fields with large datasets such as computer vision, the situation in medicine is different. Medical imaging datasets, such as orthopedic X-Rays and MRIs, typically contain only hundreds to thousands of labeled images. This limited dataset size makes it challenging to effectively employ Convolutional Neural Networks (CNNs) for medical image classification.

PROBLEM

- Medical datasets are small, imbalanced, and of a general poor quality.
- Models trained on these datasets often have poor classification accuracies
- A key focus is on exploring which machine learning tool presents a robust and effective solution to combat the dataset issue.
- This exploration will allow us to harness the potential of machine learning in medical applications.

OBJECTIVES

Our research goals are divided into three main objectives, each encompassing its own subset of specialized aims:

- Evaluate the efficacy of top DL models for medical image recognition and identify the most efficient one.
- Assess the ability of GANs to produce synthetic images from limited pathological datasets and recommend the best GAN setup.
- Examine the performance of leading NAS techniques in designing CNNs for binary medical image classification, aiming to pinpoint the optimal NAS approach and its relevance in this domain.

SUPERVISED LEARNING

Supervised Learning uses labeled data for neural network training. Data Augmentation (DA) expands small datasets. Transfer Learning (TL) applies insights from one model to another.

We formulate a multi-label classification problem and test on two medical datasets with permutations of the following experimental variables:

1. **Architectures**: ConvNeXt, Swin-T, DenseNet, Resnet, EffNet
2. **DA Methods**: No DA, Cropping, RandAug, Neural Augment
3. **TL Methods**: No TL, Pretraining and Finetuning

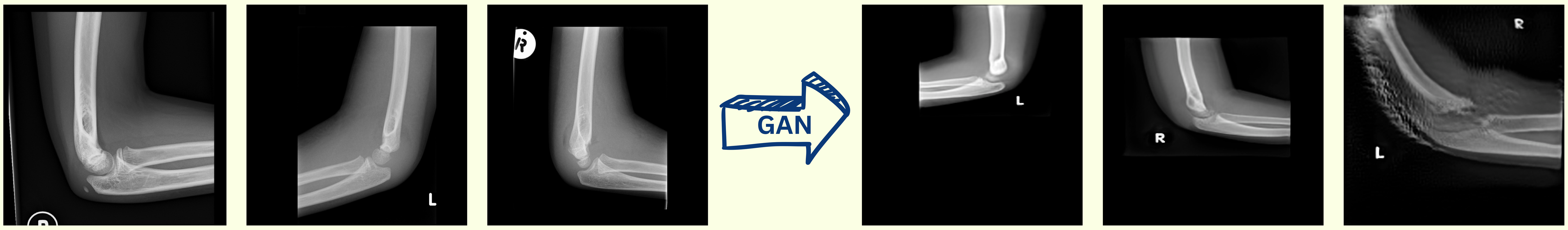
Key Findings:
We conclude that ConvNeXt is the superior architecture, Neural Augment is the most effective DA method and Transfer Learning is invaluable for classification tasks.

Elbow Top Models	Accuracy	Macro Avg.	Precision	Recall	F1	AUC	Hamming
ConvNeXt-B with TL and Neural Aug	**30.00**	**88.66**	**49.02**	**40.00**	**44.05**	**67.39**	**11.34**
ResNet50 with TL and Neural Aug	27.50	88.57	48.37	35.60	41.01	65.41	11.43
DenseNet121 with TL and Random Cropping	27.50	86.88	36.25	23.20	28.29	59.04	13.12

Neck Top Models	Accuracy	Macro Avg.	Precision	Recall	F1	AUC	Hamming
DenseNet121 with TL and Neural Aug	**48.44**	**83.20**	**62.50**	**39.22**	**48.19**	**66.68**	**16.80**
ResNet152 with TL and Neural Aug	48.44	80.08	56.25	35.29	43.37	64.23	18.36
ConvNeXt-T with TL and Neural Aug	46.88	83.20	62.50	39.22	48.19	66.68	16.80

Top model performances, with Default Validation

Conclusions:
We find that class imbalancement is a more detrimental issue than shear dataset magnitude and subsequently develop **Default Validation**, which significantly increases reported accuracy.



NEURAL ARCHITECTURE SEARCH

Neural Architecture Search (NAS) automates the creation of optimal neural network architectures. This research delves into the potential of NAS to enhance Convolutional Neural Networks (CNNs) for medical image classification, especially in the context of sparse datasets.

Key Findings:

1. **State-of-the-Art NAS Models:** The study evaluated NAS models like DeepMAD, ShapleyNAS, and ZenNAS on pathological datasets of the neck and elbow.
2. **Performance Insights:** ZenNAS excelled in Cervical Neck X-Rays, while ResNet50 outperformed NAS for Elbow X-Rays. Our pretrained ResNet achieved 80% in weighted TPR and FPR, compared to radiologists' median of 87%.
3. **Real-World Application:** Elbow dataset models showcased potential as first-readers in resource-constrained environments.

Conclusions:
The research underscores the capabilities of NAS in medical image classification, emphasizing the importance of dataset-specific model selection and optimization. The findings suggest that NAS holds significant potential to revolutionize healthcare diagnostics, especially in addressing challenges posed by data scarcity.

GENERATIVE ADVERSARIAL NETWORKS

Generative Adversarial Networks (GANs) use a 2-neural network architecture to create *fake* images for the creation of *fake* images that closely resemble original images. This research looks into the effectiveness of GANs on small pathology datasets by contrasting Vanilla GAN, WGAN-GP and StyleGAN2.

Key Findings:

1. **Vanilla GAN** often experiences mode collapse and other training instabilities producing unrecognisable images.
2. **WGAN-GP** regularization and an improved loss function stabilizes training however visual quality is still poor.
3. **StyleGAN2** produces high-quality imagery but often overfits on a small dataset.
4. **MedFID** we propose a medical adaption of FID scores trained on classifying x-ray images.

Conclusions:
StyleGAN2 creates high-quality imagery on small datasets with shorter training times than typical StyleGAN2 tasks. However, it tends to capture minority elements in inconsistent datasets. We show that high-quality synehtic images are possible but is limited by dataset inconsistencies.

The Team:
Winner Kazaka | KZKWIN001@myuct.ac.za
Tomas Slaven | SLVTOM001@myuct.ac.za
Shaylin Chetty | CHTSHA042@myuct.ac.za
Supervisor: Geoff Nitschke | geoff.nitschke@uct.ac.za

Department of Computer Science
University of Cape Town
www.cs.uct.ac.za
Tel: 021 650 2663

National Research Foundation