Classification of Tuberculosis using Deep Learning: Comparative Analysis of CNN Models

Muhumuza Joshua [1] Joyce Nabende Andrew Katumba

[1]Makerere University

Introduction

Early detection of Tuberculosis (TB) is pivotal to effectively managing and curbing the spread of this pervasive health concern that affects millions of people worldwide. This paper delves into the application of Convolutional Neural Networks (CNN) for the classification of TB from chest X-ray images. We offer a comprehensive comparative analysis of multiple deep learning models, including plain CNN, VGG-16, DenseNet-121, ResNet, Inception, and Xception models. These investigations were conducted on the EUCREI and DOHA datasets. The results show that the Inception and Xception models exhibited exceptional performance in the classification task

Datasets

Our research utilized the EUCREI and DOHA datasets, with each dataset being subject to similar preprocessing and augmentation procedures. The Ernest Cook Ultrasound and Research Education Institute (EUCREI) is a Uganda based leading center at tertiary level in the training of Ultrasound, Medical Imaging and other health-related courses in Sub Saharan Africa[1] . This data has two main categories of TB patients either sick or normal.

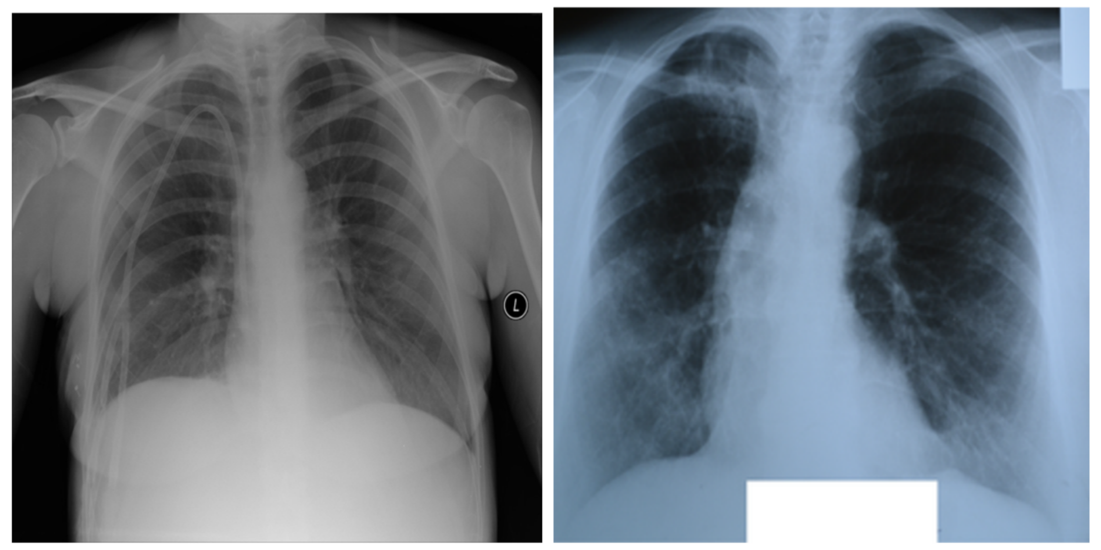

Figure 1. A normal and TB chest x-ray image

- **EUCREI** 907 chest x-ray images, TB patients categorized as sick or normal
- **DOHA** 3674 normal images and 700 sick images

The DOHA dataset was compiled by a team of researchers from Qatar University, Doha, Qatar, and the University of Dhaka, Bangladesh along with their collaborators from Malaysia in collaboration with medical doctors from Hamad Medical Corporation and Bangladesh. It includes 3674 normal images and 700 sick images[2]

Methodology

A methodology from image preprocessing, augementation, model evaluation and dataset mixing was followed

- **Image Preprocessing** Image preprocessing was employed to maintain consistency across the images and reduce computational complexity. The preprocessing steps included resizing, grayscale conversion, contrast enhancement, normalization, and data type conversion.
- **Data Augementation** To address class imbalance, data augmentation techniques were employed. These techniques, including gamma contrast adjustment, additive Gaussian noise, and brightness multiplication.
- **Model training and Evaluation** Our experiments utilized both random and stratified k-fold cross-validation for model training and evaluation. Each model's performance was assessed based on accuracy, precision, recall, and the F1-score.

The two datasets were then mixed and the above steps repeated.

Evaluation

We evaluated six different CNN models on the EUCREI and DOHA datasets using different sampling techniques. All the other CNNs apart from the plain CNN, were implemented using transfer learning where each of the model's top layers were customized for binary classification

1. **Random Sampling**, Under random sampling, we selected a sample of images for training, testing and validation from the eucrei dataset. Each individual was chosen entirely by chance using a train,test, validation ratio of 80:15:5.
2. **Stratified Sampling**. The datasets were partitioned into stratas using a k-fold of 2. We randomly selected and independently sampled from each stratum, maintaining the proportion of each category in the dataset. This helped us guarantee that each category is adequately represented in the sample, reducing sampling error and bias.We then evaluated the two sampling methods and their respective effect on overfitting

Model comparisons

The plain CNN model achieved a high accuracy of 99% on the EUCREI dataset and a moderate 63% on the DOHA dataset with random sampling. However, precision, recall, and F1-score results suggest a potential overfitting issue, which was substantiated when employing stratified sampling resulted in reduced accuracies (EUCREI: 62%, DOHA: 67%). The VGG-16 model demonstrated robust performance on the EUCREI dataset (accuracy: 98%), but performance dropped on the DOHA dataset (accuracy: 68%). Notably, stratified sampling also led to decreased performance, suggesting possible overfitting

MODEL	Accuracy		Precision		Recall		f1-score	
	EUC	DOH	EUC	DOH	EUC	DOH	EUC	DOH
CNN	0.62	0.67	0.46	0.82	0.46	0.76	0.46	0.81
VGG-16	0.88	0.51	0.77	0.95	0.76	0.44	0.76	0.54
RESNET	0.76	0.84	0.77	0.88	0.75	0.88	0.75	0.88
DENSENET-121	0.84	0.64	0.77	0.82	0.76	0.76	0.76	0.78
INCEPTION	0.83	0.84	0.77	0.88	0.74	0.88	0.75	0.88
XCEPTION	0.85	0.82	0.80	0.89	0.75	0.88	0.75	0.88

Figure 2. Model performance using stratified sampling

The ResNet model demonstrated lower performance compared to other models, particularly under stratified sampling (accuracy: EUCREI: 76%, DOHA: 46%), while the Inception and XCeption models demonstrated superior performance, particularly on the DOHA dataset with stratified sampling (Inception: 84%, XCeption: 82%). On the mixed dataset, Xception's performance (accuracy: 88%, precision: 89%, recall: 87%, F1-score: 88%) exceeded that of Inception (accuracy: 83%, precision: 77%, recall:74%, F1-score:75%)

Mixing the datasets

The datasets were then mixed and performance evaluated on both inception and Xception models. This provided a new set of images: 935 normal and 935 Tuberculosis images. The same metrics were used

Confusion matrix upon mixing

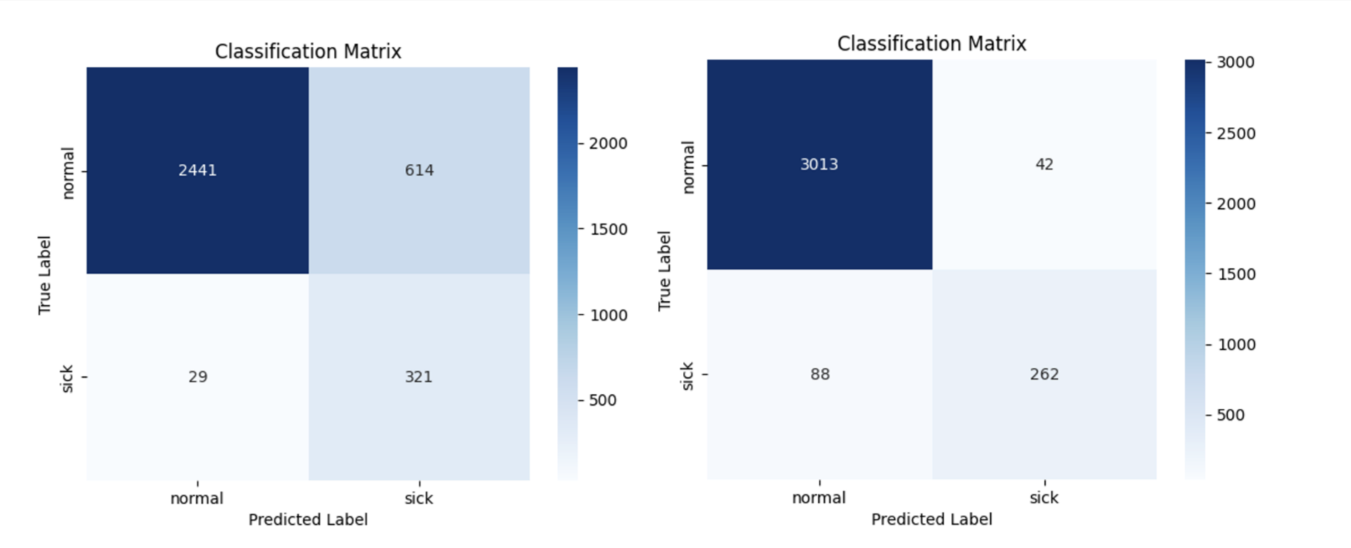

Figure 3. Inception and Xception on an unknown dataset respectively

Discussion

Inception and Xception outperformed other models, especially on the mixed dataset. On an unknown dataset, Xception achieved 96% accuracy, while Inception achieved 81%.

MODEL	Accuracy	Precision	Recall	f1-score
INCEPTION	0.83	0.77	0.74	0.75
XCEPTION	0.88	0.89	0.87	0.88

Figure 4. Model performance using stratified sampling

This study highlights the potential of deep learning models, particularly the Xception model, in assisting radiologists with TB diagnosis from chest X-ray images. Future research should focus on addressing the limitations related to overfitting in some models, potentially through further data augmentation, regularization techniques, or advanced model architectures

References

[1] https://ecurei.ac.ug/

[2] Tawsifur Rahman, Amith Khandakar, Muhammad A. Kadir, Khandaker R. Islam, Khandaker F. Islam, Zaid B. Mahbub, Mohamed Arselene Ayari, Muhammad E. H. Chowdhury. (2020) "Reliable Tuberculosis Detection using Chest X-ray with Deep Learning, Segmentation and Visualization". IEEE Access, Vol. 8, pp 191586 - 191601. DOI. 10.1109/ACCESS.2020.3031384

 muhumuzajoshua95@gmail.com