A Self-Supervised Hybrid Deep Learning Framework for Multi-Class Chest X-ray Classification on Imbalanced Data

Pranto Saha, Sadman Hafij

Daffodil International university

ABSTRACT

This project proposes a hybrid deep learning framework for multi-class chest X-ray classification in low-resource settings.

The model will learn from unlabeled data using SimCLR's self-supervised learning feature before being refined on a tiny labeled sample. A hybrid encoder combining MobileNetV2, InceptionV3, and VGG16, along with LSTM and attention mechanisms, is designed to capture rich and contextual features.

Relevant image regions will be highlighted using interpretability techniques in an effort to provide a diagnostic tool that is both effective and understandable..

CONTACT

Pranto Saha

Dept. of CSE

Email: saha15-4157@diu.edu.bd

INTRODUCTION

Millions of people die each year from respiratory illnesses such pneumonia, TB, and COVID-19; pneumonia alone accounts for 15% of all child fatalities under the age of five worldwide. Even though chest X-rays (CXRs) are frequently used to diagnose these conditions, diagnostic mistake rates are still high, especially in areas with limited resources. This is because of things like radiologist fatigue and a lack of adequate annotated data. Although there are obstacles like class imbalance and a lack of labeled datasets, deep learning—in particular, convolutional neural networks, or CNNs—has demonstrated potential in automating CXR analysis.

By fusing supervised fine-tuning for multi-class CXR classification with self-supervised contrastive learning, our approach presents a hybrid deep learning system to handle these problems. To improve feature representation and classification performance, the system makes use of a new encoder that combines several CNN architectures with an LSTM layer and attention mechanism. The model uses explainable AI approaches to increase interpretability and transparency. Comparative analysis across several CXR datasets under uniform training settings will confirm its efficacy.

METHODS AND MATERIALS

Our approach integrates a hybrid encoder combining MobileNetV2, InceptionV3, and VGG16 with sequential modeling via LSTM and an attention mechanism, initially learning robust representations through SimCLR-based contrastive pretraining on a large unlabeled dataset. This pipeline effectively addresses the challenges of limited labeled data while enhancing feature extraction and diagnostic precision.

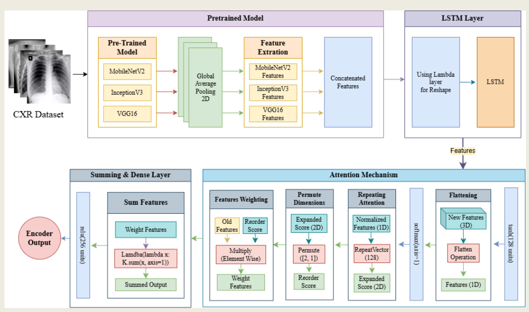

Fig.1 Encoder architecture of proposed model

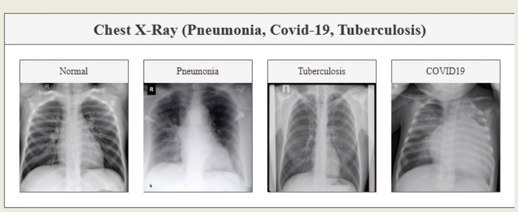

Fig. 2: Samples of CXR (Pneumonia, Covid-19, Tuberculosis)

RESULTS

Our proposed framework, combining multi-scale CNN feature extractors with LSTM and attention mechanisms, is expected to benefit significantly from SimCLR-based self-supervised pretraining. Preliminary evaluations, including ablation studies, hyperparameter tuning, and k-fold cross-validation, suggest that this hybrid approach will offer robust performance, enhanced computational efficiency, and improved explainability.

The framework is anticipated to be particularly effective for clinical deployment in resource-limited settings, providing scalable solutions for multi-class chest X-ray classification.

DISCUSSION

The proposed hybrid deep learning architecture efficiently handles the problems of multi-class CXR classification, such as class imbalance and insufficient labeled data. In order to enhance feature representation and performance, the model combines several CNN architectures with LSTM and attention techniques. By reducing the need for labeled data, SimCLR-based pretraining increases the framework's adaptability for medical imaging.

While ablation tests demonstrate the importance of each model component, the computational overhead of hybrid encoders and LSTM may limit scalability in low-resource environments. Efficiency optimization and model evaluation using external clinical datasets should be the main goals of future research. Confirming the attention mechanism's clinical efficacy also requires expert evaluation.

CONCLUSIONS

This approach presents a hybrid deep learning system for CXR image categorization that combines pre-trained encoders, LSTMs, and attention methods. Through the use of SimCLR-based self-supervised learning, the system seeks to improve performance and interpretability by learning robust features from unlabeled datasets, hence addressing the problem of limited labeled data.

Future research will concentrate on enhancing the framework's effectiveness, verifying it using outside clinical datasets, and investigating how to incorporate it into actual clinical procedures. These initiatives aim to improve early diagnosis of lung illnesses and assist radiologists.

REFERENCES

[1] "FastStats," https://www.cdc.gov/nchs/fastats/pneumonia.htm.

[2] "Coronavirus Disease (COVID-19) Situation Reports," https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports.

[3] "Global Health Estimates."

[4] "Pneumonia in Children Statistics - UNICEF DATA," https://data.unicef.org/topic/child-health/pneumonia/.

[5] "Global Tuberculosis Report 2024."