Enhancing Lung Diseases Recognition Through CNN-RNN Methodologies

by

Md. Fardin Ahsan
20101208
Ramisa Anan Orni
20301363
Israt Ara Zahin
20301383
Adiba Hossain
20301344
Muntaha Tabassum
20101378

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
Brac University
February 2025

Declaration

It is hereby declared that

1. The thesis submitted is our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Ahsan

Md. Fardin Ahsan
20101208

Israt Ara Zahin
20301383

Adiba

Adiba Hossain
20301344

Orni

Ramisa Anan Orni
20301363

Muntaha Tabassum

Muntaha Tabassum
20101378

Approval

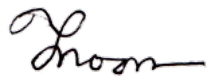
The thesis titled "Enhancing Lung Diseases Recognition Through CNN-RNN Methodologies" submitted by

1. Md. Fardin Ahsan (20101208)
2. Ramisa Anan Orni (20301363)
3. Israt Ara Zahin (20301383)
4. Adiba Hossain (20301344)
5. Muntaha Tabassum (20101378)

Of Fall, 2024 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on February 03, 2025.

Examining Committee:

Supervisor:
(Member)



Jannatun Noor
Associate Professor, CSE
Director, BSDS
United International University

Thesis Co-ordinator:

Md. Golam Rabiul Alam, PhD
Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Sadia Hamid Kazi, PhD
Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Abstract

Diagnostics of respiratory disorders greatly benefit from medical imaging, especially X-ray imaging, which offers important information about the anatomical anomalies of the lungs. As we explore deeper into the field of lung illness recognition, it becomes clear that using multiscale Deep Convolutional Neural Network techniques has the potential to transform the detection of pneumonia and tuberculosis from X-ray pictures. In this paper, we will classify images through a process that requires only chest-xray images. We have proposed a Deep Learning (DL) based algorithm for lung disease detection, which we termed as Convolutional Recurrent Network (C-RNet). In our research, we classify chest X-ray images into four categories according to the publicly available dataset. Our proposed model can calculate the dependency and continuity properties of the intermediate layer output very precisely. At the same time, the features of these intermediate layers can be combined with the final fully-connected network for classification prediction, resulting in better classification accuracy. We have explored the potential of combining CNN and RNN with XAI to identify lung diseases from chest radiographs to improve diagnostic accuracy compared to traditional single-scale methods. Upon comparing our suggested model with the current models, we discovered that, with an accuracy of 93.51% on the full dataset, our suggested model achieved the best accuracy of all the architectures we compared. Moreover, our suggested model C-RNet was observed to accurately categorize and detect the regions of disease through approaches such as Grad-CAM.

Keywords: Lung diseases, X-ray, CNN, RNN, XAI, Grad-CAM, LSTM.

Acknowledgement

Firstly, all praise be to Almighty Allah (SWT) for whom our thesis has been completed without any major interruption. Secondly, to our thesis supervisor, Dr. Jannatun Noor Mukta ma'am for her humble cooperation and guidance in our research. She has helped us enormously whenever the need arose. She was cooperative and helped us troubleshoot throughout all the complexities that we faced. Finally, without our parents' unwavering support, it may not be conceivable. We are currently on the verge of graduating for which we are thankful to their kind support and prayer.

Table of Contents

List of Figures

List of Tables

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

$Bi - LSTM$ Bidirectional Long Short-Term Memory

CNN Convolutional Neural Network

CXR Chest X-ray

$Grad - CAM$ Gradient-weighted Class Activation Mapping

$LSTM$ Long Short Term Memory

$ReLU$ Rectified Linear Unit

RNN Recurrent Neural Network

TB Tuberculosis

ViT Vision Transformer

XAI Explainble Artificial Intelligence

Chapter 1

Introduction

1.1 Motivation

Lung diseases rank among the most infectious health conditions and have become the leading causes of mortality worldwide. Developing countries like Bangladesh, are particularly vulnerable to these diseases due to environmental hazards. These diseases stem from various causes. Exposure to pollutants , dust, chemicals, infections, and smoking can cause lung diseases like pneumonia, asthma, tuberculosis, influenza, COVID, and other chronic obstructive pulmonary disease. Due to the rising of air-pollution, the number of respiratory diseases is also increasing. Prolonged exposure to polluted air often leads to chronic respiratory conditions, weakening individuals over time. The overall number of patients suffering from respiratory and infectious diseases in Bangladesh has increased drastically over the previous five years [1]. Experts warned that the infection could be disastrous for individuals older than 50, since over 70% of Bangladesh's elderly population already has comorbid conditions and non-communicable diseases [1].

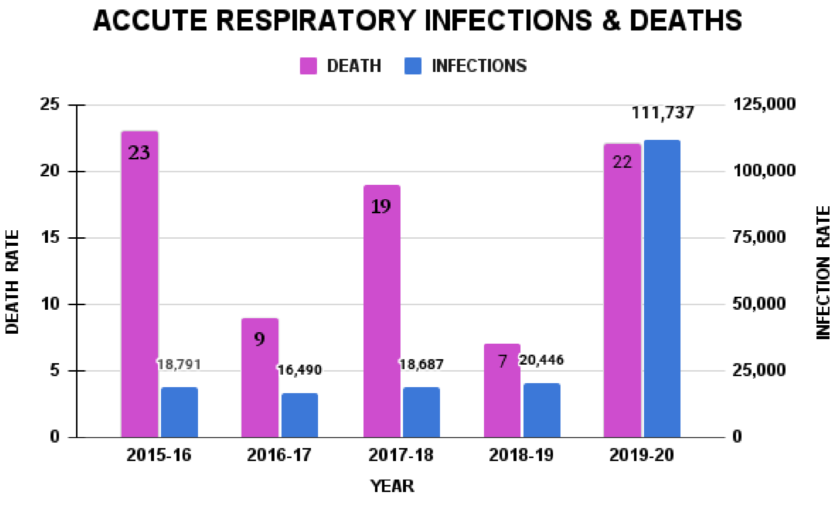


Figure 1.1: Increasing Rate Of Resperitory Diseases [1]

Pneumonia killed over 740,180 children in 2019 and is responsible for 14% of mortality among children below five. Annually, pneumonia claims the lives of about 12,000 children in Bangladesh under five [2]. In Bangladesh, pneumonia causes the deaths of about 24,000 children annually, making about 24% of all deaths among children under five, which is higher than the global scale average [3]. This means that two to three kids pass away from pneumonia per hour. Even with these concerning figures, only 60% of parents take their child to the doctor when they show signs of pneumonia. 40% of children do not obtain necessary treatment since there is a lack of awareness [3]. Moreover, lung diseases can interrupt the vital process of people's respiratory systems, causing uncomfortable symptoms, such as fatigue, wheezing, coughing, shortness of breath, and chest pain. The signs and symptoms of these diseases are quite identical. Hence, it is challenging to diagnose different diseases and critical to identify and treat them early to enhance patient outcomes and lessen the burden of these conditions on society. Addressing this growing health crisis requires urgent action, from reducing pollution levels to enhancing diagnostic techniques and treatment accessibility.

X-ray radiography is frequently the favored method for identifying lung diseases, as it is both feasible and inexpensive [4]. While CT scans provide higher detection sensitivity, X-ray radiography remains the preferred choice in clinical settings due to its affordability and widespread accessibility in general hospitals [5]. Traditional lung disease diagnosis relies mainly on radiologists ability to analyze radiological images. A reliable diagnosis might be difficult since symptoms and imaging characteristics can overlap among different lung diseases. Examining and analyzing medical images requires careful attention to precision in order to prevent important information from being overlooked. Furthermore, the increasing need for diagnostic services causes delays in testing and medication. These problems are made worse by the lack of qualified radiologists in many areas, which emphasizes the need for more effective diagnostic instruments. To assist radiologists and accelerate the identification process, researchers have built deep learning algorithms. In order to assess and classify the X-ray pictures according to unique patterns and characteristics linked to different lung diseases, these models make use of advanced machine learning techniques and neural networks.

Our suggested model also utilized advanced machine learning algorithms. We have developed our model using a number of techniques based on the application of CNN and RNN methodologies to CXR data. To train our C-RNet model, we have used an open-sourced dataset from Kaggle [6] consisting of CXR images containing cases of COVID-19, Lung Opacity, Normal and Viral Pneumonia. We integrated RNN methods like Bi LSTM techniques with CNN and used XAI methods like Grad-CAM to develop our model. Subsequently, we record every step of the procedure, including gathering data, preprocessing, training the model, evaluation and interpreting the outcomes. By creating a multiscale CNN, RNN and XAI customized model, C-RNet, our proposed model aims to bridge the gap between cutting-edge AI methodologies and real-world clinical requirements.

1.2 Problem Statement

Existing deep learning models for lung illness diagnosis still have serious drawbacks, despite notable developments in medical imaging and AI-driven disease classification. When it comes to extracting spatial data from chest X-ray pictures, CNNs have shown remarkable performance. Nevertheless, they frequently overlook contextual information and sequential dependencies that could improve diagnostic precision. When it comes to identifying minor illness progressions, when temporal or region-specific correlations are critical, this constraint is especially troublesome [7].

Additional issues including dataset imbalance, noise in X-ray pictures, and lack of interpretability further impede AI-based diagnosis in practical applications. Biased model predictions result from the excessive number of normal cases in large medical datasets relative to ill samples [8]. Furthermore, model generalization is challenging due to X-ray pictures' frequent artifacts, illumination fluctuations, and inconsistent scanning techniques across various healthcare facilities. A significant obstacle to clinical adoption of deep learning models is their explainability, since medical practitioners find them less reliable when they provide black-box predictions without explanation.

In order to overcome these difficulties, this study integrates sequential pattern recognition and spatial feature extraction to develop a hybrid CNN-RNN framework (C-RNet) for lung disease classification. By showing the model's decision-making process, the use of XAI approaches like Grad-CAM improves clinical interpretability and transparency. Our method seeks to provide greater accuracy, improved generalization, and increased reliability in practical medical applications by optimizing model architecture based on assessment metrics and training efficiency.

1.3 Research Contribution

This research provides a valuable solution through multi-scale CNN and RNN methodologies in order to identify lung diseases from X-ray images for developing countries, including Bangladesh, where there is a dearth of radiologists in the sector of medical, mainly in rural or remote areas. Multiscale CNN is utilized for effective feature collection at different dimensions to improve the identification and categorization of lung disorders and Bi-LSTM networks are integrated to identify spatial patterns and related dependencies for more precise disease identification. Moreover, the model's ability to utilize multiscale processing enhances its ability to distinguish and classify among multiple lung diseases, resulting in increased diagnosis accuracy when compared to conventional single-scale approaches. Furthermore, including XAI methods in the suggested model will enhance its credibility and comprehensibility. Grad-CAM were used to see and understand the process of how the model is making decisions.

The primary objective of this research is to develop an enhanced deep learning model that improves the accuracy of lung disease detection using chest X-ray images. To achieve this, we set the following specific objectives:

- We aim to develop an advanced patch-based augmentation technique that improves generalization across different imaging conditions without any metadata.
- Our model is designed with computational efficiency in mind, making it more suitable for real-world deployment in clinical settings.
- Implementing an optimal architecture selection strategy for C-RNet, balancing accuracy, training efficiency, and interpretability.
- Integrating explainable AI techniques like Grad-CAM to improve model transparency, aiding in clinical decision-making and trustworthiness.

C-RNet model will assist physicians in diagnosing patients quickly and accurately by giving them data that is easy to understand. Additionally, it will make sure that the model produces findings that are easy to read and help radiologists and other medical professionals comprehend the diagnostic observations which also increase dependability and assurance. This will increase the reliability of the process in the medical field as a whole. The main goal of this study is to increase the general efficacy and efficiency of accurately and efficiently identifying lung diseases. In addition to helping with the timely identification of all lung disorders, this will contribute to public health.

1.4 Thesis Oraganization

The remaining sections of this work are organized as follows. Chapter 2 provides an overview of relevant work, concentrating on existing CNN and RNN architectures, Grad-CAM approaches, and past fusion models while describing the reason for our research. Chapter 3 introduces CNN, RNN, and Grad-CAM, outlining their designs and applications. Chapter 4 discusses the study methodology, which includes the workflow, data description, preprocessing procedures, and the description of our C-RNet model. The experimental assessment is presented in Chapter 5, which covers the experimental setup, evaluation metrics, performance analysis, and results comparison, as well as Grad-CAM based visualization. Finally, Chapter 6 addresses key findings, highlighting the favorable aspects of our proposed model above others, contextualizing the findings, and offering recommendations for further research followed by the Bibliography section, which lists all the references utilized in this study.

Chapter 2

Literature Review

The goal of this review of the literature is to examine the developments in multi-scale deep learning approaches for detecting lung illness from chest X-ray pictures. We selected a few studies such as CNNs, RNNs, hybrid models, and XAI models in order to analyze the accuracy, efficiency, and readability of different models. We have attempted to figure out the main obstacles to lung disease recognition using deep learning and investigate potential solutions. Furthermore, we have chosen research papers published between 2022 and 2024 to provide a thorough and current study that takes into account the most recent developments and patterns in machine learning for image analysis in medical imaging. We have investigated papers that present new architectures, optimal methods, or upgrade procedures including multi-scale pattern extraction, data augmentation and transfer learning.

Islam et al. [9] carried out research to assess the feasibility of different architectures of the DCN on various types of abnormalities seen inside the frontal CXR images. Trial was done on chest X-rays from Shenzhen dataset, JSRT and the PCA center at Indiana University using the model. As stated by Pan and coworkers, using an ensemble of the DCN models give a higher accuracy rate than just using the individual models. However, combining two rule-based and DCN ensemble models reduced the accuracy. However, sensitivity-based localization gives out the right localization in the case of spatially disseminated sicknesses. Enhancing the accuracy by 17% points more than the previous techniques for cardiomegaly classification, the deep learning method yields the best accuracy for the SHENZHEN dataset in tubercular detection. However, their study through localization failed and remained limited to rounded contours such as a lung nodule or a bone fracture.

Islam et al. [10] offer a novel method to strengthen the identification and categorization of lung issues through the use of Explainable Artificial Intelligence (XAI) approaches in conjunction with a Deep Convolutional Neural Network and Gated Recurrent Unit (DCNN-GRU). Their suggested model uses the GRU to simulate the chronological correlations and linear details present in clinical imaging data and it makes use of the potent pattern retrieval abilities of DCNN to identify worldwide as well as local spatial abnormalities in lung CT images. Five convolutional layers with equivalent layers for pooling build up the DCNN module. These layers are used to collect multi-scale traits which are then processed by a GRU layer with 1024 neurons to find intricate periodic structures. A softmax layer for multi-scale categorization is used to generate the finalized identification, which differentiates between benign, malignant, and normal lung diseases. They have uses XAI techniques such as Gradient-weighted Class Activation Mapping (Grad-CAM), SHapley Additive exPlanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME) to enhance system readability and foster medical professional reliability. These methodologies point out important areas in the X-ray pictures which influences the predictions of the model. Two datasets were used to test the model: the COVID-19-CT dataset, which included 2,618 Community-Acquired Pneumonia (CAP) images, 7,593 COVID-19 images and 6,893 normal images and the IQ-OTH/NCCD lung cancer dataset which included 2,073 CT scans from 110 individuals. The DCNN-GRU model outperformed cutting-edge models like ResNet50, VGG16 and DenseNet121 with exceptional success rates of 98.97% for lung cancer identification and 99.30% for COVID-19 identification. Additionally, because of the effective GRU layer the model reduced learning time and showed higher efficiency across a variety parameters including precision, recall, F1-score and Cohen's Kappa. Regardless of XAI integration the trained model still has limitations that make it difficult to entirely perceive deep learning actions which requires a large amount of computational resources for learning and requires additional verification over a variety of datasets to assure universality. All things considered, the DCNN-GRU model presents a viable option for precise and effective lung disease identification due to its sophisticated deep learning structure and improved accessibility.

It has become a fact that medical image analysis particularly based on deep learning algorithms have made tremendous strides. According to Dayan et al. [11], excellent ability in recognizing different kinds of lung disorders. Validation accuracy scores for well known models like AutoThorax-Net, ChexNet demonstrate the effectiveness of CNNs in this area. However, the complex nature of diseased patterns in chest X-rays makes the accurate interpretation difficult and this highlights the requirement of improved models. Although CNNs have achieved high accuracy on lung disease detection even small performance improvements can have significant therapeutic implications. Follow-up studies have focused on transformer based topologies that have proven to perform better than the previously mentioned CNN approaches in increasing the accuracy of certain diagnosis. Transformer based topologies have been investigated as a result and they have demonstrated potential in improving diagnostic accuracy. Segmentation was assigned more importance concerning clinically relevant areas in chest X-rays with the concept of more localised categorical analysis to reduce noise and present a more detailed panel for specific lung pathologies. Indeed, research findings suggest that transformer oriented configurations are taking the lead over typical CNNs in medical imaging analysis especially for lung disease detection. The performance revealed by the ViTs is an accurate model that can capture intricate features without explicit segmentation thus simplifying the preprocessing. The very evidence from this study supports the effectiveness of transformer based models in clinical practice in terms of how much they could potentially improve patient outcomes as well as its diagnostic competences.

Lung diseases have been classified from chest X-ray images at the medical imaging domain by using deep learning algorithms. CNNs can differentiate COVID-19, pneumonia and TB better than manual interpretation methods, according to studies. Quality training datasets are crucial to these model's performance. As noted by Ifty et al. [12] curation guarantees a balanced lung condition representation while normalisation and data augmentation boost model performance and flexibility. Research has been conducted on both hybrid and transformer based models. The ability of the Xception model to identify lung diseases relatively well explains how important model selection and hyperparameter tuning are. More layers make the model deep therefore, deep learning models must give a better explanation of their decisions. Grad-CAM and LIME-based explanations can be helpful in developing the model prediction in a manner suitable for adoption by the clinician. Increased complexity with more hidden levels thus, requiring an explainable system for making decisions in a deep learning model. By making learnings visible for better predictions via its XAI models of Grad-CAM and LIME it would be possible to win over the clinician. To categorize lung conditions with chest X-rays, deep learning and explainable AI (XAI) is being adopted with increasing frequency. Novel architectures, vast datasets and explainability methods possess the ability to enhance diagnostic performance and support professional judgment. Future work should develop robust, interpretable models for healthcare professionals.

The application of CXR images in order to identify COVID-19-related pneumonia is covered by Bukhari et al. [13]. They emphasize the significance of radiological features in differentiating COVID-19 pneumonia from other forms of pneumonia caused by microbiological infections. The study employed a collection of 278 CXR images, which was subsequently separated into three categories: normal, pneumonia (caused by illnesses other than COVID-19), and COVID-19. The aim of this research was to ascertain the testing integrity of a convolutional neural network architecture known as ResNet-50 for COVID-19 infections as per CXR images. A potential constraint on the study's robustness and generalizability is its dependence on a very small dataset of 278 CXR pictures. Thus, additional testing with larger and more diverse datasets will be required in subsequent research to ensure the validity and appropriateness of the proposed computer vision-based diagnostic system in real-world clinical settings.

Medical research needs lung cancer identification because it stands as one of the primary global killers among cancer patients. According to Patil et al.[14] an early diagnosis which works efficiently determines how patients fare in terms of survival rates and medical outcomes. Subjective image interpretation through CT scans and X-ray machines remains problematic because expert diagnostic skills are required and treatment delays frequently occur. New machine learning and deep learning technologies drive a transformation of contemporary lung cancer diagnostic procedures. Research indicates that convolutional neural networks (CNNs) learn complex image features effectively enough to surpass standard analysis methods. Studies that use large datasets including LIDC-IDRI deliver trustworthy models which demonstrate outstanding performance in detecting lung nodules. Multiple CNN architectures appear in academic research where VGG-19 demonstrates superior performance when extracting features from CT scans. Studies confirm that lung nodule recognition structures achieve detection success rates exceeding 90 percent. Multiple model integration through hybrid approaches now produces effective predictions by solving individual design weaknesses. Machine learning applications for lung cancer detection demonstrate potential to transform diagnostic clinical practices by shortening diagnostic times. The use of these technologies transforms achievable patient outcomes because they enable better medical choices along with earlier detection opportunities. Technical applications require additional research to solve generalization problems between datasets and model integration issues. Fundamental research shows significant advancement in using ML and DL approaches to detect lung cancer. Novel lung cancer detection solutions create a revolutionary pathway toward speedier medical treatment by using efficient diagnostic techniques. Studies need immediate growth because this research discovery demonstrates potential to enhance lung cancer survival rates and improve patient life quality.

Diagnostic methods notably medical imaging advanced with COVID-19. Speed and access limit conventional diagnostic procedures like RT-PCR which are effective. The quick adoption of imaging techniques, notably radiographic imaging on chest X-rays as the most promising approach for early COVID-19 diagnosis is not surprising. Khadhim et al. [15] has demonstrated that deep learning networks notably CNNs can accurately identify COVID-19 in CXR images. Since limited labeled medical pictures most employ using pre-trained models for transfer learning to improve diagnostic accuracy. Several machine learning algorithms have shown great accuracy throughout separating COVID-19-positive individuals from those with other respiratory illnesses. However, dataset representativeness and class imbalance remain issues. Most datasets favor certain demographics or circumstances, causing overfitting and misclassification. New techniques stress balanced datasets with a broad variety of patient demographics consisting of children and adults. Thus, combining CNNs and Recurrent Neural Networks (RNNs) improves diagnostic performance by leveraging both architectures strengths to improve feature extraction and temporal analysis of imaging data. The population of RNNs may learn sequential data dependencies, which helps identify abnormal development. The results show a major trend in CXR imaging COVID-19 diagnosis utilizing sophisticated deep learning algorithms. Though several accomplishments have been documented further research is needed to improve models, solve dataset limits, and improve clinical research accuracy and dependability. That hybrid approach might lead to better COVID-19 diagnostics in future research. Recurrent neural networks learn sequential data relationships that help comprehend illness progression. There is an increasing interest in the following publications for COVID 19 diagnosis using CXR imaging through the use of current deep learning algorithms. Despite several successes research is still needed to improve these models, overcome dataset restrictions, and improve clinical accuracy and dependability. Hybrid model research has considerable potential for improving COVID-19 diagnostic tools.

In this paper Mangeri et al. [16] studied Chest disease prediction using X-ray pictures highlighting deep learning advances, notably Convolutional Neural Networks. Many researchers have examined CNN architectures ability to diagnose pneumonia, TB, and COVID-19 using X-rays. Medical image classification accuracy has improved substantially with CNNs. They better distinguish healthy and unhealthy states by automatically extracting and learning complicated picture information. Comparing pre-trained model designs shows that some obtain higher classification accuracy than others. The size and quality of the dataset also affect model performance. Although much research demonstrates remarkable accuracy rates with smaller datasets expanding to bigger datasets can be difficult and inefficient. Effective data augmentation and preprocessing are needed to improve model results. To evaluate these diagnostic models accuracy, precision, and recall are stressed. Deep learning approaches can improve diagnostic skills, showing a trend toward more automated and reliable medical imaging systems. This study shows that CNNs can alter chest illness diagnosis and suggests topics for future research to address limitations and improve healthcare applications.

In another study, Sharma et al. [17] introduced two CNN architectures which were developed specifically for the intent of detecting pneumonia from chest X-ray pictures. They introduced two variants of CNN networks, one with a dropout layer and the other without dropout layer. Both CNN are composed of a convolution layer, max-pooling layer, and classification layer. For this evaluation, an available dataset in Kaggle was utilized and comprised of 5863 chest X-ray pictures. The data augmentation methods are used to reduce over-fitting. In this way, the data augmentation ensures that CNN will see some variations of the data at every epoch of the training. Among the few data augmentation methods that are applied, we have Rescale Rotation Range, Zoom Range, Width Shift Range, Height Shift Range, etc. Experiments show that dropout and training using augmented data leads to higher loss and lower training accuracy than other models but has a better performance in the validation. Using the augmentation and dropout parameters, the testing accuracy is 0.90 while when the augmentation and dropout parameters are removed, the accuracy is 0.74. In this case, it is worth using batch normalization to minimize the extent of overfitting on the outcome.

Abubakar et al. [18] talks how artificial intelligence and machine learning classify pneumonia and tuberculosis using CXRs. This study shows that CNN models can segregate TB in CXR pictures and will increase in accuracy over conventional approaches. The paper discusses the assemblage of a set of X-ray images from distinct references, including Montgomery county x-ray set, Shenzhen China set and Kaggle, made so that a diverse dataset is created for training the algorithms. The Montgomery County X-ray Set was one of many X-ray picture sources used in this work. Moreover, it includes the study of medical image classification architectures and strategies. For example, one study used the Densenet-169 architecture and Support Vector Machines (SVM) that achieved an adequacy of 80% in predicting features from images of CXR . In contrast, another study discussed normal and abnormal organ segmentation techniques for x-ray images pointing out that achieving credible outcomes relies heavily on dataset size in the end. The research underlined the significance of a large dataset for credible and validated outcomes.

Duong et al. [19] used state-of-the-art Machine Learning and Computer Vision algorithms to demonstrate a workable method for detecting tuberculosis from CXR images. Three modern deep neural networks—modified versions of EfficientNet, modified Vision Transformer, and modified Hybrid EfficientNet with Vision Transformer were chosen for three main categorization algorithms when they designed the framework. Furthermore, they used multiple augmentation approaches to strengthen the learning process. The suggested method was assessed by the authors using a sizable dataset that was assembled by combining multiple publicly available datasets. The authors also compared their suggested method with two cutting-edge systems. Additionally, the study employed pre-trained weights in the

PyTorch framework to speed up the training of the EfficientNet and Vision Transformer models. Their research expanded on the more general uses of AI in illness detection and medical imaging.

In another study, Malik et al. [20] used dilated convolution (CDC_Net) and residual network theory to design an end-to-end CNN model that categorizes CXRs between COVID-19, pneumothorax, pneumonia, lung cancer and tuberculosis. For depth evaluation the research utilizes three baseline classifiers Vgg-19, ResNet-50, Inception v3 alongside state-of-the-art classifiers. Twelve distinct databases on chest illnesses gathered from diverse sources were used to train and evaluate the CDC_Net. In this study, the CDC_Net model is cross-validated ten times. As a result, 90% of the data were used for learning, while 10% were used for classifier testing. To ensure that the CDC_Net was trained correctly for this study, the data was normalised using pixel normalisation techniques.

Another study conducted by Kundu et al. [21] focused on developing an effective computer-aided diagnosis (CAD) system for pneumonia by analyzing CXR images. An illustration of a pneumonic and a healthy lung X-ray is shown in Figure 2.1. Pneumonic X-rays are distinguished from healthy conditions by the white areas known as infiltrates, which are indicated by red arrows. To improve pneumonia categorization, the proposed system used ensemble methodologies along with deep transfer learning. The system achieves a high level of accuracy on two pneumonia CXR datasets, namely the RSNA Pneumonia Detection Challenge dataset and the Kermany dataset, by utilizing transfer learning models as base models and a decision-scoring method that combines the base models to make final predictions. They developed a set of three convolutional neural network models: GoogLeNet, ResNet-18, and DenseNet-121. Additionally, they used an ensemble technique based on weighted average. The weight vector is created by fusing the scores of four standard evaluation metrics: precision, recall, f1-score, and area under the curve. However, this method is prone to error. On the RSNA and Kermany datasets, the suggested method produced sensitivity rates of 87.02% and 98.81%, and accuracy rates of 86.85% and 98.81%, respectively.

(A) (B)

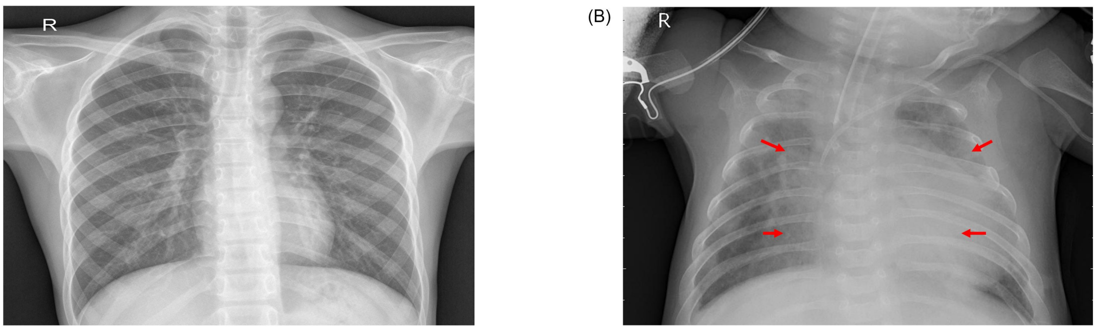


Figure 2.1: Examples of two X-ray plates that display (a) a healthy lung and (b) a pneumonic lung. [22]

For COVID-19 identification, Atitallah et al. [23] provide a novel randomly initialised CNN (RND-CNN) architecture. The network consists of several hidden layers of varying sizes, each of which was constructed from the ground up. The performance of this RND-CNN is assessed using two publicly available datasets: the

COVIDx and the revised COVID-19 datasets.Three different medical dataset types include X-ray images showing either normal chests or COVID-19 or pneumonia conditions. The proposed RND-CNN contains four hidden blocks for feature extraction and learning together with two fully connected layers and an input layer and SoftMax layer for case classification (classes: COVID-19/Pneumonia/Normal). The Keras library version 43 served to program neural networks alongside the TensorFlow backend version 44. Future work should investigate the efficacy of RND-CNN by evaluating its performance using CT and MRI image formats according to the authors' recommendation. Additional research will enhance the model's capacity to identify images according to several labels including fibrosis, emphysema, pneumothorax and others.

Bhosale et al. [24] presented the use of CNNs in identifying COVID-19 from pictures from radiology X-rays. One of the suggested models was using statistical models like the Logistic regression classifier to address this issue with the help of feature extraction using CNNs like Resnets or Inception. They are then used in the classification with the help of SVM methods. The workflow of this system utilizes portable devices to transcode X-ray pictures into the DeepX model in addition to employing the system's AWS Cloud9 EC2 environment to analyze DeepX via computers. Thresholding and segmentation of images through the use of deep learning algorithms is another process in data collection and pre-processing for deep learning model training. Model which was suggested seemed to be interesting and had a function of good test accuracy of 94%. The persuasive precision of the algorithm effectively identifying COVID-19 and classifying it as 95% of the small list of instances recognizably from radiology X-ray images shows the algorithms' ability to spot Covid-19 out of the images of CXR.

Gour et al. [25] created an advanced stacked convolutional neural network structure to automate COVID-19 diagnosis by uniting X-ray and CT scan images of the chest. The proposed approach during training obtained several sub-models from both VGG19 and Xception networks. The softmax classifier executes the combination process of the constructed sub-models. A stacked CNN model combines different sub-model CNN networks to establish COVID-19 detection capabilities in radiological images. The authors collected X-ray images from three public databases for creating their X-ray image dataset while building their CT image dataset from CT images. The researchers improved the VGG19 and Xception models by adding CT and X-ray data to them. A stacked ensemble model underwent training through multiple sub-models which stemmed from VGG19 and Xception architecture design. They examined six pre-trained CNN models for their performance when detecting COVID-19 through CXR images. A division of the COVID19CXr dataset into test and training and validation sets took place at this point. Xception and VGG19 received training through 3192 iterations of CXR images from the training set. The authors addressed class imbalance through class weight assignment during their network training phase.

A mechanized system for identifying COVID-19 using CXR pictures was proposed by Muralidharan et al. [26]. Modes were retrieved from the CXR images using the Fixed Boundary-based Two-Dimensional Empirical Wavelet Transform (FB2DEWT). One CXR picture was split categorised into seven modes by researchers. Using the evaluated modes as input CNN classifies images into three categories: normal, pneumonia, and COVID-19. In future, the efficiency and accuracy of the model can be increased by applying evolutionary computing algorithms to automatically choose the X-ray picture modes for the multiscale deep CNN.

Ahmed et al. [27] delved into the importance of medical imaging in early diagnoses. To use diagnosis in amelioration of treatment outcomes and selection of treatment options. It had also been demonstrated that deep learning architecture in combination with the CNN can help reduce some of the impacts of false positives or mislabeling on error maps and categories. Deep learning helped traditional features-based algorithms distinguish TB and pneumonia. Such techniques involve assessing if radiologists used the proper annotation for stimuli and comparing it to NLP-annotated photos. Both criteria could be used to distinguish SDR or DR TB prevalence using CXRs. Their work reduced the hype of quick machine learning (ML) and AI technique development for medical imaging for early respiratory illness detection and differential analysis. The proposed mechanism complements recent advancements by combining components of many systems and mixed characteristics to diagnose pneumonia or TB in X-ray pictures.

Research done by Ahmed et al. [28] had made significant progress. Various image processing techniques, including the Weber local descriptor, local binary patterns, contextual style transfer, and intensity distribution remodelling, had been examined in studies to enhance the accuracy of categorization. Joint illnesses were extremely important, especially in light of the COVID-19 pandemic. Here, the authors emphasized transfer learning and cite examples of how transfer learning models might diagnose illnesses. Because medical imaging studies generated a lot of data, which needed to be efficiently managed in order to apply the acquired knowledge to other processes. The assessed study indicated the advancement of deep learning technologies for simultaneous respiratory problem diagnosis utilizing medical imaging.

Nafisah et al. [29] described how researchers have worked hard to employ deep learning in medical imaging, notably for TB identification in chest radiographs. The U-Net has been widely utilized to divide biomedical pictures, allowing accurate identification of areas like lung anomalies in chest radiographs. It was possible for health care practitioners to verify and interpret the results of tuberculosis through visual perception of tuberculosis infected lungs. Research had demonstrated the possibility of diagnosing TB through the segmentation of CXRs by deep learning models (Section networks) alongside explainable artificial intelligence algorithms. This improvement in technology was designed to make more accurate and faster TB diagnoses possible.

In this study, Iqbal et al. [30] created a unique TB-UNet based on X-ray images for precise lung region classification. The developed CAD method consists of two phases: (TB) categorization and lung area identification. Using the dilated fusion block (DF) and attention block (AB) block as its foundation, it generated the greatest results in terms of accuracy (0.9770), precision (0.9574), recall (0.9512), F1score (0.8988), and IoU (0.8168). The authors had introduced TB-DenseNet, which was built on five dual convolution blocks, DenseNet-169 layer, and a feature fusion block for precise categorization of tuberculosis images. Three datasets of CXR were used in the study, and segmented and original CXR pictures were added to TB-DenseNet for improved categorization. In this study, researchers presented TB-UNet, an entirely new UNet variant for segmenting human lungs in the images of CXR. The emphasis block (AB) was additionally incorporated in order to focus greater focus on areas of interest (ROI) and retrieve additional relevant data from CXR images. A TB-DenseNet with five convolutional dual blocks with varying filter widths was proposed as the second option. For enhanced feature retrieval and to avoid overfitting problems, DenseNet-169 was employed. Additionally, gradient-based labeling was employed to closely inspect the lung injury locations using the CXR images.

In another study Sharma et al. [31] developed a CNN model for the recognition of CXR images in order to detect pneumonia. The model was suggested for the identification of CXRs, which were divided into two groups: pneumonic and healthy X-rays. For recognition, a CNN model with two optimizers, RMSprop and Adam, was utilized. The Kaggle dataset, which contained 5000 image samples for every class, was used for recognition. The CNN model consists of three primary layers: convolution, pooling, and the fully connected. Based on the collected findings, it was found that the maximum validation accuracy was attained by the Adam optimizer and the RMS prop optimizer with an LR of 0.01. The RMSprop Optimizer for the provided dataset yields the highest accuracy and lowest loss rate of 90% and 0.33, respectively, based on the validation accuracy achieved using the two optimizers. As a result, RMSprop had outperformed Adam Optimizer in identifying the provided dataset. Future initiatives could concentrate on cutting processing time in order to increase efficiency. Large-scale data processing demands a significant amount of computational power, which might restrict the model's scalability and its suitability for contexts with limited resources.

Chapter 3

Background

3.1 Convolutional Neural Network (CNN)

It is an advanced deep learning model that has been specifically developed to process and analyze the visual data and is called as Convolutional Neural Network or CNN [5]. In a very strong image identification and categorization activities, CNNs are incredibly successful. Because of their capacity to dynamically and adaptively absorb spatial patterns of specifications, which is designed after the arrangement and performance of the people's visual cortex. Convolution is a CNN's core function that employs filtration systems, often referred to as kernels, to feed images in order to generate features visual representations that gather important features like borders, surfaces, and formats. These filters multiply each element of the provided image separately and then aggregate them over the region the latter performing equivalently to swiping over the given image.

CNNs analyze and adapt from picture input employing a variety of layers that coordinate with each other. A detailed explanation of each CNN architecture is provided below:

3.1.1 Architecture Overview:

Input Layer:

The input layer is where the data is entered. Pixel metrics from images are accepted as input by this layer's parameters. The input size for a 28 x 28 grayscale image is (28, 28, 1), 3D matrix (Height $\times$ Width $\times$ Channels) for RGB pictures. The input size for a 64x64 RGB picture is (64, 64, 3). Where,

Convolutional Layer (Conv Layer):

These layers use convolution procedures to extract features from the initial data received from the input layer. By using numerous filters to identify unique types of infections, the model begins learning stacking interpretations that span superficial characteristics (edges) to advanced characteristics (objects). It performs convolution processes across a number of filters and the provided input image. As each filter moves along the picture, dot products among the initial patches and the filter are calculated. Stride, padding, and filters/kernels are important elements of this convolutional layer. Here's a brief description of these elements:

Filters/Kernels: Minimal vectors that can identify features like edges or patterns, such as 3x3 or 5x5.

Stride: Controls the number of units the filter travels throughout the convolution process.

Padding: Regulates the outcome's spatial dimensions. There is valid padding that diminishes output. On the other hand, same padding keeps the size of the input same, add zeros to make sure the dimensions of the input and output are equivalent.

The following is the mathematical computation performed by the convolutional layer:

$$\mathbf{Y(i,j)=(I*K)(i,j)+b} \tag{3.1}$$

Where I is the input, K is the filter, b is the bias, and * denotes convolution [5].

Activation Layer (ReLU Layer):

Non-linear activation operations that include ReLU come after convolutional layers, adding non-linearity as well as allowing the network to absorb intricate patterns. When the value provided is positive, the ReLU activation function transmits it directly; otherwise, simply adjusts values that are negative to zero. The ReLu function's mathematical expression [5] is

$$\mathbf{f(x)=max(0,x)} \tag{3.2}$$

Pooling Layers:

Pooling layers reduce computing costs and the possibility of overfitting by shrinking the visual qualities of feature representations. They eliminate minor essential data while keeping the most essential details. It functions autonomously on every layer of the feature map. Additionally, it uses a fixed-size frame (such as 2x2) for combining data. Two different kinds of pooling are employed here.

- **Max Pooling:** Utilizes the maximum value of a window. For instance, by choosing the maximum value of each window, a 2×2 max pooling minimizes a 4×4 feature map to 2×2.
- **Average Pooling:** Calculate and select the average value of a window.

Flatten Layer:

This layer creates a 1D vector from multi-dimensional maps of feature. Additionally, it flattens the outcomes of the pooling and convolutional layers into a single smooth vector. Next, it gets the data ready for the levels that are fully connected. For instance, a feature map with dimensions of (8, 8, 64) is compressed into a 1D vector with dimensions of $8 \times 8 \times 64 = 4096$.

Fully Connected Layers:

Fully connected layers perform high-level reasoning and categorization by analyzing the gathered features after feature extraction and size reduction. Every neuron in this layer is connected to every other neuron within the layer above it. The high-level characteristics are then processed for regression or classification. The mathematical Operation of this layer is,

$$\mathbf{y} = \mathbf{W} \ . \ \mathbf{x} + \mathbf{b} \tag{3.3}$$

Here, x denotes features from the previous layer/ inputs, W is weight matrix and b Bias vector. Finally, y gives the results of the linear transformation, processed further by an activation function [5].

Dropout Layers:

Dropout layers periodically disable a portion of neurons to avoid overfitting during forward and backward passes. Additionally, it fosters the network's expansion of stronger features. For instance, Dropout(0.5) means in every iteration randomly it will adjust 50% of neurons to zero.

Output Layer:

This layer provides the ultimate predictions and generates the outcome (class labels or regression results). This task is essential for many neurons. One neuron has sigmoid activation in binary classification, N neurons have Softmax activation in multi-class classification and one neuron has linear activation in regression.

Batch Normalization Layer:

Batch Normalization Layer enhances performance and expedites training by normalizing inputs to layers. In order to scale and adjust inputs, it then takes into account the sample mean and variance. Additionally, this layer aids in lowering internal covariate shift.

Since CNNs are effective at processing multidimensional data and are capable of extracting pertinent features without human assistance, they are extensively utilized. They are consequently perfect for uses including object detection, image categorization, and medical image processing.

3.2 RNN Architecture

An artificial neural network model called a Recurrent Neural Network (RNN) is developed to identify patterns in data series, including time series, text, audio, and video [32]. It is referred to as ‘recurrent’ because it depends on earlier calculations to get the same result for each member of a series. RNNs are perfect for scenarios in which context and sequence are important since they accomplish this by storing details regarding previous inputs in internal memory. RNNs are useful for sequential data handling because in contrast to conventional feedforward neural networks they contain loops that permit information to remain. RNNs are good at identifying patterns within a period of time which makes them appropriate for detecting anomalies in sensor or financial data.

3.2.1 Architecture Overview:

Input Layer: RNN input layer accepts input data sequentially. One by one, over various time steps, every single component of the input sequence x= x_1, x_2,..., x_T is provided into the system. For example, every term (or character) in an expression is an intake at a time phase.

Hidden Layer:

This layer uses its own memory configuration to handle inputs and record periodic dependencies. The present input x_t and the preceding hidden state h_{t1} are used to modify the hidden state h_t at time step t. Information from earlier steps is stored in the hidden state which functions as a memory.

$$h_t = f(W_{xh}x_t + W_{hh}h_{t-1} + b_h) \tag{3.4}$$

where, W_{xh} stands for input to hidden state weight matrix, W_{hh} is hidden to hidden state weight matrix, b_h is the bias vector and f is a commonly used activation function, tanh or ReLU [32].

Output Layer:

This layer uses the hidden state to generate outputs or predictions. Using bias vector, weight matrix and activation functions, this layer generates output.

- **Activation Functions:** In order to enable the neural network to absorb intricate patterns, it introduces non-linearity. Some of the common activation functions are tanh, softmax, ReLu.

Loss Function:

It calculates the discrepancy between expected and real results. For instance, regression is performed using Mean Squared Error (MSE) and classification is performed using Cross-Entropy Loss.

Backpropagation Through Time (BPTT):

It uses gradients to update weights in order to train the network. After that, the RNN is unrolled during time phases, gradients are calculated at every time phase and weights are updated to reduce loss. Additionally, it is vulnerable to vanishing or fulminating gradients which makes learning long-term dependencies challenging.

Types of RNN Variants

Vanilla RNN: It is the Basic RNN with the previously mentioned architecture. Because of diminishing gradients it faces challenges with long-term correlations.

Long Short-Term Memory (LSTM): Its purpose is to get over fading gradients. It presents gates and memory cells. For example, Forget Gate specifies what data to ignore. What new data should be stored is decided by the input gate and what should be output is decided by the output gate.

Gated Recurrent Unit (GRU): Having lesser gates, it is the simplified LSTM variant. For instance,Update Gate integrates input and forget gates and Reset Gate Regulates the flow of previous data. When managing long-term dependencies it is also quicker and easier than LSTM.

Flow of Data in RNN:

During time t, input x_t joins the network. After that, x_t and h_{t-1} are used to modify the hidden state h_t. Output y_t is produced from there. In order to modify weights, losses are then back propagated over time.

RNN networks can take inputs of variable length, adapt from sequential patterns and simulate dependencies over time features that classic neural networks do not have therefore RNNs are frequently employed.

3.2.2 RNN (Bi -LSTM)

An improved version of conventional LSTM networks which is called bidirectional long short-term memory (Bi-LSTM) networks use data movement in both forward and backward directions to improve model efficiency [32]. Traditional RNNs struggle to handle correlations over time because of issues like shrinking and inflation gradients. Long Short-Term Memory (LSTM) networks, which make use of storing neurons and circuits and overcome this problem. By handling input in both forward and backward orders bidirectional LSTM (Bi-LSTM) advances one step farther and captures additional details. Bi-LSTMs handle data concurrently for both directions allowing the system to record information across the previous and subsequent time phases in contrast to normal LSTMs that analyze data linearly from the previous to the future. Bi-LSTMs are very good at sequence categorization and prediction activities because of their parallel execution.

A Bi-LSTM's processing under every LSTM cell operates similarly to that of conventional LSTMs, utilizing gates such as input, forget, and output gates to control the movement of data. To decide which data to keep or delete, each gate employs the sigmoid activation function. Therefore, the concealed condition and cell condition are refreshed. A thorough description about every Bi-LSTM layer and its function inside the model is given below:

Input Layer:

After receiving unprocessed sequential data, the input layer makes it ready for additional analysis. After that, it sends the input to the LSTM layers for analyzing in both forward and backward direction simultaneously. Series of text integration or one-hot encoded matrices are frequently used in text analysis, whilst integer patterns may be used in time-series data processing. The input data usually takes the form of (batch size, sequence length, features), in which attributes are the data at every interval and sequence length is the number of time steps.

Forward LSTM Layer:

Beginning with the initial time frame to the last, the given input series is processed by the forward LSTM layer. In order to comprehend how prior data sets affect the next stage, it records details from previous ones (earlier time steps). With its cell state for storage management and its input, output, and forget gates, this level is structurally similar to a typical LSTM. As the sequence progresses, this layer detects patterns and constraints. Additionally, it uses storage cells and gates to keep essential data. Every gate's arithmetic calculation and operational changes continue to follow the conventional LSTM version:

- **Forget Gate:**

$$f(t) = \sigma(W_{fx} \cdot x(t) + W_{fs} \cdot s(t-1) + b_f) \tag{3.5}$$

  This specifies which data from the prior state should be kept or discarded [32].

- **Input Gate:**

$$i(t) = \sigma(W_{ix} \cdot x(t) + W_{is} \cdot s(t-1) + b_i) \tag{3.6}$$

$$c(t) = \tanh(W_{cx} \cdot x(t) + W_{cs} \cdot s(t-1) + b_c) \tag{3.7}$$

  These formulas determine what fresh data should be stored in the cell and calculate the input variation [32].

- **Cell State Update:**

$$C(t) = f(t) \cdot C(t-1) + i(t) \cdot c(t) \tag{3.8}$$

  This updates the cell state by combining the provided inputs and forgetting gate outcomes [32].

- **Output Gate:**

$$o(t) = \sigma(W_{ox} \cdot x(t) + W_{os} \cdot s(t-1) + b_o) \tag{3.9}$$

$$s(t) = o(t) \cdot \tanh(C(t)) \tag{3.10}$$

These formulas use the current cell state and concealed state to calculate the ultimate result [32].

Backward LSTM Layer:

In order to enable the system to utilize the setting that takes place afterwards in the series to notify prior predictions, the backward LSTM layer gathers attributes proceeding backward using the sequence and provides upcoming setting to the forward LSTM. It handles the input sequence backwards, from the final time phase to the initial and records details from the subsequent in contrast to all orientations.

The forward and backward stages in Bi-LSTM models function simultaneously while making the same contributions to the final result. The sum of their contributions is shown as [32]:

$$Output_{Bi-LSTM}(t) = Concat(Forward(s(t)), Backward(s(t))) \tag{3.11}$$

Concatenation/Merging Layer:

The outcomes of the forward and backward LSTM levels merge once the input has been analyzed. Usually, fusion is used to achieve that, combining the concealed states from each side for every time step. As a result, the input sequence is represented more thoroughly and extensively. This layer integrates data from the past and the future to deliver a comprehensive contextual perspective. Dropout Layer (Optional): The dropout layer which occasionally sets a portion of cells to zero during learning that is frequently employed following the combination layer to avoid overfitting. Consequently, the network becomes obligated to learn more robust and widely applicable patterns.

Fully Connected (Dense) Layer:

The fully connected or dense layer receives the concatenated outputs obtained from the Bi-LSTM stages and transforms them into the required output dimension. It is the layer that makes the ultimate predictions. The extremely dimensional Bi-LSTM results are converted into the desired output shape. This layer integrates extracted features for decision-making and connects absorbed characteristics to the outcome layer for the final prediction.

Output Layer:

The output layer uses the fully connected layer's result to generate the final prediction, which dictates which activation function is utilized such as Linear for regression operations, Sigmoid for binary classification, and Softmax for multi-class categorization. This layer transforms discovered patterns into useful predictions and generates the ultimate result in the shape of preference.

The Bi-LSTM model is perfect for tasks where comprehension of combined directions is required since it incorporates both previous and future context. It increases productivity in sequential tasks like text categorization, speech recognition, and machine translation. For improved competence, this kind of approach can be integrated with other deep learning models or emphasis processes.

3.2.3 GradCAM (Gradient-weighted Class Activation Mapping)

A method in XAI called Grad-CAM [33] aids in the interpretation of deep learning models, especially CNNs. By creating a visualization called a heatmap, it illustrates the regions of the image being used that have the most influence on a model's prediction. It gives information about things the model perceives as significant by highlighting the areas of an input image that it concentrates on when generating a certain prediction.

Explanation of Grad-CAM Layers and Their Roles

- **Input Layer:** The input image is inserted into the model at this layer. All graphic data in a particular format (RGB images scaled according to particular ratios) can be used for the input. This layer initiates the foundation for feature collection and gradient calculation and supplies the data set that the CNN will handle.

- **Convolutional Layers:** Hierarchical details are extracted from the original image by the convolutional layers. As the layers get deeper, they pick up designs like edges, textures, and intricate shapes. Grad-CAM focuses on the last convolutional layer since it preserves the spatial data required to locate objects and These layers create visual representations that draw attention to particular regions of the picture.

- **Activation Maps:** The outcomes of the convolutional layers are called activation maps. They show that particular characteristics are visible in various areas of the picture. Each attribute map's significance for the target class is used for weighing the activation maps from the last convolutional layer. The Grad-CAM visualization is built on top of this spatial data.

- **Fully Connected Layers:** Convolutional layers capture features, which are then aggregated by fully connected layers to generate predictions. Spatial data is not retained by these layers. Due to their low spatial resolution, Grad-CAM omits these layers when creating heatmaps. In order to determine the significance of every detail map, gradients are rather back-propagated to the last convolutional layer.

- **Gradient Calculation:** Grad-CAM calculates the gradients of the ultimate convolutional layer's feature mappings in relation to the desired class score (the softmax layer's output). The significance of every map of features for target class prediction is indicated by these gradients.

  The gradient for the k-th feature map is given below:

$$\alpha_k^c = \frac{1}{Z}\sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \tag{3.12}$$

Where, Where Z = is the normalization factor, y^c is the score for the target class c, and A^k is the k-th feature map [33].

- **Global Average Pooling:** To determine an unique important weight for every feature map, the gradients are taken as an average spatially across the feature map's width and height. It calculates each feature map's influence to the prediction of the desired class.

- **Weighted Feature Maps:** To create a weighted set of the feature maps, all feature maps are multiplied by the associated relevance weight and then added together. Employing the ReLU function, it merges the main pertinent feature maps and suppresses those that have no significance (to focus solely on positive effects).

- **Heatmap Generation:** Bilinear interpolation is used to increase the sample of the weighted feature map to equal the actual input image's resolution. The areas which were particularly relevant to the model's judgment are highlighted in the heatmap that results. It enables the model's decision-making procedure to be readable and offers a visual depiction of its focus. The final Grad-CAM map is given by [33]:

$$L^c_{\text{Grad-CAM}} = \text{ReLU}\left(\sum_k \alpha_k^c A^k\right) \tag{3.13}$$

Grad-CAM can demonstrate if the model is examining the disease or unimportant areas in medical imaging. It assists in determining whether the model is concentrating on the appropriate regions. By clearly explaining predictions, Grad-CAM increases user confidence in AI systems which is crucial in delicate sectors like medical. By indicating if the model concentrated on inaccurate or unnecessary regions, it can be utilized to identify model problems. Grad-CAM findings can direct architectural modifications or data augmentation to enhance the model. Grad-CAM makes deep learning models predictions easier to understand, which improves their functionality and dependability.

Chapter 4

Research Methodology

4.1 Workflow

In our research paper we are aiming to develop a hybrid model, C-RNet, combining CNN-RNN architecture to enhance the overall performance for lung disease categorization based on chest X-ray images. Here in the Figure 4.1, we walk through the steps of how we have achieved the goal of developing our model.

In the initial step, we start by collecting our dataset. For data collection, we use publicly available COVID-19 Radiography dataset [6] where each image is labeled and then divided into classes like COVID-19, Lung Opacity, Normal, and Viral Pneumonia. After gathering our dataset, we have preprocessed the dataset. In this process, we preprocess it by resizing images for consistency, normalizing to improve model convergence, and applying rescaling and zooming to prevent overfitting. We have also addressed missing values to avoid any bias in the model's learning process. We also extract patches from images to improve model performance. Additionally, we divided the image into small patches for the RNN branch. Finally, we split data for training, testing, and validation and by following these steps our dataset is now ready for further model training.

Right after preprocessing, the images are passed into the proposed CNN, and the pathed images are passed into RNN models and finally these two models are merged. This merging step integrates spatial features from images and temporal features from sequences effectively, ensuring that our model performs well on new, unseen, and diverse data.

In the next step, the data from merged step goes to the training and evaluation phases where the performance is assessed using accuracy, F-1 score, recall, precision, and other evaluation metrics. Once evaluated, the trained model is prepared for the classification which categorizes input chest X-ray images into one of the following classes COVID, Lung Opacity, Normal, Viral Pneumonia. We have also used Grad-CAM to highlight the most important regions in the images that are the reason or influence the model's decision-making.

Finally our trained model is compared with Xception (K-fold), ViT, VGG19, and

Inception LSTM architectures which ultimately helps us identify the best-performing model.

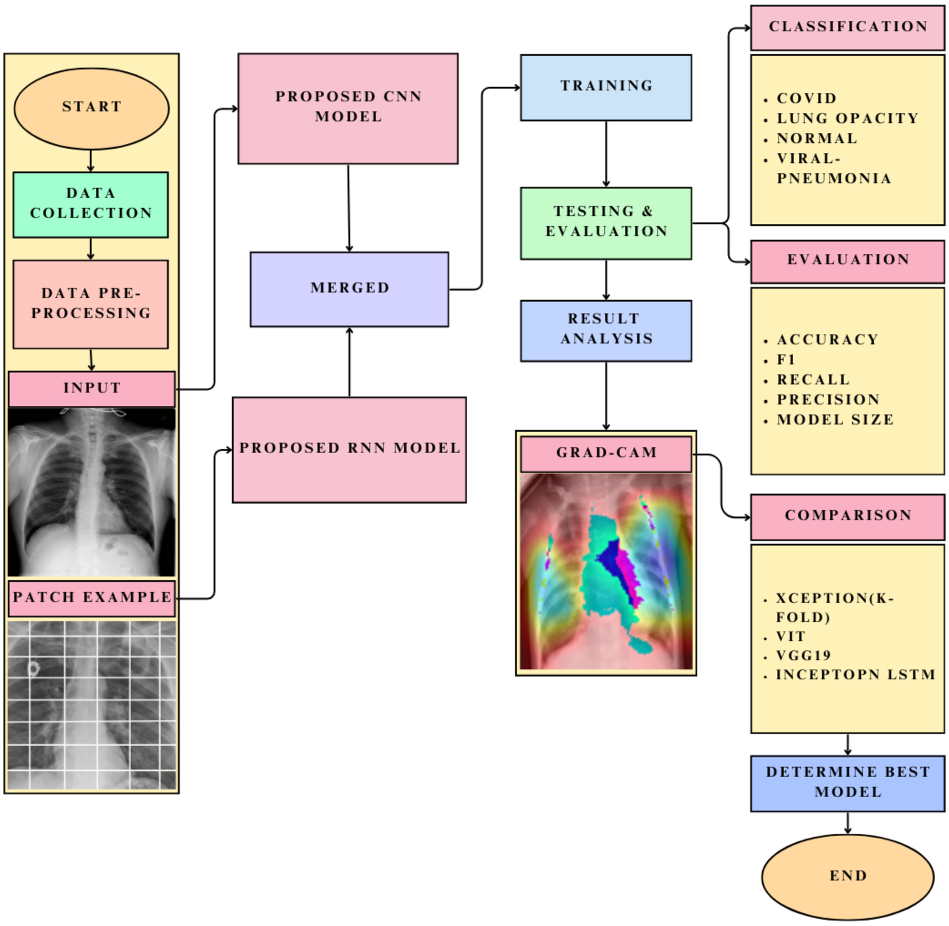


Figure 4.1: Workflow of proposed C-RNet model

4.2 Data Description

The dataset described is a comprehensive and meticulously curated collection of chest X-ray (CXR) images, developed collaboratively by researchers from Qatar University, the University of Dhaka, and other institutions in Pakistan and Malaysia, alongside medical professionals [6]. This database aims to facilitate impactful scholarly work and AI-based advancements in COVID-19 detection and other pulmonary conditions. It is notable for being awarded the COVID-19 Dataset Award by the Kaggle Community, highlighting its utility and importance in research. The dataset is divided into 4 basic classes.

- COVID-19
- Viral pneumonia
- Lung opacity

- Normal

The initial release of the dataset included 219 COVID-19 cases, along with 1,341 normal and 1,345 viral pneumonia cases. Subsequently, the dataset underwent multiple updates, significantly expanding its scope. Currently, it comprises 3,616 COVID-19 cases, 10,192 normal cases, 6,012 lung opacity cases (representing non-COVID lung infections), and 1,345 viral pneumonia cases. Additionally, ground-truth lung masks for these images are provided, aiding in segmentation tasks.

The authors have also created the COVID-QU-Ex dataset, which is a significant improvement compared to the previous one. In its extended version, it includes 33,920 CXR images, consisting of 11,956 COVID-19 cases, 11,263 non-COVID infections (including viral and bacterial pneumonia), and 10,701 normal ones. This dataset also provides ground-truth lung segmentation masks for all images and 2,913 more infection segmentation masks coming from the QaTaCov project. It is described as the largest lung mask dataset ever created, which underlines its value for machine learning applications that require precise lung segmentation.

To offer an entire understanding of the dataset used in this research, Figure 4.2 presents a classification chart summarizing all accessible data types. This figure illustrates the distribution of data across various categories, including vital attributes like the amount of samples per category. Additionally, Figure 4.3 represents a visualization of the dataset images used as input in our model.
m

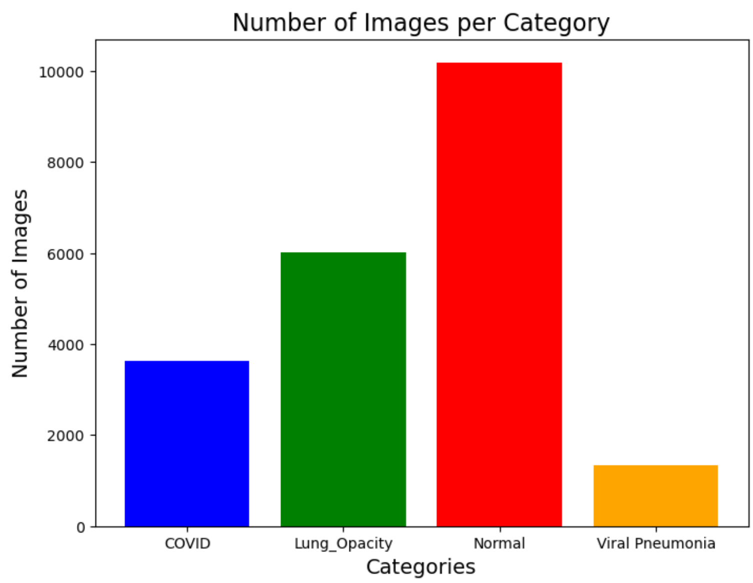
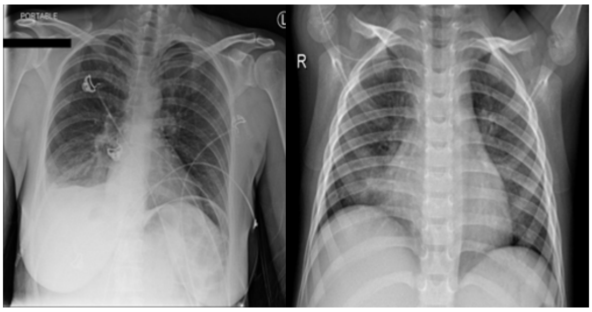
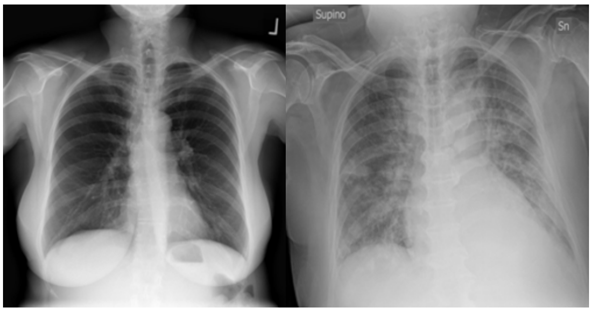


Figure 4.2: Classification of Dataset Categories



(a) Viral Pneumonia (b) Lung Opacity



(c) Covid (d) Normal

Figure 4.3: Visualization of Dataset Images

4.2.1 Data Pre-processing

To prepare the dataset for effective model training, we have performed image resizing, labeling, data scaling, dropout, and patch cutting. Preprocessing has started by loading the dataset and all images are resized to a specific image size i.e. 128x128. However, images that are not resized or returned 'None' are dropped. After resizing all images, a label is assigned to each category where 'COVID' = 0, 'Lung-Opacity' = 1, 'Normal' = 2, 'Viral Pneumonia' = 3. Then, the data is divided into training and test sets with 70% data in the training set and 30% in the test set. For the CNN part, we have performed image rescaling where each pixel value was brought between 0 and 1.

Since our model incorporates an RNN, the dataset images are patched by converting them into 8×8 size patches. This approach allows the model to focus on smaller regions, improving training accuracy by isolating relevant parts of the image. Additionally, data augmentation was not required for the RNN component, as patch creation itself acts as a form of augmentation. This structured preprocessing method preserve the data in a format that can be immediately fed into the deep learning model.

4.3 Model Description

4.3.1 CNN branch:

The CNN architecture workflow starts with an input image of size 128×128×3, which represents a color image. This input is normalized using a rescaling layer that scales the pixel values to the range [0, 1]. There are total 5 convolutional blocks in the CNN part and the data is passed through these five sequential convolutional blocks.

The first convolutional layer uses 32 filters which use a 3×3 kernel size to extract features from small parts of the image. Similarly, the remaining 4 convolutional layers use 64, 128, 256, 512 filters respectively to extract more complex features and also keep the padding size which helps to keep the output image size the same as or very close to the input image. L2 regularization is used to reduce overfitting in all convolutional layers which helps to protect the model from overfitting. After the 5 convolutional layers, batch normalization is performed which normalizes the output of the layer and helps in making the training time fast and stable. We have used ReLU function for fast training which makes negative values zero and positive values unchanged. The last part of the convolutional part uses 2x2 pooling window which reduces the size of the layer input image (downsampling). This preserves important spatial features and makes the model more compact and efficient. Then we used Dropout layer through which we close the connections of 30% of the neurons. This helps in protecting the model from overfitting.

After passing through these five convolutional blocks, the extracted features undergo a global average pooling operation. It averages the values of each feature map in the output of the last convolutional block so that the layer reduces the size of the output and creates a more compact feature representation. A dense layer with 256 units and ReLU activation is applied to further process the feature vector. A dropout layer with a rate of 0.5 is followed to improve the generalization ability.

The result is the final output of the CNN branch, a condensed feature vector ready to be integrated into the broader model or used for downstream tasks such as classification. Figure 4.4 provides a overview of the CNN branch of our model.

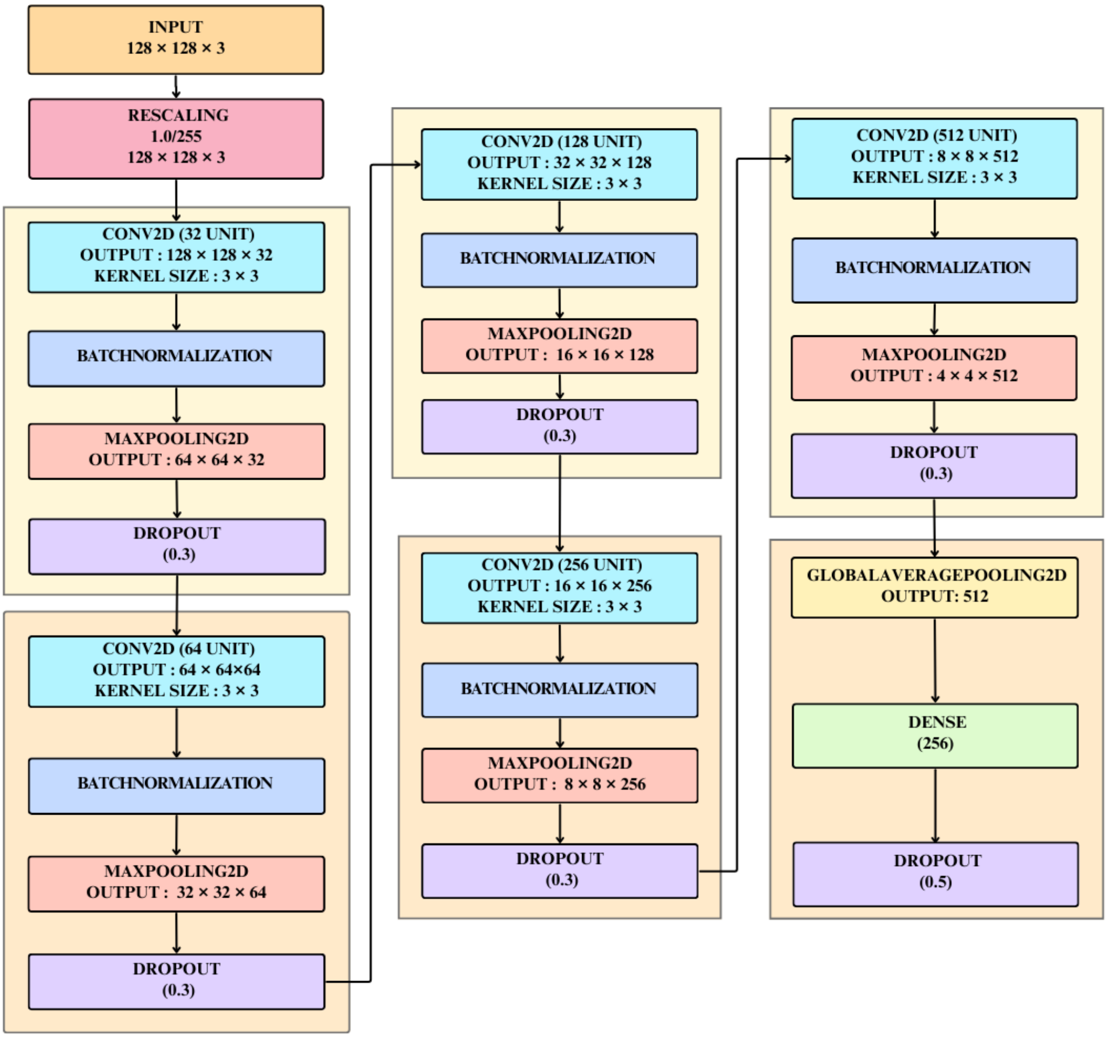


Figure 4.4: CNN Branch of C-RNet Model

4.3.2 RNN branch:

The design, depicted in Figure 4.5, provides a overview of the RNN branch of our model. The RNN branch starts by receiving an input tensor with the shape (256, 192). This input represents sequential or temporal data that will be processed to extract meaningful features. The first step involves passing the input through a Bidirectional LSTM layer with 64 units, utilizing the hyperbolic tangent (tanh) activation function. This layer enables the model to learn temporal dependencies in both forward and backward directions, capturing contextual information effectively.

In the case of RNN, the image is not being used directly as input. It is used to create small patches from the image data. The main purpose of using patches is to analyze the spatial features of the image by breaking them down into small parts. There is a lot of information in a whole image which can sometimes be difficult to process together and many CNNs cannot detect that information. Although

CNN can learn spatial features from patches very well, RNN is able to learn long-term relationships and sequential features between them. By converting the entire image into small patches instead of the whole image, each patch can be processed separately which helps the model to learn features faster and easier. Each image from the dataset i.e. each input sample is divided into 256 patches and each patch contains 192 features (specifically RGB pixel values of each patch). Later when these patches are sent as input to the RNN, the shape of the input is (256, 192). Here 256 is the length of the sequence i.e. 256 patches and 192 is the number of 192 features present in each patch. Basically here we have taken each patch as a timestep. Then this shape goes directly as input to the RNN. Through this input, the RNN model is receiving sequential data and learning the relationship between them.

Then the feature Bidirectional layer goes to an LSTM network which is being processed from both directions i.e. forward and backward sequence. It processes the image patches like a time series so that it is able to capture the hidden relationship between each patch.

After taking the output of the LSTM, it goes to a dense layer which has 128 units and the ReLU activation function is used. It converts the output of the LSTM into a feature vector which can be processed more complexly in the next layer. L2 regularization is used in this layer so that the model is protected from overfitting. Finally, a dropout layer is used to control the model and prevent overfitting. Hence, 30% of the neurons are randomly deactivated in this layer.

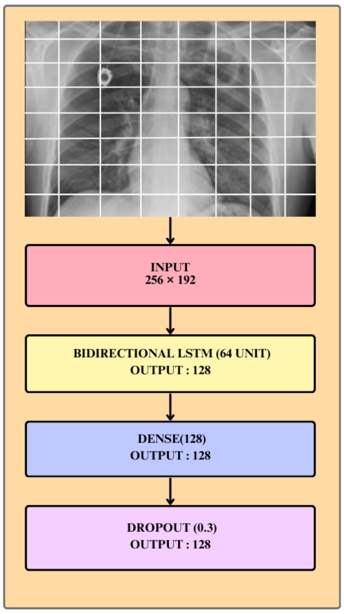


Figure 4.5: RNN Branch of C-RNet Model

4.3.3 Final Model (CNN + RNN):

The complete model architecture starts with two separate branches: the CNN branch for processing image data and the RNN branch for handling sequential data. These branches are designed to extract meaningful features from their respective inputs before merging into a unified representation for the final classification task.

In the CNN branch explained above, the input images are normalized and processed through five convolutional blocks with filters from 32 to 512 respectively. The ReLU activation, batch normalization, max pooling and dropout are used to prevent overfitting. A global average pooling layer is extracting features, followed by a dense layer, producing high-level feature representation.

In parallel, the RNN branch takes as input a sequence of size 256×192. This sequence is processed using a bidirectional LSTM layer with 64 units to capture temporal dependencies in both directions. The LSTM output is then passed through a dense layer with 128 units and ReLU activation to enhance the learned features. A dropout layer of 0.3 is applied to regularize the RNN branch and reduce overfitting.

Finally from the output of both CNN and RNN branches are merged into a single combined representation. This merged representation is then passed through a dense layer with 128 units and ReLU activation to integrate the features from both branches further. Finally, the output layer consists of 4 units with softmax-activation, which produces probabilities for each of the four output classes.

This architecture efficiently combines spatial features from images and temporal features from sequences, making it well-suited for tasks that require multimodal data processing. The network is designed to be robust through regularization techniques such as L2 regularization and dropout layers, ensuring it generalizes well to unseen data.

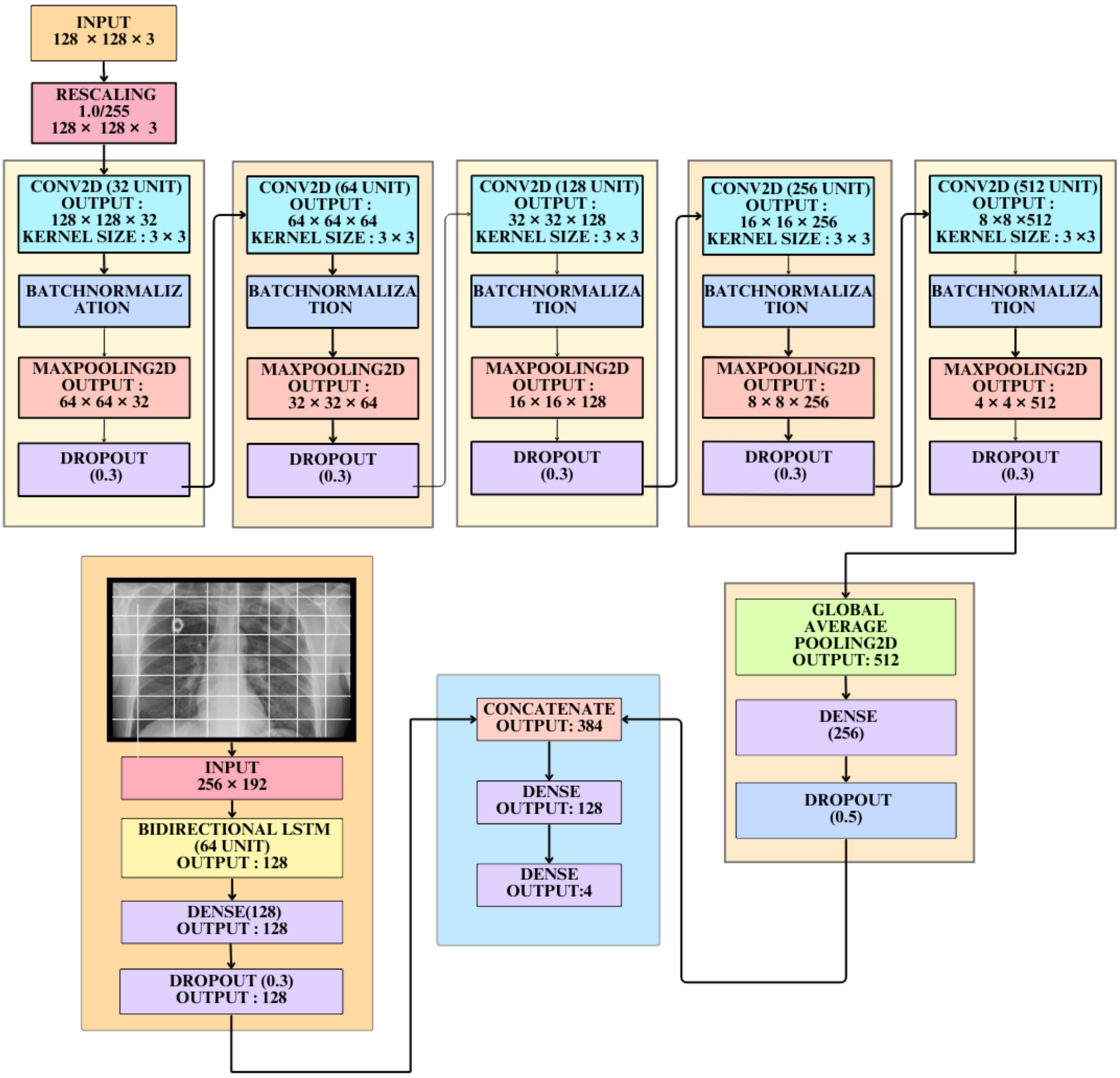


Figure 4.6: C-RNet Model

Chapter 5

Experimental Evaluation

5.1 Experimental Setup

The proposed CNN-RNN model is developed and evaluated in an environmental computer environment with the following specifications: 13th Gen Intel(R) Core(TM) i5-13600KF, 16 GB DDR4, NVIDIA GTX 1650 (4 GB VRAM). The project was developed using Windows 11 as the operating system and Python 3 as the primary programming language. Several libraries and tools were utilized, including zipfile, os, shutil, random, numpy, pandas, skimage, matplotlib, tensorflow, and keras, to facilitate data processing, image manipulation, and model training.

Although this setup is sufficient for smaller datasets and models, it may struggle with very large datasets or complex models that require higher VRAM. The experiment was done using the deep learning framework for Python, TensorFlow, and Keras.

5.2 Evaluation Metrics

To ensure a robust assessment beyond a single measure, we have utilized multiple metrics including accuracy, precision, recall, and F1-score. These metrics comprehensively evaluate our model's performance. The various evaluation metrics utilized in this research can be represented through the following equations:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5.1}$$

$$Precision = \frac{TP}{TP + FP} \tag{5.2}$$

$$Recall = \frac{TP}{TP + FN} \tag{5.3}$$

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5.4}$$

Here, TP stands for correctly predicted positive cases or true positive. TN is correctly predicted negative cases. FP and FN are false positive and false negative. These are the cases that are incorrectly predicted. Although accuracy is a widely used metric, it does not always reflect the model's true performance, especially in imbalanced datasets. By incorporating precision, recall, and F1-score, we ensure a more reliable evaluation that considers both false positives and false negatives. These metrics help in optimizing the model for real-world applications where misclassification costs vary depending on the context.

5.3 Experimental Findings

5.3.1 Performance Analysis:

Class	Precision	Recall	F1-Score	Support
COVID	0.98	0.96	0.97	1080
Lung Opacity	0.90	0.91	0.90	1835
Normal	0.94	0.94	0.94	3049
Viral Pneumonia	0.95	0.97	0.96	386
Accuracy			0.94	6350
Macro Avg	0.94	0.94	0.94	6350
Weighted Avg	0.94	0.94	0.94	6350

Table 5.1: Performance Metrics Value of the C-RNet Model

The analysis result provides significant details regarding how well our Customized CNN-RNN model classified lung diseases. In four categories— COVID, Lung Opacity, Normal and Viral Pneumonia—the model exhibits outstanding precision, recall, and f1-scores.

According to the categorization findings, the COVID class performs exceptionally well, attaining 98% precision and 96% recall, which results in a f1-score of 97%. This demonstrates the model's high accuracy in detecting COVID cases with few false positives. With 90% precision and 91% recall, Lung Opacity again exhibits strong performance, providing a f1-score of 90%. The model may misclassify certain situations, though, as indicated by the minor decline in precision, which points to a possible area for additional development.

With 94% precision and 94% recall, the Normal category does well overall, generating a 94% f1-score. In medical diagnostics, this is especially important because correctly identifying normal situations lowers the chance of unwanted medications.

The model performs exceptionally well for viral pneumonia, achieving 95% precision and 97% recall with a 96% f1-score. This shows how well the model can identify cases of viral pneumonia while minimizing false negatives.

The C-RNet model's average accuracy of 94% shows how reliable and strong it is as a method of diagnosis in medical imaging. The model's reliability in accurately categorizing all categories is further supported by the high weighted average and macro scores (both 94%).

Moreover, our C-RNet model has an AUC-ROC score of 0.9904, demonstrating high discriminatory performance in diagnosing lung illnesses. This score indicates that the model minimizes false positives and false negatives by successfully distinguishing between diseased and normal cases. This performance demonstrates our hybrid CNN-RNN architecture's ability to handle a wide range of imaging situations and variances in X-ray datasets. Additionally, the confusion matrix of our proposed C-RNet model can be found in the figure 5.1. It is promising for practical medical use in lung disease identification utilizing X-ray pictures is highlighted by its success across all classes, particularly in accurately differentiating between various lung diseases.

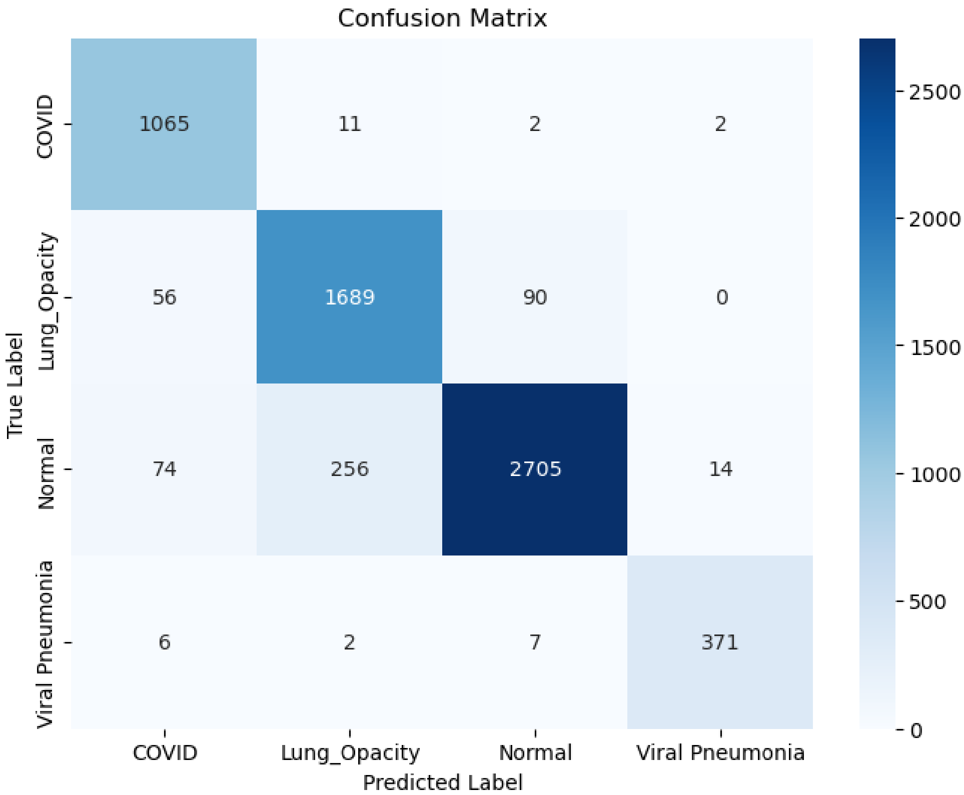


Figure 5.1: Confusion Matrix of C-RNet Model

5.3.2 Result Analysis

We utilized various CNN models in our research on lung disease classification using CNN and RNN approaches. The models tested include Xception (k-fold), ViT, VGG19, Inception LSTM, and our proposed C-RNet model. The accuracy of each of these five models was evaluated. The accuracy of the models is visually shown in the Figures 5.2–5.6.

The proposed **C-RNet** approach for detecting lung diseases performed exceptionally well, striking a balance between accuracy and effectiveness. With a validation accuracy of 93.51% and a training accuracy of 96.88%, the model demonstrated a remarkable capacity for generalization to new data. This excellent performance demonstrates how well the model detects lung conditions while preserving computational performance. Our model's improved convolutional architecture, which is created especially to identify complex patterns in CXR images, takes responsibility for its performance.

The depthwise segmented convolutions in the **Xception** model, an incredibly effective machine learning architecture, improve efficiency while lowering computational expenses. The model's outstanding capacity to apply generalization successfully to unseen data was demonstrated by its 94.50% training accuracy and 91.16% validation accuracy. This result demonstrates how well Xception balances accuracy and computational economy in the diagnosis of lung diseases.

This investigation also showed promising results for the **ViT** paradigm, which uses self-awareness strategies to process images comprehensively. The ViT model achieved 79.95% validation accuracy and 86.12% training accuracy. The effectiveness of ViT is attributed to its capacity to identify extended relationships in medical imaging, providing an alternative yet successful method for classifying lung diseases.

The **VGG19** method that adds more layers to the classic VGG19 architecture, demonstrated excellent outcomes in lung disease identification, with training accuracy of 83.17% and validation accuracy of 86.99%, demonstrating the model's capacity to retrieve while keeping complex features from X-ray images. However, its effectiveness might still be affected by overfitting, as evidenced by the small discrepancy between training and validation accuracy, indicating a usual issue in deep learning methods for medical image categorization.

Inception LSTM together showed a marginally different success curve. Both the validation and training accuracy of this model are 64.89% and 61.94%, respectively. When paired with temporal modeling using LSTM, the accuracy rates demonstrate the intricacy of implementing deeper, hybrid architectures like Inception, even though they are lower than those of more conventional methods like VGG19. The findings imply that although the hybrid model is capable of capturing spatial characteristics from the X-ray pictures, the temporal dependencies provided by the LSTM would not be as useful for this specific purpose and would need to be further adjusted or refined.

To conclude, our research offers a thorough evaluation and comparison of various CNN architectures for the purpose of classifying lung conditions from X-ray images. The outcomes reveal that although all models have benefits, the Custom C-RNet model that was developed specifically for this purpose, exhibits the most potential in terms of accuracy and efficiency. The findings of this study could greatly progress the creation of automated diagnostic tools for the medical imaging sector.

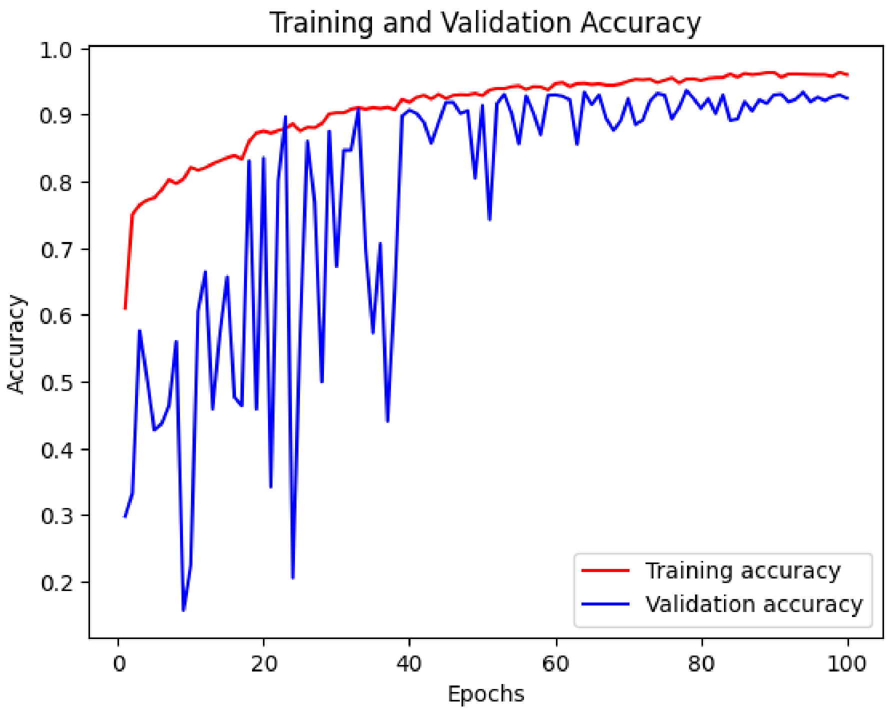


Figure 5.2: Accuracy of C-RNet Model

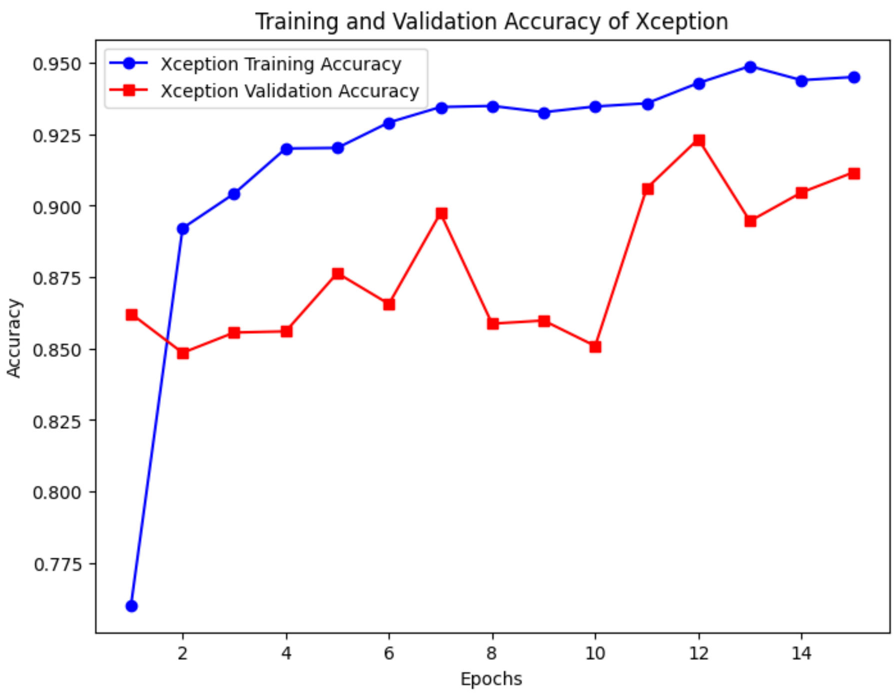


Figure 5.3: Accuracy of Xception Model

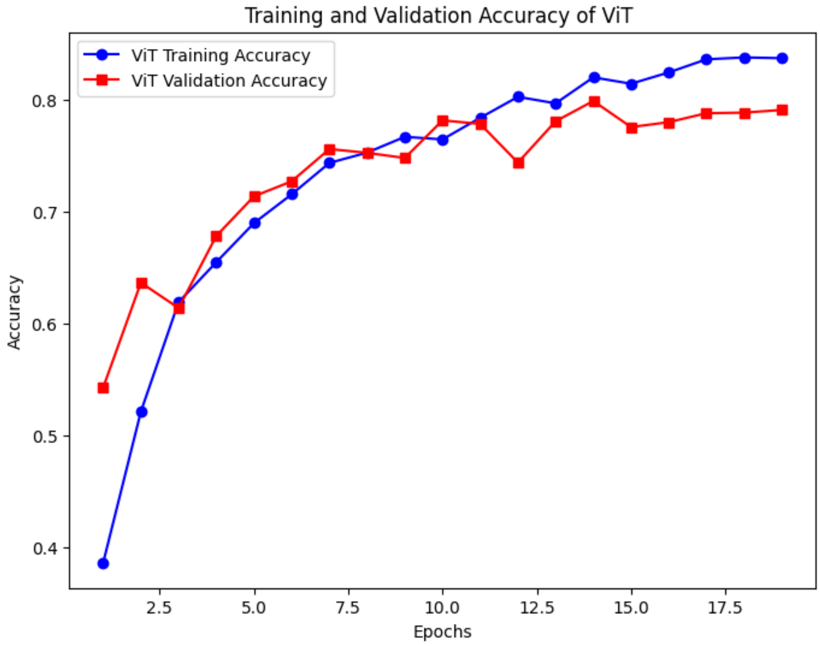


Figure 5.4: Accuracy of ViT

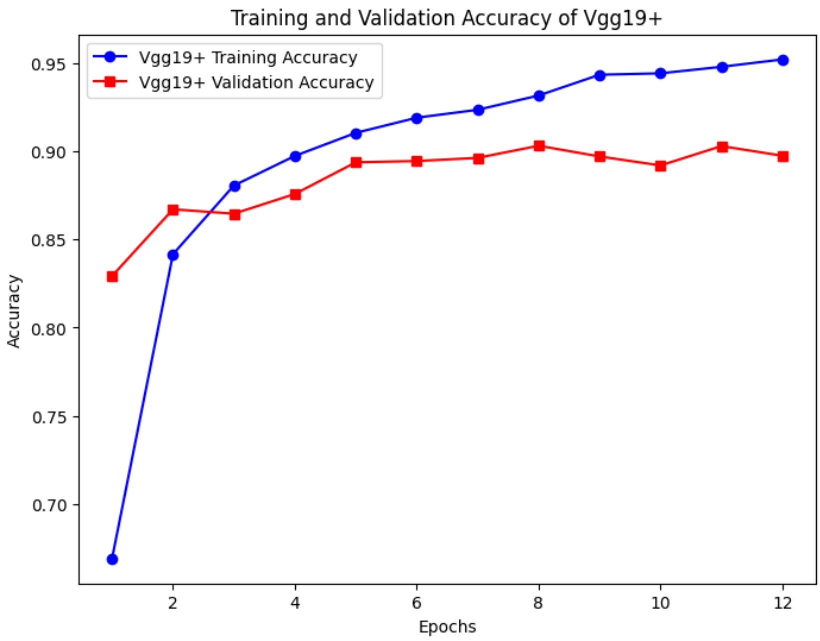


Figure 5.5: Accuracy of VGG19

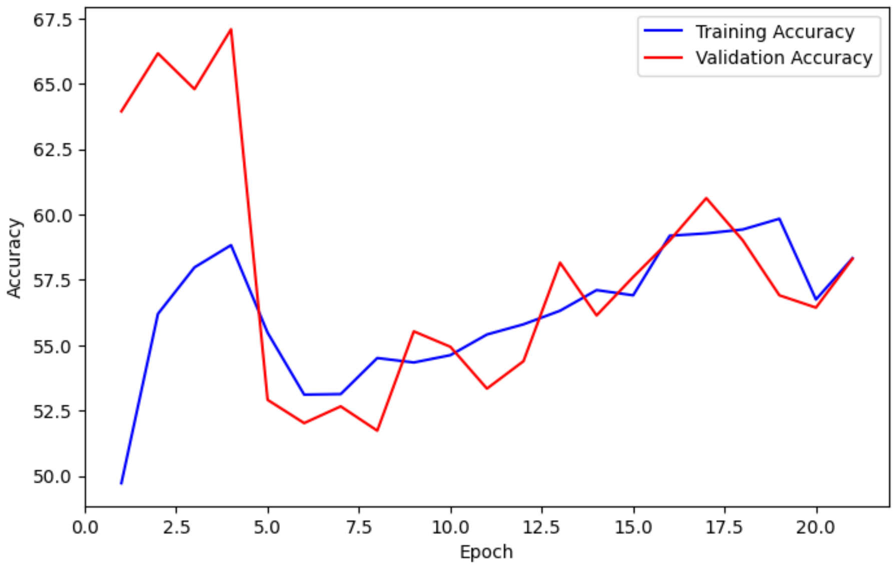


Figure 5.6: Accuracy of Inception LSTM

5.4 Results Comparison

This section compares how well different deep learning methods perform when used to classify lung diseases from X-ray pictures. When compared with the different designs that were examined, such as Xception, ViT, VGG19, and Inception LSTM, our suggested model performed significantly better. The table 5.2 clearly demonstrates the accuracy value of all the models.

Model	**Training Accuracy**	**Test Accuracy**	**f1-score**	**Model Size**
C-RNet	96.88%	93.51%	94%	7.25 MB
Xception	94.50%	91.16%	92%	81.59 MB
ViT	86.12%	79.95%	80%	26.09 MB
VGG19	83.17%	86.99%	89%	82.89 MB
Inception LSTM	61.94%	64.89%	58%	91.64 MB

Table 5.2: Comparison of Models Performance

C-RNET: Our suggested C-RNet model demonstrated its remarkable capacity to classify lung disorders with an impressive test accuracy of 93.51% and training accuracy of 96.88%. Additionally, its 94 f1-score shows high recall and precision, guaranteeing accurate prediction. Its small model size of only 7.25 MB is another significant benefit. Model's remarkable capacity to correctly detect lung illnesses from X-ray pictures is demonstrated by these excellent accuracy ratings. The remarkable outcomes imply that the model's architecture was extremely tailored for the particular objective of identifying lung diseases, which qualifies it for use in practical medical environments.

Xception: With a f1-score of 92% and test and training accuracy of 91.16% and 94.50%, respectively, the Xception model demonstrated strong performance. Although it did effectively, it fell a little short of our suggested model, suggesting that although Xception is effective at feature extraction. Furthermore, its higher model size (81.59 MB) raises the possibility that it may need more processing power, which would make it less suitable for deployment that is lightweight. There might be space for improvement in terms of accuracy and efficiency for the categorization of lung diseases.

ViT: With a f1-score of 80, ViT obtained 79.95% test accuracy and 86.12% training accuracy. Despite using a transformer-based strategy, it performs worse than Xception and C-RNet. Although its model size of 26.09 MB is reasonable, the lower cosine similarity (83.32) indicates that the feature extraction is not as good as it may be. The reduced validation accuracy raises the possibility that ViT needs to be further improved and adjusted in order to adequately capture the subtleties of medical image classification. In general, ViT is less accurate and reliable than our suggested C-RNet model, which makes it less useful for classifying lung diseases.

VGG19: VGG19 had a f1-score of 89% and demonstrated test and training accuracy of 86.99% and 83.17%, respectively. Its large model size of 82.89 MB is a major disadvantage despite its excellent performance. Even though the model produced encouraging outcomes, our suggested model and Xception outscored it, suggesting that while VGG19 is a strong model, its structure may require additional tuning to meet the requirements of the lung illness classification challenge. Hence, it is ineffective for real implementation due to its significantly bigger size in contrast to our suggested C-RNet model.

Inception LSTM : With a f1-score of 58%, the Inception LSTM model only achieved test accuracy of 64.89% and training accuracy of 61.94%, indicating a severe under performance. Although the model is renowned for its effectiveness in specific tasks, it performed noticeably worse in this situation than the other models. Furthermore, model's size of 91.64 MB is very huge, rendering it inappropriate for efficient lung disease classification. This shows that the complicated character of the task and model's constraints in addressing this particular area may prevent Inception LSTM's complicated structure from being properly adapted for lung disease categorization from X-ray pictures.

In summary, our suggested model performs better than the others in terms of loss metrics and accuracy. The structure and specialist modification probably contributed to its enhanced lung disease identification abilities using X-ray pictures. The outcome of our suggested model, which stands out as the best option for lung illness identification in this environment, was superior to that of models like Xception and VGG19, despite the fact that they both performed well. Therefore, The most suitable choice for identifying lung diseases from X-ray images is our suggested C-RNet model, which strikes the ideal combination between effectiveness, reliability, and efficacy. The Figure 5.7 visualize the comparison of testing, training and f1 score metrics of all the models.

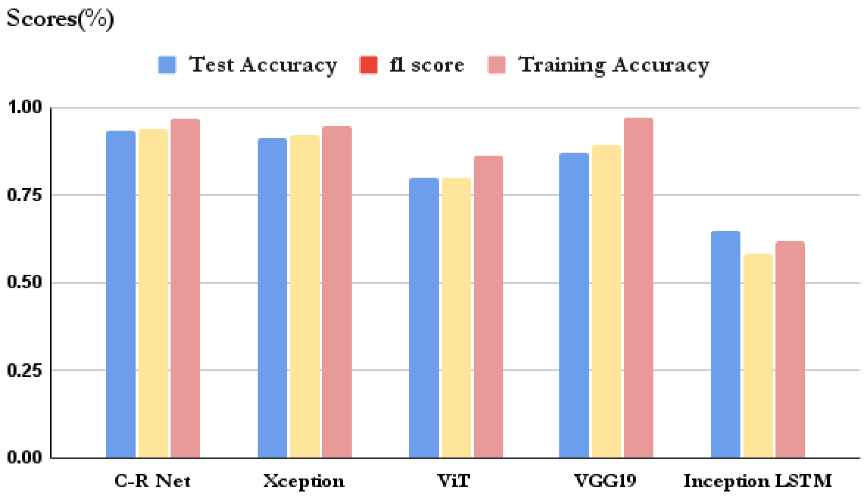


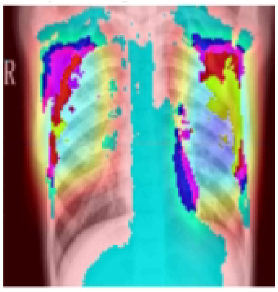
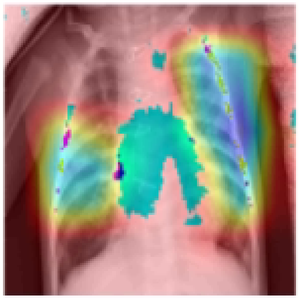
Figure 5.7: Testing, Training and F1 Score Comparison of Models

5.5 Grad-CAM Comparison and Analysis:

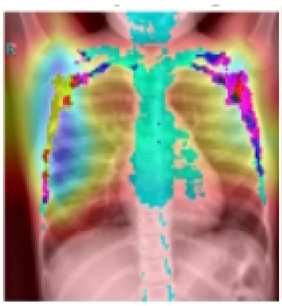
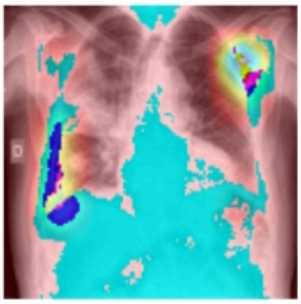
In order to improve the evaluation of our C-RNet model, we have used Grad-CAM in our study. Grad-CAM is an effective technique for identifying and analyzing the areas of a picture that have a significant effect on the model's predictions. This method displays how particular picture attributes influence the CNN-RNN model's decision-making process and offers an understanding of the internal functioning of the system.

We concentrate on the 'conv2d' layer of our architecture while using Grad-CAM. We create a visualization called heat map that shows the most important regions in the X-ray pictures by calculating the gradients of the projected class in relation to the feature maps of the corresponding layer. We are then able to visibly connect the model's concentration with real lung features by superimposing this heat map across the initial X-ray image.

This visualization method is essential for confirming the model's precision and dependability. Grad-CAM increases trust in the CNN-RNN model's clinical skills and improves the understanding of its outputs by visibly displaying the areas of significance. This feature is especially crucial in medical imaging, where attaining high accuracy is only as crucial as comprehending the logic underlying a model's predictions. As a result, our customized C-RNet model for classifying lung diseases from X-ray pictures is now much more transparent, reliable and interpretable because of the inclusion of Grad-CAM. The Grad-CAM visualization of our proposed model can be seen in the Figure 5.8.

(a) Viral Pneumonia (b) Lung Opacity

(c) Covid (d) Normal

Figure 5.8: Grad-CAM Visualization of Diseases Using C-RNet Model

Chapter 6

Discussion

6.1 Key Findings

Our proposed model classifies lung conditions including COVID-19, lung opacities, normal and viral pneumonia from the dataset and to achieve our goal we integrate CNN and BiLSTM networks. Our research uses CNN to extract spatial features from chest X-ray images and BiLSTM for sequential pattern recognition. The primary outcomes of our study indicate the efficiency of our C-RNet model in classifying lung diseases. The CNN module captures local spatial information, guaranteeing that complicated patterns in radiographic pictures are adequately represented. Furthermore, the biLSTM branch improves performance by examining the dependencies between extracted features, allowing the model to grasp temporal interactions. This combination allows the RNN to identify variations in pixel flow, which is essential for disease detection. These findings clearly answer the original study question by illustrating how our model uses both geographical and temporal information to increase classification accuracy, resulting in a viable tool for lung disease diagnosis.

6.2 Why Choose This Model Over Others?

Although there are several deep learning models for medical image classification. But our method is completely different from the rest. This is because our method improves both spatial and sequential feature extraction at the same time. Most traditional CNN-based models rely on spatial feature extraction without considering contextual relationships [34]. Typically, CNNs extract features from images and pass them to RNNs as input. In this method, sequence size matching is required, and sometimes spatial relationships can be lost due to the sequence size sorting. In contrast, our model incorporates an advanced patch-based augmentation technique that improves generalization across different imaging conditions. Hong et al. [35] also proposed a multi-class classification method of lung disease using CNN model. But, their model only collected CXR pictures that are only in the PA (Posteroanterior) and AP (Anteroposterior) positions, with no indication of the disease's exact location. In contrast, our C-RNet model solves this restriction by using Grad-CAM visualization, which highlights specific parts of the X-ray where abnormalities are found. This not only enhances interpretability, but also aids in clinical decision-making by providing visual evidence of impacted areas. As a result, our model improves both classification accuracy and explainability, making it better suited for real-world medical applications. The study by Sulaiman et al. [36] does not address picture quality variability or the resemblance between normal and abnormal lung areas, and the ideal choice of k for a CNN model depends on variables such as dataset size. Our approach, on the other hand, improves its robustness in a variety of scenarios by handling image quality variations and performing well across several datasets [37]. Moreover, another study by Khan et al. [38] performs well, but it requires a lot of processing power, which can make it less useful in actual clinical settings. Our C-RNet model, on the other hand, is made with computational efficiency in mind, guaranteeing quicker processing without sacrificing accuracy. This makes our method better suited for implementation in healthcare environments with limited resources, where efficiency and performance must be balanced.

6.3 Contextualizing Findings

Our findings show that the C-RNet model surpasses earlier techniques for lung disease categorization. Traditional studies have generally used CNNs for feature extraction or CNN-extracted features as RNN inputs [39]. In contrast, our hybrid model improves performance by processing entire images in the CNN branch while supplying 256 image patches as time-series data to the RNN. This dual-stream technique considerably increases the model's robustness to different image resolutions and noise distortions, making it more versatile across a wide range of datasets.

6.4 Recommendations for Future Research

Future work should explore incorporating additional pre-processing techniques such as histogram or various filters, contrast enhancement, and segmentation-based feature extraction to improve model input quality. Additionally, integrating multi-modal data sources including CT scans, clinical metadata, and laboratory results can significantly improve classification performance.

Chapter 7

Conclusion

This study suggests a new deep learning-based method for identifying lung diseases by combining RNN with CNN from chest X-Ray images. We created a customized Convolutional Recurrent Network (C-RNet) that successfully combines CNN for global feature collection and RNN for sequential reliance training in order to improve the precision and effectiveness of lung disease diagnosis. Our work employed a publicly accessible data set to categorize chest X-ray pictures into four different groups. The suggested approach improves the precision of classification by effectively capturing the continuity and dependency characteristics of subsequent levels. The primary finding of this research demonstrates the combined strategy functions better than conventional single-scale techniques. With a 93.51% rate of success on the entire dataset, our model shows promise for improving diagnostic performance. Our suggested C-RNet model's compact size of just 7.25MB is another significant advantage; it is noticeably smaller than other pretrained models that we trained and assessed on the same dataset. Because of its lightweight design, it is ideal for practical clinical implementation in Bangladesh's rural areas, where clinical resources and processing power are scarce. Additionally, we guarantee improved readability and dependability for clinical imaging-based disease detection by utilizing XAI solutions such as Grad-CAM. The evaluation criteria confirms that our strategy succeeds. We think that our research will help improve lung disease early identification and medication, especially in places with limited resources. To ensure greater practicality in medical applications, our future research will focus on further optimization of our proposed model's performance and investigating its adaptability to various respiratory illnesses.

Bibliography

[1] Directorate General of Health Services (DGHS). *DGHS Data Portal*. Accessed: [insert date]. 2025. URL: https://old.dghs.gov.bd/index.php/en/data.

[2] *icddr,b - Pneumonia and other respiratory diseases — icddrb.org*. https://www.icddrb.org/news-and-events/press-corner/media-resources/pneumonia-and-other-respiratory-diseases. [Accessed 23-05-2024].

[3] *Pneumonia kills 24,000 children in Bangladesh every year: icddr,b — tbsnews.net*. https://www.tbsnews.net/bangladesh/health/least-2-3-children-die-pneumonia-bangladesh-every-hour-icddrb-736410. [Accessed 23-05-2024].

[4] Ali Narin, Cihat Kaya, and Zeki Pamuk. "Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks". In: *Pattern Analysis and Applications* 24 (2021), pp. 1207–1220. DOI: 10.1007/s10044-021-00984-y.

[5] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. http://www.deeplearningbook.org. MIT Press, 2016.

[6] Tawsifur Rahman. *COVID-19 Radiography Database*. Accessed: 2025-01-29. 2020. URL: https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database.

[7] Pranav Rajpurkar et al. "Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists". In: *PLoS medicine* 15.11 (2018), e1002686.

[8] Alistair EW Johnson et al. "MIMIC-CXR-JPG, a large publicly available database of labeled chest radiographs". In: *arXiv preprint arXiv:1901.07042* (2019).

[9] Mohammad Tariqul Islam et al. "Abnormality detection and localization in chest x-rays using deep convolutional neural networks". In: *arXiv preprint arXiv:1705.09850* (2017).

[10] Md Khairul Islam et al. "Enhancing lung abnormalities detection and classification using a Deep Convolutional Neural Network and GRU with explainable AI: A promising approach for accurate diagnosis". In: *Machine Learning with Applications* 14 (2023), p. 100492. ISSN: 2666-8270. DOI: https://doi.org/10.1016/j.mlwa.2023.100492. URL: https://www.sciencedirect.com/science/article/pii/S2666827023000452.

[11] Baljinnyam Dayan. "Lung Disease Detection with Vision Transformers: A Comparative Study of Machine Learning Methods". In: (2024).

[12] Tanzina Taher Ifty et al. "Explainable Lung Disease Classification from Chest X-Ray Images Utilizing Deep Learning and XAI". In: (2024).

[13] Syed Usama Khalid Bukhari et al. “The diagnostic evaluation of Convolutional Neural Network (CNN) for the assessment of chest X-ray of patients infected with COVID-19”. In: *MedRxiv* (2020), pp. 2020–03.

[14] Nandkishor Chhagan Patil and Nitin Jagannath Patil. “Lung Cancer Detection Using CNN VGG19+ Model”. In: *Journal of Electrical Systems* 20.3 (2024), pp. 541–550.

[15] Ban Jawad Khadhim et al. “Diagnose COVID-19 by using hybrid CNN-RNN for chest X-ray”. In: *Indones. J. Electr. Eng. Comput. Sci* 29.2 (2023), pp. 852–860.

[16] Isoon Kanjanasurat et al. “CNN–RNN network integration for the diagnosis of COVID-19 using chest X-ray and CT images”. In: *Sensors* 23.3 (2023), p. 1356.

[17] Harsh Sharma et al. “Feature extraction and classification of chest x-ray images using cnn to detect pneumonia”. In: *2020 10th international conference on cloud computing, data science & engineering (Confluence)*. IEEE. 2020, pp. 227–231.

[18] M Abubakar, I Shah, W Ali, et al. “Classification of Pneumonia and Tuberculosis from Chest X-rays”. In: *arXiv preprint arXiv:2103.14562* (2021).

[19] Linh T Duong et al. “Detection of tuberculosis from chest X-ray images: Boosting the performance with vision transformer and transfer learning”. In: *Expert Systems with Applications* 184 (2021), p. 115519.

[20] Hassaan Malik et al. “CDC_Net: Multi-classification convolutional neural network model for detection of COVID-19, pneumothorax, pneumonia, lung Cancer, and tuberculosis using chest X-rays”. In: *Multimedia Tools and Applications* 82.9 (2023), pp. 13855–13880.

[21] Rohit Kundu et al. “Pneumonia detection in chest X-ray images using an ensemble of deep learning models”. In: *PloS one* 16.9 (2021), e0256630.

[22] Daniel Kermany, Kang Zhang, Michael Goldbaum, et al. “Labeled optical coherence tomography (oct) and chest x-ray images for classification”. In: *Mendeley data* 2.2 (2018), p. 651.

[23] Safa Ben Atitallah et al. “Randomly initialized convolutional neural network for the recognition of COVID-19 using X-ray images”. In: *International journal of imaging systems and technology* 32.1 (2022), pp. 55–73.

[24] Yogesh H Bhosale et al. “Deep convolutional neural network based Covid-19 classification from radiology X-Ray images for IoT enabled devices”. In: *2022 8th international conference on advanced computing and communication systems (ICACCS)*. Vol. 1. IEEE. 2022, pp. 1398–1402.

[25] Mahesh Gour and Sweta Jain. “Uncertainty-aware convolutional neural network for COVID-19 X-ray images classification”. In: *Computers in biology and medicine* 140 (2022), p. 105047.

[26] Neha Muralidharan et al. “Detection of COVID19 from X-ray images using multiscale Deep Convolutional Neural Network”. In: *Applied Soft Computing* 119 (2022), p. 108610.

[27] Ibrahim Abdulrab Ahmed et al. “Multi-techniques for analyzing x-ray images for early detection and differentiation of pneumonia and tuberculosis based on hybrid features”. In: *Diagnostics* 13.4 (2023), p. 814.

[28] Mohammed Salih Ahmed et al. “Joint diagnosis of pneumonia, COVID-19, and tuberculosis from chest X-ray images: a deep learning approach”. In: *Diagnostics* 13.15 (2023), p. 2562.

[29] Saad I Nafisah and Ghulam Muhammad. “Tuberculosis detection in chest radiograph using convolutional neural network architecture and explainable artificial intelligence”. In: *Neural Computing and Applications* 36.1 (2024), pp. 111–131.

[30] M Abubakar, I Shah, W Ali, et al. “Classification of Pneumonia and Tuberculosis from Chest X-rays”. In: *arXiv preprint arXiv:2103.14562* (2021).

[31] Sandhya Sharma et al. “Performance evaluation of the deep learning based convolutional neural network approach for the recognition of chest X-ray images”. In: *Frontiers in oncology* 12 (2022), p. 932496.

[32] Rowel Atienza. *Advanced Deep Learning with TensorFlow 2 and Keras: Apply DL, GANs, VAEs, deep RL, unsupervised learning, object detection and segmentation, and more*. Packt Publishing Ltd, 2020. ISBN: 978-1-83882-165-4. URL: https://www.packtpub.com/product/advanced-deep-learning-with-tensorflow-2-and-keras/9781838821654.

[33] United Nations in Bangladesh. *The disease burden of air pollution on children continues to rise in Bangladesh, according to latest State of Global Air 2024 report*. Accessed: 2025-01-31. June 2024. URL: https://bangladesh.un.org/en/272029-disease-burden-air-pollution-children-continues-rise-bangladesh-according-latest-state.

[34] Samiul Islam, Md Maksudul Haque, and Abu Jobayer Md Sadat. “Capturing spectral and long-term contextual information for speech emotion recognition using deep learning techniques”. In: *arXiv preprint arXiv:2308.04517* (2023).

[35] Min Hong et al. “Multi-class classification of lung diseases using CNN models”. In: *Applied Sciences* 11.19 (2021), p. 9289.

[36] Adel Sulaiman et al. “A convolutional neural network architecture for segmentation of lung diseases using chest X-ray images”. In: *Diagnostics* 13.9 (2023), p. 1651.

[37] Tawsifur Rahman. *Tuberculosis (TB) Chest X-ray Dataset*. Accessed: February 6, 2025. 2023. URL: https://www.kaggle.com/datasets/tawsifurrahman/tuberculosis-tb-chest-xray-dataset.

[38] Ali Yousuf Khan et al. “X-Ray Image-Based Real-Time COVID-19 Diagnosis Using Deep Neural Networks (CXR-DNNs)”. In: *Journal of Imaging* 10.12 (2024), p. 328.

[39] Jaehoon Koo, Diego Klabjan, and Jean Utke. “Combined convolutional and recurrent neural networks for hierarchical classification of images”. In: *2020 IEEE International Conference on Big Data (Big Data)*. IEEE. 2020, pp. 1354–1361.