Addis Ababa University

College of Technology and Built Environment

School of Electrical and Computer Engineering

Automated Classification and Localization of Thoracic Diseases from X- Ray Images Using Deep Learning

By: Biruk Tena Siyoume

Advisor: Dr. Menore Tekeba

A Thesis Submitted to Addis Ababa College of Technology, School of Graduate Studies, Addis Ababa University in Partial Fulfillment of the Requirement for the Degree of Master of Science in Computer Engineering

Addis Ababa, Ethiopia

June, 2025

Addis Ababa University

College of Technology and Built Environment

School of Electrical and Computer Engineering

This is to certify that the thesis prepared by Biruk Tena, entitled: *Automated Classification and Localization of Thoracic Diseases from Chest x- ray Images Using Deep Learning,* and Submitted in partial fulfilment of the requirements for the Degree of Master of Science in Computer Engineering compiles with the regulations of the University and meets the accepted standards with respect to originality and quality.

Approval by Board of Examiners:

Dr. Sosina Mengistu		
Dean, School of Electrical & Computer Engineering	Signature	Date
Dr. Menore Tekeba		
Advisor	Signature	Date
Dr. Bisrat Derebssa		
Internal Examiner	Signature	Date
Dr. Fitsum Assamnew		
External Examiner	Signature	Date

Declaration

I, the undersigned, certify that research work titled Automated Classification and Localization of Thoracic Diseases from chest x- ray images Using Deep Learning is my own work. The work has not been presented elsewhere for assessment. Where material has been used from other sources, it has been properly cited.

Biruk Tena Siyoume		
Name	Signature	Date

Acknowledgement

First of all, I would like to thank the Almighty God for giving me the strength and His showers of blessing through all challenging journey of the thesis.

My deepest gratitude is to my advisor, Dr. Menore Tekeba, for his encouragement, guidance, support and enthusiasm with his knowledge. His guidance helped me in all the time of the research and writing of this thesis.

My sincere thanks also go to my wife Hiwot Mamo and My son Nathan Biruk for your patience and for supporting me throughout all my studies. My special thanks to My sister who is everything to me and my brothers.

Last but not the least, I would like to thank all my friends and classmates for the time we had together.

Table of Contents

List of figures

List of Tables

List of Abbreviation

CXR	Chest X-Ray
CAD	Computer-Aided Detection
TDCL	Thorax Disease Classification and Localization
ACC	Accuracy
CAM	Class Activation Map
CNN	Convolutional Neural Network
DNN	Deep Neural Network
CT	Computer Tomography
MRI	Magnetic Resonance Imaging
US	Ultrasound
ILSVRC	ImageNet Large-Scale Visual Recognition Competition
ROI	Regions-of-Interest
COPD	Chronic Obstructive Pulmonary Disease
DCNN	Deep Convolutional Neural Network
SVM	Support Vector Machine
NLP	Natural Language Processing
NIHCC	National Institute of Health Clinical Center
PACS	Picture Archiving and Communication System
CSV	Comma-Separated Values
DIP	Digital Image Processing
JPEG	Joint Photographic Experts Group
CDF	Cumulative Distribution Function
ML	Machine Learning
MLP	Multi-Layer Perceptron
RNN	Recurrent Neural Network
DBN	Deep Belief Network
ReLUs	Rectified Linear Units
BPNN	Backpropagation Neural Network
MSE	Mean Square Error

TPR	True Positive Rate
FPR	False Positive Rate
CTR	Cardiothoracic Ratio
MSRG	Multiscale Region Growing
KNN	K^{th} Nearest Neighbor

Abstract

In developing countries, thorax diseases are the common cause of death. Chest X-ray (CXR) is the most common and effective procedure for the diagnosis of a wide variety of thorax diseases. However, an efficient chest x-ray interpretation requires experienced radiologists, which is particularly lacking in a developing country such as Ethiopia. Efficient computer-aided detection (CAD) of Chest X-ray that can assist radiologists to diagnose lung diseases and offer huge benefits to public healthcare. Currently ChestX-ray14 is the largest public dataset for CXR that includes 14 different pulmonary diseases and one normal class (such as atelectasis, cardiomegaly, consolidation, edema, effusions, emphysema, Fibrosis, Hernia, infiltrations, masses, nodules, no-finding, pneumonia, pneumothorax and Pleural Thickening); This large data set has opened the possibility of creating a better CAD system in radiology. In this study, the proposed Automated Thorax disease Classification and localization (ATDCL) has four main stages, the preprocessing stage, the segmentation stage, the classification stage and the localization stage. Image scaling, normalization, histogram equalization, and median filtering were employed in the preprocessing stage to improve the quality and remove noise from the input images. We have used U-net deep neural network for segmentation. The network composes of two paths: the down-sampling path and the up-sampling path, also known as encoder and decoder. The encoder part used to extract features from the image and then from these latent features, the decoder path will learn to reconstruct the high-resolution binary mask. Then the region of interest was taken out from the chest x-ray image to utilize as input for the next sub stages. In the thorax classification stage, we employed the inception model, a novel deep learning approach. We have achieved the Average accuracy value (Acc) of 0.8489 and we have used a Class Activation Map (CAM) method that can significantly improve the understanding of radiologists about the approximate location of the existed pathology. Our approach can provide radiologists a way to make quicker and more reliable decisions, and providing greater health care system.

Keywords*: Automated Thorax disease Classification and localization (ATDCL), Chest X-ray (CXR), Deep Learning (DL), Receiver Operating Characteristic (ROC), Area Under Curve (AUC), Class Activation Map (CAM), Computer-aided Detection (CAD), Convolutional Neural Network (CNN), deep neural network (DNN).*

Chapter 1

Introduction

1.1 Background

The rapid growth of storage technologies and digital image collection in a range of computer vision applications, computer programs' image understanding have become an interesting and active topic for machine learning field and in application-specific studies [1]. Significantly enhanced quantitative performance in object recognition, detection, classification, and segmentation has been demonstrated in compared to earlier machine learning approaches up on hand-crafted image features. Medical x-rays are a part of medical imaging that are generally used to diagnose some sensitive human body parts such as lungs(chest), bones, teeth, skull, etc. Chest x-ray images can show chest diseases like atelectasis, cardiomegaly, consolidation, edema, effusions, emphysema, Fibrosis, Hernia, infiltrations, masses, nodules, pneumonia, pneumothorax and Pleural Thickening. Radiologists challenge to classify abnormalities on chest x-rays images. As a result, throughout the last few decades, computer-aided diagnosis (CAD) systems have been created to extract meaningful information from X-rays in order to assist doctors in gaining a quantitative understanding of an X-ray. In recent years, there has been a significant increase in the use of machine learning applications in the healthcare sector.

Medical imaging is a method used to generate images of the human body for the purpose of clinical diagnosis. Medical imaging techniques such as Computer Tomography (CT), Magnetic Resonance Imaging (MRI) and Ultrasound (US) are widely used for the diagnosis of abnormalities like lung nodule detection, skin lesion detection, breast cancer detection. CT can produce images with significantly higher spatial resolution in much less time. MRI and US take up the advantage of not exposing the subject to ionizing radiation [2]. In the framework of medical imagery, segmentation is a separation of an image domain into non-overlapping areas of pixels that are consistent with important anatomical or abnormal structures such as tumors or lesions. In many diagnostic and surgical procedural processes, the segmentation of anatomic organs is vital and it is also useful for computer-aided diagnosis. Segmentation algorithms are categorized into three types: supervised, unsupervised, and interactive. Supervised segmentation methods require manually labeled training data to recognize a specific area of image interest that

may restrict the scope of the method [2]-[3]. Uncontrolled or unsupervised (automatic) methods give segmentation results with no previous input image information and do not require manual intervention [4]. In general, the Uncontrolled or unsupervised methods of segmentation limit the segmentation of complex areas of interest[5, 6]. The combination of human experts and machine intelligence has strengthened the accuracy, precision and efficiency of the interactive segmentation [3]. Interaction segmentation takes an important place for various applications such as the localization of tumors, tissue volume estimation, computer-assisted operation and disease diagnosis [7]. The features are extracted for gradation and recognition from a segmented region of interest (e.g. classification in medical exam results like normal, benign and malignant) [8].

Medical images typically have interior structures with varying textures and pixel density; if we solely employed traditional features to categorize medical images, it would be impossible to efficiently characterize specific classes [9]. In the past few years, deep learning has become one of the hottest researcher areas in computer science and computer applications. As deep learning has advanced, numerous researchers have attempted to apply this new technique to nonmedical imagery. The study in [10] primarily addressed the deep model's framework. To solve imaging difficulties, a number of deep models have been proposed. [11] developed a deep convolutional neural network to identify images in the ImageNet large-scale visual recognition competition 2010 (ILSVRC - 2010), attaining world-class performance. In [12], the deep-model depth impact on image recognition performance was discussed by the Authors. This approach has already been inspired such successful research. Significant efforts to leverage this new approach to the problem of classifying medical images. The classification of medical images has been one of the major issues in the field of imaging, with the objective of classifying medical images into various categories to help physicians in the diagnosis or research of diseases. In general, medical imaging classification can be divided into two parts. The first step is extracting effective features from the image. The second step is using the features to build models that classify the image dataset. Typically, doctors used their professional experience to extract features from medical images in order to classify them into distinct classifications, which is a complex, time-consuming, and tedious task. This approach is prone to instability or repetitive results. In view of the research so far, research into the classification of medical images was extremely useful. A number of published studies were carried out by the researchers in this field. But we cannot still carry out this task effectively at present.

consider past studies that [13]-[14] used shallow models for the classification of medical images which mainly depend on the shape, color, texture feature and their combinations before the appearance of deep architectures. But, for all these models, the major problem is that the features extracted are often referred to as low-level features, which lack the ability to represent high-level problem domain concepts and have a low general capacity. In contrast, in deep architectures [15]-[16] in the field of non-medical images, there was considerable success. Deep learning methods, the most amazing branch of the field of machine learning, provide an effective way to build a complete end-to -end model that compute final classification labels using the raw pixels of medical images.

In the medical image analysis domain, the application of deep models requires considerable effort to match other imaging aspects, since large datasets are required for deep architectures to obtain excellent features. Interestingly, medical images are usually difficult to obtain and therefore medical data sets are usually relatively small. Thus, the method can lead to overfitting of the model if we use a deep model directly to address a small dataset. Apart from these problems, the accuracy of the model has proved quite poor and training a deep model usually takes a high amount of computation. We used a moderate deep model to solve the demerit of the traditional features to overcome these concerns about traditional methods and can utilize deep architectures to automatically extract high-level features to classify medical images.

Because of advancements in digital medical imaging techniques and patient image-screening protocols over the last 10 years, the quantity of images acquired every day in any modern hospital has increased enormously. One immediate consequence of this trend is the enormous increase in the number of images that must be reviewed by radiologists. This has led to a concomitant demand for computerized tools to aid radiologists in the diagnostic process. Automated systems assist the radiologists in the diagnostic process by classifying images into predefined categories. This is done by learning from a large archive of image examples that are annotated by experts.

Image classification or categorization is the classification or labeling of images into predetermined categories. The principal challenge of image categorization is the capture of the most significant features within the images that facilitate the desired classification. A single image can contain a large number of regions-of-interest (ROI), each of which may be the focus

of attention for the medical expert, depending on the diagnostic task at hand. A single chest image, for example, includes the lungs, heart, rib cage, diaphragm, clavicle and blood arteries, anyone of which could be the center of attention. Focusing on a ROI inside the image that is related to the assumed pathology is one technique to improve the image categorization process. The advantage of the ROI approach is that the rest of the image can be ignored leading to computational advantages and increased accuracy in the classification.

Clinical decision support techniques based on automatic classification algorithms can produce a strong need to localize the area that is most relevant to the diagnostic task. A diagnostic system that provides the image region that was utilized to make the choice in addition to the decisions on the existence of a pathology can help radiologists to better comprehend the decision and assess its reliability. Another benefit of making a ROI-based decision is that we can create a detailed representation of the local region. We refer to this approach as ROI based image categorization. Of course, it is not suited to every image categorization task; not every pathology includes a significant and identifiable ROI that appears across the data set. In such cases, a global, full-image categorization is appropriate.

Image segmentation is a critical image processing method. Many applications, such as object synthesis or computer graphic images, necessitate precise segmentation.

1.2 Motivation of the Study

In our country, Ethiopia, TB and Lung cancer are become the worst lethal disease for humans from time to time, and their treatments started during at higher stage of the conditions or the abnormalities. This is due to manual based analysis of the lung cancer, tuberculosis and the symptoms complexity of the thoracic diseases [17]. Identifying of the diseases in the early stage and threatening it on time is very vital. Patients do not experience any signs of the disease until it spreads to the entire lung or chest organ, at which point treatment becomes extremely challenging. Thus, lung nodule detection is one hot research area on which many papers done on it and tuberculosis identification is another area on which motivates me all together to do this work. This paper classifies each chest or thoracic diseases into its patient categories by using multi label deep learning techniques.

Medical image analysis gets a new breakthrough technique due to deep learning. It has been used in the development of CAD systems that lower the lung cancer rate of death. To enhance early

diagnosis, anomalies in several types of medical imaging can be identified through deep learning. The use of deep learning methods for feature extraction improves the accuracy of medical image analysis in systems that detect chest diseases. Research[18, 19] suggests that deep learning continues to be the main focus of studies in order to improve the outcomes of traditional learning techniques and provide effective medical image analysis for x-ray images.

Problem Statement

Lung cancer and Chronic Obstructive Pulmonary Disease (COPD) become the world's fast growing and most serious issues, now a day. Especially, in Sub-Saharan African countries it becoming a risk challenge for health services due to rising numbers of patients. In 2011, the World Health Organization reported that 34% of Ethiopians died with non - communicable diseases, with a nationwide cardiovascular disease prevalence of 15%, cancer prevalence of 4%, and chronic obstructive pulmonary disease prevalence of 4% [20]. The Chest X-ray (CXR) is the most commonly requested radiologic examination because it is inexpensive, quick, and reliable in diagnosing many lungs and heart failure. Currently, Chest X-ray is the best available screening diagnosis for many diseases including pneumonia and cardiomegaly. However, an efficient interpretation of CXR requires radiologist years of experience and a tremendous amount of time. Furthermore, radiologist performance is frequently measured in terms of quantity [21] rather than the quality of their reports. This can lead to misinterpretation, low-quality report, unnecessary sub-sequence procedure and have a direct negative impact on patient health and medical infrastructure. An effective automatic CXR interpreter could provide a huge impact to improve radiologist workflow, patient prioritization and support decision making.

On the other hand, this reliable amount of data can help researchers develop a better computer aided diagnosis (CAD) system to filter obvious normal images and allow radiologists to concentrate themselves on the hard case, saving the doctor time and avoiding misdiagnosis. In [22] the largest publicly available data set had published ChestX-14, with 112,000 frontal x-ray images and a label on 14 different lung diseases and one no-finding (i.e., Atelectasis, Cardiomegaly, Effusion, Infiltration, Mass, Nodule, no-finding, Pneumonia, Pneumothorax, Consolidation, Edema, Emphysema, Fibrosis, Pleural Thickening, and Hernia). This huge dataset had provided opportunities to develop better CAD system that can greatly help the radiologist. Thus, because of the huge volume of data to be examined and the less accuracy of x-ray images

chest disease screening puts big pressure and burden on radiologists. Detecting pneumonia in chest radiography can be difficult for radiologists. In X-ray images, the appearance of pneumonia is frequently ambiguous, can overlap with other diagnosis, and can resemble a variety of other benign abnormalities. These inconsistencies result in significant variation in the diagnosis of pneumonia among radiologists [23]. Therefore, a fully-automated diagnosis system and consistent techniques needed to simultaneously reading a Chest X-Ray and diagnosis.

We employ Deep Convolutional Neural Network (DCNN) architecture for classification, different pooling algorithms, and various multi-label CNN losses due to the large image capacity. Deep learning consists of deep neural networks with a wide range of hidden layers that are used very effectively to computer vision problems in both classification and object detection especially where large set of benchmark training data are available.

Doctors and radiologists still use manual and visual methods to diagnose chest radiographs. As a result, there is a need for an autonomous system capable of diagnosing the chest X-ray [24].

So typically observed problems by this researcher are:

- Shallow network (like SVM and Logistic regression [17]) techniques get lower features of the chest x-ray images which misclassification of each thoracic diseases into correct categories of the diseases. Due to incapability of traditional methods extracting of complex feature of medical images.
- Analyzing large amount of thoracic x-ray images is a vast load for radiologists and causes them make an error. In general, manual diagnosis radiologists lack to consider knowledge about characteristics of each thorax diseases and they made biasness for some specific reasons, very time consuming, tedious process, not accurate and costly, Reliability of the manual classification depends on the knowledge and skill of the radiologists.
- Existing DNN works focus on binary classification of thorax diseases as normal and abnormal in [24] which is difficult to classify each patient to the pre-defined diseases classes for multi feature diseases.

For these reasons, the aim of this research is to design an autonomous system capable of diagnosing the chest X-ray that can withstand all of the limitations listed above, and we attempted to address the following research questions (RQs).

1. Does the Inception-v3 model we proposed better than the baseline approach of weakly-supervised classification and localization of common thorax diseases?
2. What is the impact of lung segmentation to enhance the computational speed of our system?
3. Do the preprocessing techniques enhance the classification performance of our approach?

1.3 Objectives

1.3.1 General Objective

The overall goal of this study is to design a deep neural network (DNN) for thoracic disease classification and location identification.

1.3.2 Specific Objectives

- To select the preprocessing technique and conduct effective pre-processing of data.
- To compare the impact of data pre-processing on the classification and location identification performance.
- To select and examining lung segmentation method.
- To evaluate the accuracy of classification and localization method.

1.4 Significance of the Study

Chest X-ray is one of the most popular examinations for identifying heart and lung diseases. Based on chest X-ray images many automated diagnosis algorithms have been proposed as a result of the vast number of clinical cases that exist. To the best of our knowledge, nearly no previous auto-diagnosis algorithms take into account the effect of relative location information on disease incidence. In this study, we propose to employ relative location information to assist the identification of thorax diseases.

In Ethiopia, as in [20] there are several lungs cancer and heart diseases victims , and Since the diagnosis provided by radiologists is not yet facilitated with CAD techniques, our effort encouraged to create automated systems using publicly available datasets.

- It delivers an investigate outcome for chest or thorax diseases categorization and localization for investigators in the building of thoracic diseases classification methods using X-ray images.

- The result of this study assists in reducing false positive rates in thoracic diseases classification and localization systems.
- The study helps in identifying chest diseases and find the exact location of the diseases from thoracic X-ray images and it also encourages investigators to do further investigations on thoracic diseases classification and localization.
- Some significance of this paper is giving insight for different applications areas of medical image processing like: for classification of pathological regions, to detect the level of cancer, level of diabetes, for classification of anatomical regions, to classify different bones.

1.5 Contribution of the Thesis

In researches done before for thoracic diseases classification and localization systems, different researches done only on thoracic diseases identification as normal or abnormal (thorax with disease) as binary classification but in this thesis work, we try to classify and locate fourteen different types of thoracic diseases like (atelectasis, cardiomegaly, consolidation, edema, effusion emphysema, fibrosis, hernia, infiltration, mass, nodule, pleural thickening, pneumonia and pneumothorax), this paper focus on multi-label classification and location of thoracic diseases. The other contribution of this study is conducting different preprocessing methods to get a better classification and localization performance and conduct segmentation to speed up the tuning of the model. As various papers [1, 6] suggest feature extraction of the input data using deep learning from chest X-ray section enhances chest diseases classification, this improves the overall performance of the classification and localization processes.
The implementation of the DCNN algorithm for the classification and location detection of chest diseases is done with the applicability of image processing techniques.

This paper extends the field of classification and localization of thoracic diseases from X-ray images with the following contributions:

- Classification of different pathological abnormalities: to classify (multi-label classification) fourteen different types of thoracic diseases like (atelectasis, cardiomegaly, consolidation, edema, effusion emphysema, fibrosis, hernia, infiltration, mass, nodule, pleural thickening, pneumonia and pneumothorax)
- Localization of pathological region: After obtaining a trained model, heatmaps are generated on test images that point out the exact location of the thoracic diseases. We calculate the

heatmap and generate the B-Box for pathological candidates by using activations from the gradient of the transition layer and weight from the prediction layer.

- Improve performance using different image preprocessing techniques.

1.6 Challenges and Scope of the Thesis

1.6.1 Scope of the Thesis

This thesis ultimate goal is to design and implement a DCNN model to classify and localize the type of the thoracic diseases which helps to find categories and location of the thorax disease from thorax X-ray images.

Generally, this thesis could be far from images format conversion like gray scale to color conversion and vice versa, preprocessing, prepare data, loading and randomly transforming images, design CNN model, train CNN, classify each pathology into its predefined classes and indicate the approximate location of each thorax diseases.

1.6.2 Challenges of the Study

we can generalize the challenges we face with medical data in this thesis work into two broad categories: the first challenge we face during this thesis work is on the labeling accuracy (loosely labeled) of the medical dataset. Because of the dataset is the text-mind disease image labels (each image might have several labels), from the corresponding radiological reports extracted by using natural language processing. We also show that these common thorax diseases can be recognized and located using a multi-label disease classification and localization framework. In this challenge category we have seen the visual labeling accuracy and the meaning of the labels medically. So, we found that labeling accuracy of the medical data is questionable. Working with all the datasets found in national institute of health (NIHCC) and picture archiving and communication system (PACS) is the second challenge, because the size of the data (about 45.7 GB, 112,120 frontal-view X-ray images of 30,805 unique patients), handling the data and carrying out the training process require an extensive amount of processing power.

1.7 Methodology

We used the steps bellow in order to carry out and implement our research.

1.7.1 Data Collection

The methodology employed in this thesis consists of collecting image of thoracic X-ray images from national institute of health clinical center (NIHCC), loading and preprocessing of thoracic X-ray images, model design for segmentation and classification. The following are list of the key steps we performed.

1.7.2 Design and Implementation

Literature Review - studying journals, articles, research papers and other relevant information assists in the process of choosing effective algorithms.

Data collection - Collecting medical image data from national institute of health clinical center (NIHCC) and picture archiving and communication systems (PACs) and extracting altogether for training and testing.

Preprocessing – is the process of creating the right input data for the segmentation stage, such as equalizing the histogram, resizing images, filtering out noise and histogram equalization.

Model design – Create the deep neural network that will be utilized, choose the training procedure, then use the deep neural network to the classification and localization process.

Implementation tools– This study have been conducted on the following hard ware and software resources:

Hardware specification: the algorithm run on Intel® core™ i5-6300HQ CPU@2.30, 2.30 GHz processor, GPU: Nvidia GeForce GTX 950M 4 GB dedicated graphics.

Software: cuda-ubuntu1604-amd64.deb, python 3.6(Anaconda3 3.6): Tensor flow-GPU, SimpleITK, NumPy, Pandas, Matplotlib with an editor Jupyter Notebook.

Result evaluation and conclusion – The evaluation is performed by computing the accuracy and loss function. we evaluate the final result by computing training-loss vs epochs, training-accuracy vs epochs, validation-loss vs epochs, validation-accuracy vs epochs using Categorical cross-entropy classifier.

1.8 Thesis Organization

There are five chapters in this thesis. Chapter one comprises the background information, problem description, objective, contributions, study scope and limitations. Chapter two includes back ground of x-ray images and a detailed understanding of thoracic diseases classification and localization from x-ray images using machine learning. Chapter three is about literature review on diagnosis of chest diseases from thorax x-ray images using different classical and state of the art deep neural network systems which has been done abroad researchers. Chapter four is about the design and implementation of the paper. Chapter five presented about data setup, experiments and result discussion. The last chapter of the study, chapter six, delivers the conclusion of this thesis work and future directions.

Chapter 2

Theoretical Background

2.1 Introduction

Now a day, thorax diseases is a major health issue in the world as whole. Around 450 million people are affected by pneumonia alone (that is, 7% of the world's population) and results about 4 million deaths are reported every year [25]. Due to its low-cost and easy-access nature, chest radiography, colloquially called chest X-ray (CXR), is one of the most common types of radiology examinations for the diagnosis of thorax diseases [25]. The vast number of global chest x-rays are actually almost completely examined by means of slice-by-slice visual inspection. This requires a high degree of expertise and dedication and is time-consuming, costly, operator-friendly, and unable to use the immeasurable informatics of such large-scale data. In addition, since chest x-ray images are complex, even radiologists are challenged to prejudice against thorax conditions and are thus not able to interpret chest x-rays by specialist radiologists in many countries. It is therefore important to develop automated algorithms for the computer-aided diagnosis of chest X-ray diseases.

During the last few years, in many computer vision applications, in particular classification of natural and medical images, profound advances in learning techniques have been achieved [26]-[27]. This performance has led several researchers to use deep convolutional neural networks (DCNN) to diagnose chest X-ray thorax diseases. Wang et al. in [22] presented a unified, weakly supervised multi-label classification system by evaluating various multi-label DCNN losses and different pooling strategies. The deep learning model called Chex Net was developed by Rajpurkar et. al.[23] to make the optimization of such models manageable by using dense connections to normalize batch models. According to the results published, Chex Net is capable of detecting pneumonia at a level that matches or exceeds the level of radiologists.

Thorax diseases can intuitively be identified with computer assistance in two consecutive steps: identification of pathologic abnormalities and classification of them.

Intuitively, computer-aided diagnosis of thorax diseases consists of two successive steps: the detection of pathological abnormalities and classification of them. Due to the complexity and variety of thorax and poor chest x-ray quality, automatic abnormality detection in chest x-rays is

also another challenge. Manually marking the counters of irregular regions on the chest X-ray requires much more effort than labeling. As a result, few public chest X-ray data are fitted with an abnormality mask [22], leading this computer-aided diagnostic activity to a weakly supervised problem (i.e., providing only the names of the abnormalities in each radiograph but not the location of the abnormalities.

2.2 Chest X-Ray Overview

Radiography is an imaging technique that used to display the inner structure of the object within the human body by using x-rays, gamma rays and other types of ionizing and non-ionizing radiation. To create images of the body's internal structures, radiography or x-rays use a very small dosage of ionizing radiation. X-rays are the oldest and most widely used type of medical imaging. They are also utilized to assist in the diagnosis of broken bones, the search for injury or infection, and the detection of foreign objects in soft tissue. Some x-ray tests may include barium or iodine-based contrast material to enhance visibility of specific organs, blood vessels, tissues, or bones.

Radiographic imaging is utilized in a vast variety of usages such as medicine, engineering, forensics, and security. Radiographic examination is a means of checking materials for hidden defects by the penetration of things with the power of electromagnetic shortwave radiation (high energy photons). In the radiography examination, the sample portion is placed between the radiation source and the film (or detector). Material density and thickness changes in the test portion will diminish penetrating radiation through interaction mechanisms such as scattering and absorption. The absorption variations are then recorded on film(s) or electronically.

2.2.1 Formation of an X-Ray Picture

There are two fundamental methods for producing pictures with x-radiation. Passing an x-ray beam through the body portion and projecting a shadow image onto the receptor is one way. The second method, used in computed tomography (CT), employs a digital computer to calculate (reconstruct) an image from x-ray penetration data obtained by scanning a relatively thin beam over the patient's body. At this point, we are simply concerned with projection imaging, which is the fundamental method used in radiography and fluoroscopy. In addition to the x-ray penetration qualities, scattered radiation (Scattered Radiation and Contrast) and the contrast

qualities of the receptor and display system have a considerable impact on image contrast. (Film Contrast Properties). Image blurring also reduces the contrast of small things within the body and anatomical detail (Blur, Resolution, and Visibility of Detail and Radiographic Detail). As the x-ray beam emerges from the patient's body it contains an image in the form of variations in exposure across the image area. This is formed by variations in the attenuation through different parts of the body. The amount of variation in x-ray exposure between points within the image demonstrates contrast in this attenuation image; the amount of contrast produced in a specific examination is determined by both the physical characteristics of the body section and the penetrating characteristics of the x-ray beam. We explore the characteristics of both the objects within a body and the x-ray beam and show how optimum image contrast can be achieved.

2.2.2 What Exactly is a CSV Image File?

A CSV file is a plain text file that contains a list of data separated by commas. These files are frequently used to exchange data between applications. For instance, CSV files are often supported by databases and contact administrators. These files are also known as Character Separated Values or Comma Delimited files. They usually utilize the comma sign to split (or delimit) data, but they sometimes use other characters, such as semicolons. The notion is that you can export complicated data from one form to a CSV file, then import that data into another application. Since the finding labels in national institute of health clinical center, the database found as CSV format we need to convert to the appropriate format in order to be readable by the training algorithm.

2.3 Thorax Abnormalities (Diseases)

Table 2. 1 Summary of fourteen common thorax pathologies [28]

Disease	Description
Atelectasis	A complete or partial collapse of the lung, the disease can be seen in CXR at which the lung is missing completely or partially.
Cardiomegaly	Enlargement of the heart. Cardio-thoracic ratio, a useful measurement on CXR, can be used to diagnose cardiomegaly.
Pleural Effusion	An unusual amount of fluid around the lung, the disease causes the lung to appear white in CXR.

Infiltration	Abnormal substances accumulate gradually that occurs within or spreads as through the interstices of the lung.
Mass	Abnormal spot or area in the lungs that is more than 3 cm in size in CXR.
Nodule	Abnormal spot or area in the lungs that is less than 3 cm in size in CXR.
Pneumonia	Inflammatory condition of the lung cause by bacterial, fungi, etc.
Pneumothorax	Abnormal collection of air in the in the pleural space.
Consolidation	Consolidation An induration of a normally aerated lung (the swelling or hardening of the soft tissue). A CXR frequently displays a white lung area.
Edema	Fluid accumulation in the tissue and air space of the lung, CXR will show fluid in the alveolar walls.
Emphysema	A kind of chronic pulmonary obstruction (COPD). The breathing tubes in emphysema are reduced and the air bags are destroyed, which cause respiratory shortages.
Pulmonary fibrosis	Excess of fibrotic tissue in the lung, lead to scars is formed in the lung tissues.
Pleural Thickening	Extensive scarring thickens the lining of the pleural. Spotted in CXR as an irregular shadow on the pleura that extends over at least 25 percent of the chest wall.
Hernia	Protrusion of lung outside of the thoracic wall.

2.4 Jpeg Image Data

The initial stage of our work is to gather data from different source of images (or databases) to be input to the system. Direct digital JPEG capture is the preferred method for collecting images from technologies that are intrinsically digital, such CT, MRI, US, and x-rays. The process of gathering thorax x-ray images from NIHCC data sources is known as JPEG image collection for the system of categorization and localization of thoracic or chest diseases. Thus, gathering JPEG images is the first step towards addressing this problem. In a study published in [22], the researchers tested their machine learning models using data from private hospitals.

2.5 Image Processing

Image processing is a way of performing certain operations in an image to obtain or extract relevant information from the enhanced image. The goal of the digital image processing (DIP) stage is to make the data suitable so that the learning machine classifier can utilize it. Image processing is the way to manipulate images in several techniques to easily visualize and detect images.

Digital images can offer a lot of detail. In addition, a computer model may find the images very challenging to process in a reasonable amount of time due to their high complexity. Thus, in order to improve the complexity, quality, simplicity of manipulation, and decrease false positive objects outside of thoracic diseases, we used image processing techniques instead of using a single image due to their complexity, dimensionality, and file format. To identify features that are shared by the entire image rather than pixel-by-pixel, it is also helpful to identify spatial connections within the image. In our proposed design, image processing is intended to lower false positives and enhance the quality of images extracted from image files that assist us in identifying suspicious objects. It also decreases the file size and helps to get precise region of interest of abnormalities. Furthermore, it prepares the input data so that machine learning algorithms can train on it. Because the proposed DCNN method only functions with numerically valued data that comes within the same range of zero and one. Hence, the dataset needs to be properly prepared to make it suitable for the system before it is provided as input to the DCNN algorithm. As a result, we employ various image processing methods, including image preprocessing and segmentation.

2.5.1 Image Preprocessing (Image Enhancement)

Pre-processing is a critical stage in any image processing or feature extraction task, such as orientation, labeling, artifact removal, enhancement, and segmentation. The preprocessing involved in color conversion, image resize, noise removing and enhances the quality and produces an image in which lesion or abnormality can be detected correctly. The main goal of the preprocessing stage is to minimize noise, undesired artifacts, and image distortions. It is utilized on images at the least level of visuals. The primary aim of preprocessing is to improve image observation, which raises the likelihood of effective abnormality visualization. Thus, to

improve the image quality, the approach we propose applies median filter, data normalization, and histogram equalization [8, 29]. Unwanted areas in the x-ray images include contrast, noise, and high frequency components.

2.5.2 Gray Scale Conversion

Grayscale representations are preferable over colored representations because they simplify the process and reduce computational costs. Indeed, color may be of limited utility in many applications, and providing extraneous information may increase the quantity of training data required to attain high performance. For example, assume a simple de-noising application of a color image. It will need to de-noise in each channel, and this appears to be a duplicate process to be simplified and it can convert the given image to gray scale and deal with that.

2.5.3 Image Resizing

Image resizing is an important role in image processing technique, to enlarge and reduce the given image size in pixel format. Image interpolation in to two ways image down sampling and up-sampling is necessary when resizing the data to match with either the essentials of the output display. However, it is more capable to transfer low-resolution forms to the client, a guesstimate of the high-resolution of the original may be necessary when donating the final visual data. Thus, the accurate resizing of image is a necessary step in many methods, stretching from several applications for critical functions in the field of medical, defense sectors and etc. When an image is scaled, the pixel information is altered, and any unnecessary pixel information is removed. When an image is expanded, new pixel information is added. Large images take up a lot of memory, which might cause our computers to get overloaded and uploads to be extremely slow. The memory cost of an image is calculated based on its size. So, correctly resizing an image reduce memory cost and image uploads become fast by making appropriate image dimensionality which is suitable to the neural network.

2.5.4 Normalizing Input Images

The normalization of data is an important step that ensures that a consistent distribution is accessible to each input parameter (pixels in this thesis)[8, 29]. These speeds up convergence when training the network. The normalization of the data is carried out by the subtraction of each

pixel of the mean and by the division of the output by standard deviation. For input images we need pixel numbers to be accurate, we can choose to scale the normalized data in the range [0, 1] or [0, 255]. For our dataset we used the range of [0, 1]. Data normalization can be calculated using equation (2.1)

$$normalization = \frac{(x - xmin)}{(xmax - xmin)} \tag{2.1}$$

where Xmax and Xmin are the maximal and minimal values respectively[29].

2.5.5 Histogram Equalization

Histogram equalization is a method used in image processing for contrast correction using the histogram of the image [8]. This method increases the overall image contrast, particularly when the available image data is defined by close contrast values. This change improves the distribution of intensities on the histogram. The equation (2.2) can be used to compute histogram equalization. As our data are greatly affected by variations in intensity and unequal illumination, we use histogram equalization in x-ray images to enhance views and detail. This technique is used to generate JPEG images of high quality which are readily accessible for the next part. It adjusts the contrast and improves the view of the internal structures in x-ray, CT, MRI and US images, as applied in many different researches works [30]. The main benefit of using this approach is that it is very simple and invertible techniques. The equalized histogram can be calculated with the cumulative distribution function (CDF) formula S_k as follows:

$$sk = \sum_{j=0}^{k} \frac{nj}{n} = \sum_{j=0}^{k} pr(rj) \tag{2.2}$$

Where, $0<r_k<1$ is the normalized gray level and K=0,1,2…...L-1 L is the gray level number[31].

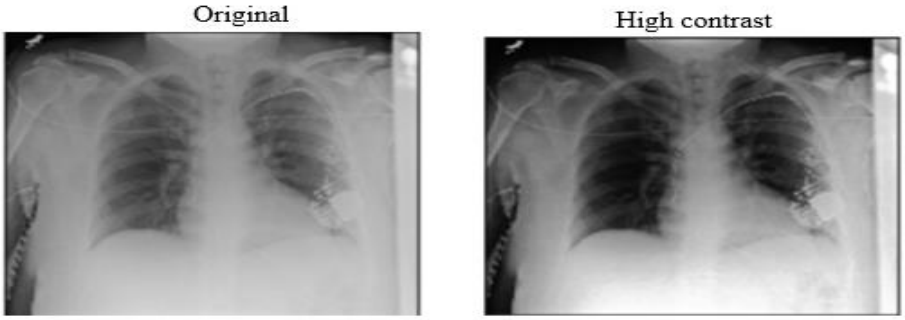


Figure 2. 1 Image processing with histogram equalization

2.5.6 Noise Removal Using Median Filter

The random variable in brightness or color information identified in medical equipment or scanners images is defined as image noise. Image noise is commonly recognized as an unwanted effect of image capture. Noise is often defined as the uncertainty in the signal due to random fluctuations in signals. There are many causes for these fluctuations. All medical images contain some visual noise. Image denoising is a critical aspect in medical image processing, in which the original images are poor due to the entering of noises and artifacts in the images acquisition systems. In the image processing filters, filtering techniques for image adjustment or enhancement are primarily employed to remove the high frequencies in the image, i.e. the smoothing of the image, or low frequencies, i.e. improving or detecting the image edges. Median filter is a nonlinear operation used to reduce artifacts, salt and pepper noises [8]. The image's noise increases as sharpening and shadows are carried out by means of a high frequency signal in the images. Noise filtering using median filters is easier to eliminate noise and retain digital images' edges and fine information of images. In [8, 29] there are a number of different filters available, such as low pas, high pass, mean, median, Weiner, gaussian etc. For our way of removing some noises, we have used median filter technique, since a median filter it retains sharp edges of images and fine details of digital images in the X - ray under certain conditions while removing noise. The median filter value is calculated using equation (2.3).

$$\mathbf{y(m,n) = median[x[i,j], (i,j) \in \omega} \tag{2.3}$$

Where ω represents a neighborhood defined by the user center around location (m, n).

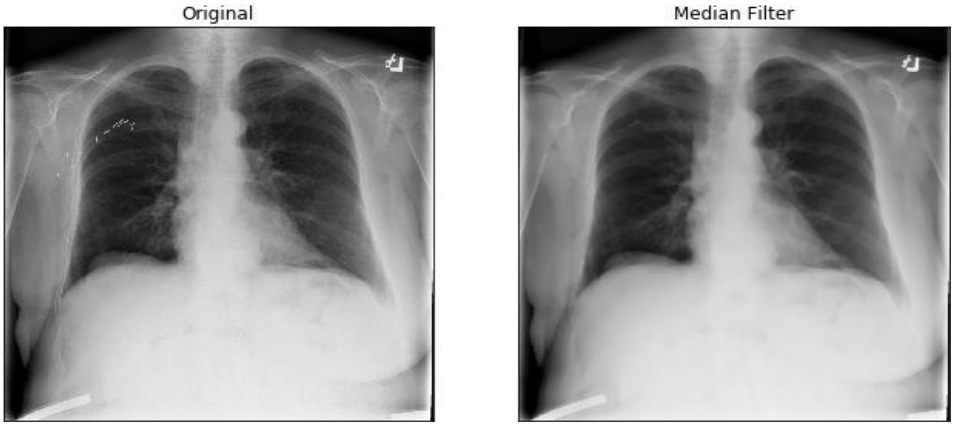


Figure 2. 2 Image processing with median filter

2.6 Lung Segmentation

Semantics segmentation aims to label every pixel in an image. This is more difficult than classification, in which the classifier learns simply "WHAT" object is in the image. Segmentation requires the model not only learns about “WHAT” object but also learn about “WHERE” the object is. In this study, our task was to segment the lung region from the CXR image. Given a CXR, we have to output a binary mask where lung pixels are 1 and other pixels are 0.

We tackled this by using U-net [34] architecture. The network composes of two paths: the down-sampling path and the up-sampling path, or also known as encoder and decoder. In the encoder stage, we used ResNet-50 as a backbone to extract features from the image. Then from these latent features, the decoder path will learn to reconstruct the high-resolution binary mask.

We use transposed convolution (also known as up convolution) to learn the up-sampling path. This approach allows the model to learn a more flexible up-sampling transformation compare to other fixed strategy methods. However, the down-sampling method employs the max pool layer, which discards the spatial information from the features map, we additionally include a short-cut link inspired by U-net. After each transpose convolution, we will concatenate the feature map from the equivalent down-sampling path that brings in spatial information to help in the reconstruction of the fine-grained segmentation.

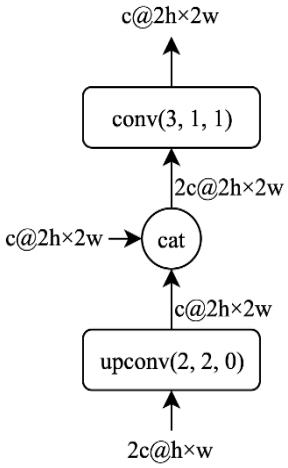


Figure 2.3: **Up Layer**

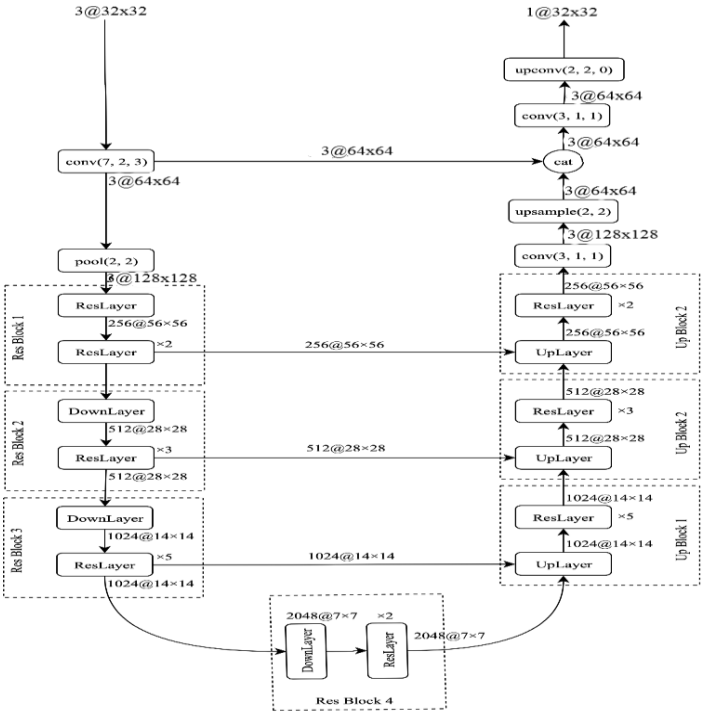


Figure 2.4: U-net with ResNet-50

The down sampling path is identical to ResNet-50 that will encode the data. The up-sampling path will learn to decode the features map in order to re-create the input resolution with skip. We formalize the segmentation problems as pixel-wise classification where each pixel is label as lung region or not. Let $y \in R^{W \times H}$, $y_{i,j} = 1$ indicates lung region and 0 otherwise, let $p \in R^{W \times H}$ which $p_{i,j}$ indicates the probability of pixel $y_{i,j}$ is lung region. W, H is the width and height of the input image and ground truth mask.

Using the above setting, the loss function is equal to the total cross entropy loss for every pixel in chest x-ray.

$$\mathrm{L(p,y)} = \sum_{\mathrm{i=0,j=0}}^{\mathrm{i\in H, j\in W}} -\mathrm{y_{i,j}} \ln(\mathrm{p}) + \sum_{\mathrm{i=0,j=0}}^{\mathrm{i\in H, j\in W}} (1-\mathrm{y_{i,j}}) \ln(1-\mathrm{p}) \tag{2.7}$$

2.7 Machine Learning for Thorax Disease's Classification

Machine learning (ML) is the methodical utilization of algorithms and mathematical models by computer systems to execute a certain activity effectively without the use of human involvement, instead relying on patterns and conclusions. It's a subfield in artificial intelligence. Machine

learning algorithms are a branch of science that tries to teach computers to function without the need for particular programming in order to learn more about solutions to various problems. The most important feature of machine learnings is that the process of decision is guided by information or data, with minimal human interference.

Machine learning algorithms develop a mathematical model sample data called "training data", to produce predictions or conclusions without explicit programming for the purpose task [35]. The system learns and predicts when new data is entered by evaluating the training samples. Machine learning and artificial intelligence have made significant advances in the medical field in areas such as medical image processing, computer-aided diagnosis, image interpretation, image fusion, image registration, image segmentation and classification, image-guided therapy, image retrieval and analysis in recent years. In machine learning there are three different types of algorithms [35]:

1. Supervised learning: The model has prior knowledge of what the output value is from the input and the expected output data and the model make the best approximates based on the input and output data relationship. This method is the most common learning method. It utilizes labeled training data to learn how to map from the input variable to the output variables. Examples: multi-layer perceptron (MLP) Neural Networks, Logistic Regression, SVM, K-Nearest neighbors, Linear Regression, Decision Tree etc.
2. Un supervised or Unmonitored learning: Only data input is used to learn the model. This method is particularly useful in practice because there is plenty of unlabeled data while labeled data is scarce and requires a great effort to label it. This method provides the algorithm input data and learns and predicts from experience, mostly through association, clustering, and reduction of dimensionality. The algorithm mostly learns the similarities between the input data to rebuild it again. Examples: K-Means clustering, hierarchical clustering, mixture models, etc.
3. Semi-supervised learning: in this approach both types of data are used to train the model. First, the model is pre-trained using unmonitored or unsupervised data and then enhanced with supervised data. When a neural network is used for classification; it first pre-trained layer by layer using unsupervised training algorithm. Finally, the network can be trained for classification or prediction with a standard training algorithm.

2.7.1 Artificial Neural Network

A neural network consists of interconnected artificial neurons that exchange messages. The way of interconnection distinguishes network models such as feed forward networks, recurrent networks, self-organizing networks, and so on [36].

A processing element (neuron) is known as a unit or a node in neural networks. A neuronal model has a set of connecting weights (corresponding to synapses in biological neurons), a summing unit, and an activation function, as shown in figure 2.5. The output y of the summing unit is a weighted sum of the input signals (features):

$$y = \sum_{i=1}^{n} \boldsymbol{wixi} + \mathbf{b} \tag{2.4}$$

where w_i are the weight matrix of each connection between input nodes, x_i are input matrix with hidden nodes and *b* is a bias of the node.

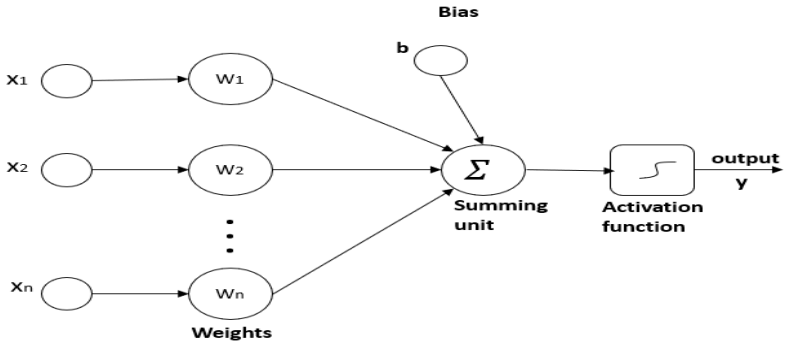


Figure 2. 5 A neuron model

Neural network architecture depends on the applications nature and complexity. However multilayer neural network with proper parameter selection is capable of classifying nearly any pattern. Between the input and output layers of multi-layer neural networks there are one or more hidden layers.

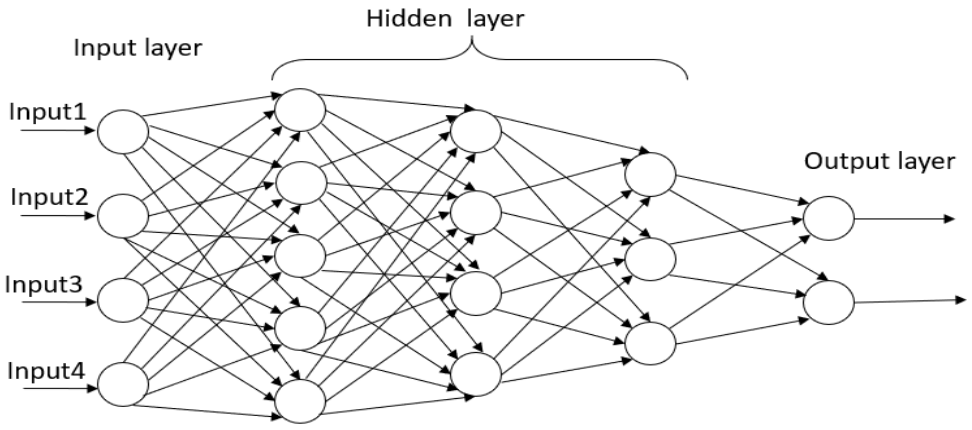


Figure 2. 6 Sample of multilayer neural networks architecture

ANN layers

As in figure 2.6 the neural network has an input layer (on the left) with four neurons, three hidden layers (in the middle) with five neurons, four neurons and three neurons respectively and an output layer (on the right) with two neurons.

Input layer: This is the first layer in the ANN neural system. This layer is important for transferring information or data to other neural system layers. The input layer is provided with a vector of predictor variable values (x1… x_p). The input layer (or processing preceding the input layer) specifies this to a -1 to 1 range for each value. The values are distributed to each of the neurons in the hidden layer by the input layer.

Hidden layer: the layer that exists between the input and output layers. When a neuron enters the hidden layer, its value is multiplied by a weight (wji), and the weighted values are summed together to produce a cumulative value uj. The weighted sum (uj) is passed into the transfer function, which returns hj. The hidden layer's outputs are distributed to the output layer.

Output layer: the final layer of the neural network that get information processed from the hidden layer. Each hidden layer neuron's value is multiplied by a weight (wkj), and the weighted values are combined together to give a cumulative value (vj). The weighted sum (vj) is passed via a transfer function, which returns the value yk. The values of y are the neural network's outputs.

2.7.2 Deep Neural Network

Deep learning refers to a subfield of machine learning inspired by the deep structure of a human brain [37, 38], that is based on learning levels of representations, corresponding to a hierarchy of features, factors or concepts, where higher-lever concepts are defined from lower-lever ones and the same lower-lever concepts can help to define many higher-level concepts. It is the growing trend towards the development of automated applications and was named one of the breakthrough technologies in 2012 [32]. Computerized vision tools based on deep learning, like CNN, RNN, DBN, DBM, and SAE, currently perform better than humans in a variety of tasks. It is an enhancement of artificial neural networks that consists of more hidden layers that allow for higher representation levels and enhanced image analysis. Due to its recent incredible results, it is widely used for various functions including imaging for medicine, recognition of speech, facial recognition, and object identification. Deep neural networks are capable of practically learning all potential mappings after a successful training process using a significant collection of data, and they can also predict intelligently, distinguishing between doubtful and non-dubious things from images of previously unknown data. Deep learning neural network techniques are divided into various categories.

2.7.2.1 Convolutional Neural Network

Convolutional neural network is one of the most important frameworks of deep learning technique and has been widely employed in feature extraction, segmentation, classification, recognition and other applications.

Deep structures include several hidden layers that allow different levels of features to be generalized. In 2006 [37] build a novel algorithm called greedy layer wise training in order to train the deep architecture neuron layers. This learning algorithm is viewed greedily as an unsupervised single-layer training in which layer by layer a deep network is trained. Because this approach has become more prominent, in practice it has started to use a lot of deep networks. The convolutional neural network is one of the most effective deep neural networks, which can include several hidden layers that performing convolution and subsampling to extract features from low to high input data [37] .This network has demonstrated great efficiency in various areas, notably, in computer vision, biological computing, fingerprint enhancement, and so on [37]. Mainly, this convolutional neural network consists of three basic layers: convolution layers,

subsampling or pooling layers, and full connection or dense layers. Figure 4 shows a typical neural network (CNN) architecture. Each type of layer is briefly explained in the following section.

Input Layer

The input layer does read the image, so there is no parameter learn in this layer. This layer takes the raw image data as it is. An input layer inputs images to a network and applies data normalization. Specify the image size. The size of an image is determined by its height, width, and number of color channels. For a grayscale image, the number of channels is 1, and for a color image it is 3.

2.7.2.1.1 Convolution Layer

In this layer, as shown in Figure below, an input image of size R*C is convolved with a kernel (filter) of size a∗a [37, 38]. Every block of the input matrix is translated to the kernel independently and a pixel is generated in the output. The result of the transformation or the convolution of the input images and kernel is used to generate or produce n output image features. A kernel of the convolution matrix is generally referred to as a filter, whereas the features of the output image obtained by the transforming kernel and the input images are referred to as the feature maps of size i∗i. Convolutional neural networks can include numerous layers of convolution; the feature vector is made up of the inputs and outputs of the subsequent convolution layers. Each convolution layer is made up of n filters. These filters are convolved or translated with the input, and the depth of the resulting feature maps (n) is proportional to the number of filters used in the convolution process. Remember that at a certain position of the input image, every filter map is considered as a specific feature [37]. The l-th convolution layer output, denoted as $C_j^{(l)}$, consists of feature maps. It's calculated as

$$\mathrm{Ci(l)} = \mathrm{Bi(l)} + \sum_{j=1}^{a_i^{(L-1)}} k_{i,j}^{(L-1)} * \mathrm{cj(l)} \tag{2.5}$$

where $B_i^{(l)}$ is the bias matrix and $k_{i,j}^{(L-1)}$ is convolution filter or kernel of size a∗a that connects the j-th feature map in layer (l - 1) with the i-th feature map in the same layer. The output $C_i^{(l)}$

layer consists of feature maps. In (2.9), the first convolutional layer $C_i^{(l-1)}$ is input space, that is, $C_i^{(0)} = Xi$.

The kernel generates feature map. After the convolutional layer, the activation function can be applied for nonlinear transformation of the outputs of the convolutional layer:

$$Yi(l) = Y(Ci(i)) \tag{2.6}$$

where $Y_i^{(l)}$ is the output of the activation function and $C_i^{(l)}$ is the input that it receives.

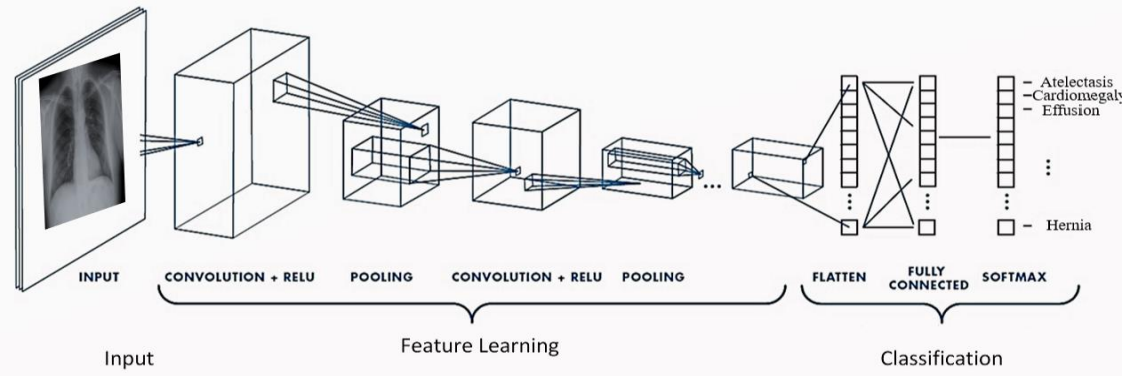


Figure 2. 7 Architecture of convolutional neural network

Typically used activation functions are sigmoid, tanh, and rectified linear units (ReLUs). In this paper, ReLUs which is denoted as $Y_i^{(l)} = \max(0, Y_i^{(l)})$ are used. This function is widely used in deep learning models Because of its help in reducing interaction and nonlinear effects. If it receives a negative input, ReLU turns the output to 0, while returning the same input value if it receives a positive input. The advantage of this activation function over other functions is that it allows for faster training due to the error derivative, which becomes very small in the saturating area, and therefore weight updates nearly eliminate. This is called the problem of the vanishing gradient.

Batch Normalization Layer

Each input channel is normalized by a batch normalization layer in a mini-batch. It was used to accelerate convolutional neural network training and decreasing sensitivity to network initialization.

Batch normalization layers are used between convolutional layers and ReLU layers. The layer begins by normalizing each channel's activations by subtracting the mini-batch mean and dividing by the mini-batch standard deviation. This layer normalizes or regularize activations

and gradients that propagate through a neural network, making training network an easier problem of optimization. Easier optimization enables the parameter updates can be larger and the network can learn faster.

Feature Maps

The same set of weights and bias are utilized for convolution as the filter proceeds along the source or input, resulting in a feature map. Each feature map is the output of a convolution with a unique weight set and bias. The number of feature maps therefore equals to the number of filters. The total number of parameters in the convolutional layer is ((h*w*c+ 1) *number of filters), where 1 represents the bias, h represents the height, w represents the width of the filter, and c represents the number of channels in the input.

Padding Size

Vertically and horizontally, you may also use the 'Padding' name-value pair option to apply zero padding onto the input image boundaries. Padding is the columns or rows of zeros added to an image input boundary. By adjusting the padding, we may control the output size of the layer.

2.7.2.1.2 Subsampling Layer (Pooling Layer)

The main objective of this layer is to minimize spatially the dimensionality of the feature maps extracted from the previous layer of convolution. There is no parameter learn in this layer. A max pooling layer down samples the input by splitting it into rectangular pooling areas and finding the maximum of each. To do this, a mask of size b∗b is chosen as shown in figure 2.8 and the sub-sampling process is performed between the mask and the feature maps. There are methods of subsampling such as averaging pooling, sum pooling, and maximum pooling have been suggested. Max pooling is the most commonly used pooling, where the maximum value of each block is the corresponding pixel value of the output image. The sample is done by dividing the input into rectangular pooling areas by average pooling layer and by calculating the mean values for each region. In the step size given in the 'Stride' argument name-value, pooling lays scan the input horizontally and vertically. The pooling areas do not overlap if the pool size is less than or equal to the stride. If there is n-by-n input into the pooling layer and the pooling region is h-by-h, the pooling layer will sample sections h for non-overlapping areas (Pool Size and Stride are equal). This means that the output of max or average pooling layer is n/h by n/h for one

channel of a convolutional layer [11]. The output of a pooling layer is (Input Size + 2* Padding) / Stride + 1, For overlapping areas. It is important to note that a subsampling layer assists the convolution layer in tolerating rotation and translation of the input images.

Dropout Layer

At the training time, the layer sets input elements to zero randomly by using dropout mask rand(size(X)) < Probability, where X is the input layer, and the remaining elements scaled by $\frac{1}{(1-probablity)}$. This process effectively changes the underlying structure of the network between iterations and helps prevent overfitting of the network [11]. During training, a higher number result in more elements being lost. At the moment of prediction, the layer's output equals its input.

Fully-Connected Layer

The convolutional neural network final layer is a traditional feedforward network with one or more hidden layers. A fully connected layer multiplies the input by a weight matrix $\boldsymbol{W}$ and then adds a bias vector $\boldsymbol{b}$. To identify larger patterns, this layer integrates all of the features (local information) learnt by the preceding layers throughout the image. The last fully connected layer in a classification issue combines the features to categorize the images. This is why the output-Size argument of the network's last fully connected layer equals the number of classes in the data set. During constructing the fully connected layer we can adjust the learning rate and regularization parameters by using associated name-value pair arguments.

Output Layers

In this layer SoftMax and classification layer are categorized under output layers. A SoftMax layer applies a SoftMax function to the input. For multi-class classification problems with mutually exclusive classes, a classification layer computes the cross-entropy loss. For classification issues, the last fully connected layer must be followed by a SoftMax layer and then a classification layer. The output layer uses SoftMax activation function:

$$y_{r(x)} = \frac{\exp\left(a_r(x)\right)}{\sum_{j=1}^{k} exp\left(a_j(x)\right)} \tag{2.7}$$

where $0 \leq y_r \leq 1$ and $\sum_{j=1}^{K} y\, j = 1$.

The SoftMax function is the output unit activation function after the last fully connected layer for multi-class classification problems:

$$P(cr|x,\theta) = \frac{P(x,cr|\theta)p(cr)}{\sum_{j=1}^{k} P(x,\theta|cj)P(cj)} = \frac{Exp(ar(x,\theta))}{\sum_{j=1}^{k} \exp\,(aj(x,\theta))} \tag{2.8}$$

Where $0 \leq p(c_r|x,\theta) \leq 1$ and $\sum_{j=1}^{k} p(c_j|x,\theta) = 1$. Moreover, $a_r = \ln\big(P(x,\theta|c_r)P(c_r)\big)$, $P(x,\theta|c_r)$ is the class prior probability.

The SoftMax function, also referred as the normalized exponential, is a multi-class generalization of the logistic sigmoid function. In most classification networks, the classification layer must come after the SoftMax layer. Train-Network in the classification layer takes the SoftMax function values and assigns each input to one of the K mutually exclusive classes using the cross-entropy function for a 1-of-K coding scheme [11].

$$\text{Loss} = \sum_{i=1}^{N}\sum_{j=1}^{K} \boldsymbol{t}\,\mathbf{ijlnij} \tag{2.9}$$

where N is the number of samples, K is the number of classes, t_{ij} is the indicator that the i^{th} sample belongs to the j^{th} class, and y_{ij} is the output for sample i for class j, which in this case, is the value from the SoftMax function. That is the probability that the network associates the i^{th} input with class j.

After finding output signals, the training of the CNN is started. Training is performed using the stochastic gradient descent algorithm. The gradients are estimated by the algorithm using a single randomly selected sample from the training set. The parameters of CNN are determined as a result of training. There have been so many parameters which control the performance of the neural network, such as:

- Number of layers in the model
- Number of epochs
- Training algorithm
- Number of neurons in each layer
- Regularization
- Optimization algorithms

2.7.3 Training Deep Neural Network

Forward Pass

Neural networks are effective tools for classifying patterns using training and testing data to build a model.

In feed forward neural network the parameter passes from input layer (edge detection) through hidden layer (complex parameter learning) to output layer(classification).

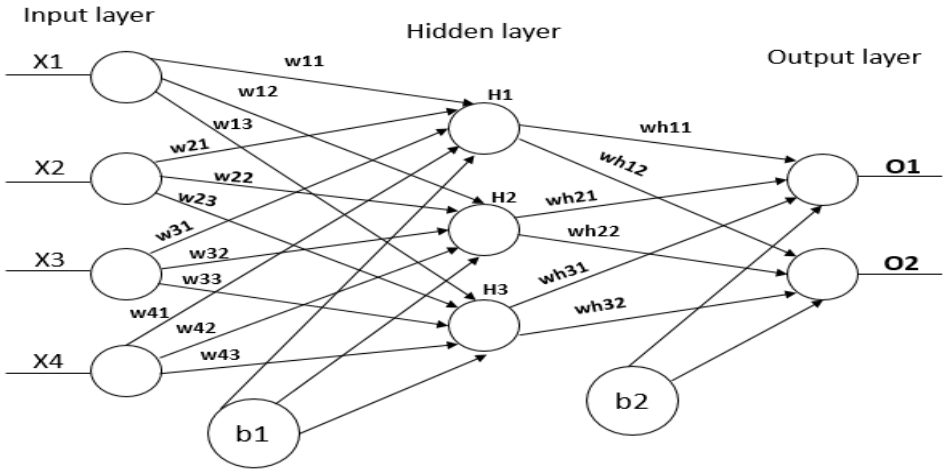


Figure 2. 8 multi-layer perceptron

Network setup:

Step 1: Initialize network parameter

The first step is to initialize the weights and biases using **rand ()** function.

Step 2: Compute net input at nodes in the hidden layer

Net Input is input multiplied by weight, then incremented by bias. Using matrix multiplication of input vector and weight vector, we get the resultant matrix using **sum-product ()** function.

Step 3: pass the net inputs through the activation function (sigmoid)

pass the output from Step 2 as input to the activation function at each neuron of hidden layer which can easily be done by $f(x) = \frac{1}{1+e^{(-x)}}$

step 4: calculate net input at output node

Now the outputs from Step 3 will act as inputs to the output node. Net input at output become the product of the output (from hidden layer) and initial weights plus bias at output.

Net input at output node = sumproduct ({output (hidden layer), {initial weights}) + bias at output

Step 5: obtain final output of neural network

pass the output received from Step 4 to the activation function as f(x), which again can be calculated using f(x)=1/(1+exp(-x)), thus resulting in the final output of neural network.

Back propagation

Backpropagation neural network (BPNN) is a multilayer feedforward neural network that uses a supervised learning algorithm known as error back-propagation algorithm. Errors at the output layer are transmitted back into the network for weight modification[35]. The algorithm starts with random weights, and the goal is to update each of the network's weights to minimizing the error for each output neuron and the network as a whole, so that the actual output is closer to the target output. Backpropagation algorithm requires a differentiable activation function, and the most popular ones are tan-sigmoid, log-sigmoid and, linear on rare occasions.

After obtaining the Feed-forward output, the next step is to evaluate the network output by comparing it to the desired outcome. The challenge here is to find the optimal weights and biases that can minimize the sum of square error (MSE):

$$E = \frac{1}{2}\sum(\text{Network output} - \text{Target output})^2 \tag{2.10}$$

Where E is error function received by the network. We need to evaluate the error produced by each of these weights and biases separately and then keep updating them to decrease the error. This process will be iterated till the convergence. The network will be called a trained network once an optimal is reached.

Step 1: updating weights at hidden [wh11, wh12, wh21, wh22, wh31, wh32]

Calculate derivative of error function **E** with respect to weights [wh11, wh12, wh21, wh22, wh31, wh32] using chain rule, which after simplification is equal to [Network output – target output] x [derivative of sigmoid at output node] x [$(\text{input})_j$ to output node] (For j = 1,2…).

were, Derivative of sigmoid function $f(x) = [f(x) * (1 - f(x))]$

then, we compute $\left[\frac{d(E)}{\mathrm{d(wH11)}}, \frac{d(E)}{\mathrm{d(wH12)}}, \frac{d(E)}{\mathrm{d(wH21)}}, \frac{d(E)}{\mathrm{d(wH22)}}, \frac{d(E)}{\mathrm{d(wH31)}}, \frac{d(E)}{\mathrm{d(wH32)}}\right]$

The new updated weights will be:

$$[\boldsymbol{initial\ weights}] - [\{(\boldsymbol{learning\ rate}) * \left[\frac{d(E)}{\mathrm{d(wH11)}}, \frac{d(E)}{\mathrm{d(wH12)}}, \frac{d(E)}{\mathrm{d(wH21)}}, \frac{d(E)}{\mathrm{d(wH22)}}, \frac{d(E)}{\mathrm{d(wH31)}}, \frac{d(E)}{\mathrm{d(wH32)}}\right]\}]$$

Step 2: update bias at output node

Calculate the derivative of error function E with respect to bias, which is equal to:

[network output – target output] x derivative of sigmoid at output node, $\frac{d(E)}{d(b1)}, \frac{d(E)}{d(b2)}$.

New updated bias [b1] new = [initial bias] – [learning rate x $\{\frac{d(E)}{d(b1)}\}$]

New updated bias [b2] new = [initial bias] – [learning rate x $\{\frac{d(E)}{d(b2)}\}$]

Step 3: update weights at input [w11, w12, w13, w21, w22,23, w31, w32, w33, w41, w42, w43]

Update weights at input to hidden layer, calculate derivative of error **E** with respect to weights [w11, w12, w13, w21, w22,23, w31, w32, w33, w41, w42, w43] after simplification it is equal to:

[Network output – target output] x [derivative of sigmoid at output node] x $[(\mathrm{weights})_j$ to output node] x derivative of sigmoid at hidden layer node x $[(\mathrm{input})_I$ to hidden layer nodes (for j = 1,2,3 and i = 1,2,3,4)

New updated weights = [Initial weights] — [learning rate * $\frac{d(E)}{d(wij)}$]

Step 4: update the bias at hidden node

calculate derivative of error **E** with respect to the biases at hidden node, which after simplification is equal to:

[Network output – target output] x [derivative of sigmoid at output node] x $[(\mathrm{weights})_j$ to output node] x [derivative sigmoid at hidden node]

New updated biases = [Initial bias] — [learning rate * $\frac{d(E)}{d(BHi)}$]

The calculation repeated until the error optimized.

Thorax Diseases localization

After obtaining a trained model, heatmaps are generated on test images that point out the exact location of the thoracic diseases. We calculate the heatmap and generate the B-Box for pathological candidates by using activations from the gradient of the transition layer and weight from the prediction layer.

Multi-label classification framework produces the heatmap which indicates the approximate spatial location of particular thoracic disease class each time. This helps to visualize the most indicative areas used by the model during classification. The heatmaps are created by pooling along the channel dimension and then averaging the class-wise features. The heatmaps show that the network that is trained on this architecture generalizes well and demonstrates good interpretation ability in localizing the areas of interest.

The deep learning model is generally viewed as a black box, which is particularly troubling in a clinical setting where physicians expect the model to clarify itself. Since the features that the model uses to perform the task can be somewhat different from human aspects. In this study, we have used the gradient Class Activation Map (Grad-CAM) method proposed by [39] to generate an interesting heatmap and to help the doctor assess the model performance. Grad-CAM is a CAM version that takes into account not only the weights but also the gradients going into the final convolution layer. Gradients have the advantage of being able to be applied to any layer of the network. Still, the last one is especially relevant for the localization of the parts of the image that contribute most to the final prediction. Furthermore, the layer used as input for the prediction can be followed by any module and not only by a fully connected layer. Grad-CAM makes use of the parameters ω_k^c, which denote neuron importance weights and are computed as:

$$\omega_k^c = \frac{1}{Z}\sum_i \sum_j \frac{\partial y^c}{\partial f_{ij}^k} \tag{2.11}$$

Where $\frac{1}{z}\sum_i \sum_j \ r$epresents the global average pooling operation (Z = i · j) and $\frac{\partial y^c}{\partial A_{ij}^k}$ denotes the back propagation gradients. In the gradient expression, y^c is the class c score and f_k is the k-th feature map. Then the Grad-CAM of class c at position (x, y), denoted as equation 2.12.

$$M^{c}_{Grad-CAM}(x,y) = Relu(\sum_{k} \omega^{c}_{k} f_{k}(x,y)) \tag{2.12}$$

where the ReLU operator convert the negative values to zero.

2.8 Performance Evaluation Metrics

In this paper, the performance of classification and localization was examined using four metrics: accuracy, precision, recall and f-score. The accuracy is the ratio of number of correctly classified samples to total samples. The number of pathological samples that are correctly identified as pathological sample by the classifier is called true positive (TP). The number of pathological samples that are incorrectly classified as normal by the classifier is called false negative (FN). The number of normal samples that are correctly classified as normal is called true negative (TN), and in a similar fashion, the number of normal samples that are incorrectly identified as pathological samples is called false positive (FP). The proportion of pathological samples that are correctly classified as pathological samples is referred to as the true positive rate (TPR), the formula is given as:

$$\text{Recall} = TPR = Sensitivity = \frac{TP}{(TP + FN)} \tag{2.13}$$

TPR is also called sensitivity which is called such as this measure shows the degree to which does not miss a pathological sample. Recall attempts to measure the number of positive class predictions made out of all positive examples in the dataset.

False positive rate (FPR) is proportion of normal samples that are incorrectly identified as pathological samples, given as,

$$FPR = 1 - Specificity = \frac{FP}{(FP + TN)} \tag{2.14}$$

The measure specificity shows the degree to which correctly identifies no-finding samples as no-finding. A classifier's goal is to have high sensitivity as well as specificity, resulting in a minimal diagnosis error.

Accuracy is a measure of how many accurate predictions the model has made for entire test datasets. The following formula is used to measure:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{2.15}$$

Precision attempts to measure the number of positive class predictions that actually belongs to the positive class

$$\text{Precision} = \frac{TP}{(TP + FP)} \tag{2.16}$$

F1 Score is the weighted average of Precision and Recall. In many cases, F1 is more valuable than accuracy, especially if the dataset has an unequal class distribution.

$$f1 - score = 2 * \frac{(recall * precision)}{(recall + precision)} \tag{2.17}$$

2.9 Summary

In this section thoracic diseases classification and location identification overview, overall deep learning algorithms, overview of the preprocessing methods are discussed in detail.

Medical images are complex processing and images are high-dimensional vectors. As a result, DNN is a far more effective and recommended approach for complicated data than shallow networks, and we anticipate better results for classification and localization of thoracic diseases from x-ray images.

Chest X-Ray (CXR) is one of radiology examinations which is most widely available today. These are typically radiologists ' first choice because of their non-invasive ability to disclose information that can often go unnoticed, such as physiological shifts, low cost and low radiation exposure.

Image acquisition is a highly important step for the automatic quality control because it provides the input data for the whole process. The preprocessing involved in color conversion, image resize, noise removing and enhances the quality and produces an image in which lesion or abnormality can be detected correctly. The main goal of the preprocessing stage is to minimize noise, undesired artifacts, and image distortions.

Convolutional neural networks are a type of deep learning techniques which has been dominant in a range of computer vision tasks and is gaining interest in a variety of areas, including radiology. Convolutional neural networks are made up of many different building components, such as convolution layers, pooling layers, and fully connected layers, and are designed to learn spatial hierarchies of features automatically and adaptively using a backpropagation algorithm. Understanding the concepts, capabilities, and limitations of convolutional neural networks is

important for realizing on its promise to improve radiologist performance and, ultimately, patient care.

Convolutional neural network has some fundamental fields such as:

- Local receptive field
- Sharing of weights
- Sub-sampling by time or space so as to extract features and reduce the size of parameters.

Chapter 3

Literature Review

Review of earlier studies on both medical image segmentation and classification individually and combined together has done under this chapter. The approach to segmentation and classification of the medical images can be mainly divided into different classes: region-based approach, texture – based approach and hybrid approaches.

3.1 Texture Classification

The objective of texture classification is to assign an unknown sample image to one of a set of known texture classes. Texture classification is one of the main domains in the field of texture analysis. Texture analysis is useful in many computer image analysis applications, such as classification and segmentation of medical images based on local specific changes in pixel intensity or color. A successful classification or segmentation requires an efficient description of the image texture [40]. A key field in the classification of texture is industrial identification and biomedical surface surveys, such as the identification of defects and diseases, ground classification and segmentation of aerial and satellite image, textured regions segmentation during document analysis, and content-based access to the image databases. Despite the fact that texture analysis has numerous potential areas of application in enterprise, there are just a few effective instances accessible. A wide range of approaches for analyzing image texture have been developed, and this texture analysis methods can be divided into four categories: statistical, geometrical, model-based, and signal processing [41].

3.2 Thresholding Approach

To create the binary images the thresholding is the basic segmentation in which the pixels having gray level intensity greater than the threshold are in the foreground region and the pixels having threshold value less than the threshold are in the background region [42]. The thresholding approach works very well for high contrast objects of a sharp edge, but it often fails when the edges vary in intensity, due to noise and blurred boundaries. The threshold algorithm is a technique of low-level segmentation based on local space domain features. The techniques in volumetric data sets are complex and time consuming.

3.3 Region-Based Approach

The region-based segmentation algorithm mainly consists of region growing, region division and region merger techniques. The region-based classification technique basically includes regional growth, regional division and regional fusion approaches.

In a region growing algorithm, the seed point will grow on the basis of similarity criteria with the next pixels, and the similarity criteria may be gray, shape, size or color, etc.[43].The region merging algorithm merges the homogenous regions based on the predefined similarity criteria. The seed points selection and criteria for similarities is vital to the segmentation results of regional-based techniques [43].

In many applications, the edges are important, and a convolution operation performed on an image by a mask with the proper values is used. The canny is a good edge detector for edge enhancement and tracing [44].

For noisy input images, a restore stage must be taken to flatter the image while maintaining the edge. The smoothing of the image can sometimes lead to a misdetection of the edges and can be avoided by multi-resolution border detection and border tracing. The region merging after region growth removes high frequency artifacts using seed point selection based on local statistics and was utilized for breast cancer and liver cyst analyses. The region-growing algorithm with edge detection and morphological operations segments lung parenchyma on CT images to diagnose lung diseases [45].

In the work [46] a region merging algorithm based on the Euclidean distance was proposed as a similarity measure for bone segmentation in radiographic hand images, taking into account similarity constraints in the local and extended regions. According to [47], the fuzzy-based Tsallis entropy Thresholding gives good results for MR brain and abdomen CT images with noise and complicated backgrounds.

The authors in [48], proposed Multiscale Region Growing (MSRG) technology for coronary artery segmentation in 2D X-ray angiograms. With the Frangi filter, the Multiscale Region grows produce good results and is not susceptible to noise and artifacts.

3.4 Graph Cut Method

Graph cut segmentation algorithm method is based on mathematical graph theory. A graph shows a set of nodes or vertices, with edges representing the connection between the neighboring vertices. When images are displayed as graphs, a pixel is defined using nodes and the boundaries are the link that connects a pixel with a weight that corresponds to the cutting cost [49]. The special point that shows object and background grading points are S (Source) and T (sink) [49]. The n-links are the edges that connect between the nearby nodes and the t-links are the edges that connect between the terminal node's pixels. The O and B represents the object and background seeds such that $O \cap B = \emptyset$. The main objective of the graph cut algorithm is to make an optimal cut that separates the object and the background.

Previous knowledge of the shape of the object (desired ROI) can be included in the graph cut algorithm to produce a robust result. The prior shape of the graph cut segmentation algorithm produces optimum results than the conventional graph cut algorithm. The star shape of the previous graph cut model includes a balloon-based objective function so that larger object segmentation can be performed. It is also efficient for multi-object segmentation in 3D images.

3.5 Artificial Neural Network Method

An artificial neural network (ANN) is a computer concept that aims to account for the human brain's parallel nature. It is a densely interconnected network of parallel processing units (neuron). These elements (neuron) are inspired by biological nervous systems [50]. The Artificial Neural Network has already made a major impact in the field of medical image processing. Artificial neural networks are frequently used for segmentation and classification by means of an adaptive learning approach. Artificial neural networks are represented by a set of nodes, often arranged in layers, and by a set of weighted direct links connecting them. In the work [51], the nodes are the information processing units and the links act as a medium of communication. There are many different types of neural networks, based on the nature of the information processing performed at individual nodes, the architecture of the connections, and the adaptation of the connection weights in the network [51]. Training and learning are two significant features of the neural network. The neural network is trained with features that may be statistical features such as mean, standard deviation, kurtosis, skewness or transform-based features by applying wavelet transform or curvelet transform. During training the neural network, the initial stage can

be described as a guessing stage and, as the training progresses, a stable state will be reached. Increased training in the neural network is easier in terms of query data. The more training the neural network receives, the better the result will be with regard to the input data. Learning is an adaptive process where the weights associated with the interlinking neurons change to provide the best response. Under the work [51], the neural networks can be categorized into supervised learning technique and unsupervised learning technique depending on the learning process. Neural Network algorithm has a main complication in the selection of different architectures, network sizes, type, number of layers and topologies. Choosing these factors has an impact on the neural network processing performance.

The Neural network based on fuzzy logic are used in broad areas of medical image applications, such as tissue classification, identification of abnormalities in mammographic images, MR brain images, etc. Fuzzy neural networks are generally classified into three classes concurrent fuzzy neural system, cooperative fuzzy neural networks and hybrid techniques. Fuzzy neural networks are noise tolerant and offer better results on segmentation. The neural network based on wavelets is commonly used in the segmentation, compression, classification of medical images. An interactive segmentation algorithm was used for the brain tumor and kernel function segmentation based on Support Vector Machine (SVM), Kth Nearest Neighbor (KNN), and Decision tree produce better result [52]. From the work [53] that the single hidden layer feedforward neural network along with a 3D fast marching algorithm was used for segmentation of the liver tumor from the MR images of the abdomen. The results were validated with a ground truth image, time complexity was significantly reduced and accurate results were obtained when compared to other semi-automatic segmentation approaches [53]. In [54], the KNN classifier provides reasonable accuracy for the identification of Abdominal Aortic Calcification (AAC) in dual Energy X-Ray Absorptiometry images compared to the SVM technique. The authors of [26] have extensively investigated that the Deep Convolution Neural Network (DCNN) based automatic brain tumor segmentation model incorporates local and global contextual features with two-phase training stages.

For the last 50 years, the idea of computer-aided diagnosis (CAD) for chest x-rays has made significant progress from rule-oriented survival prediction from lung x-rays to machine learning approaches to now deep learning [55]. Bram Ginnek, in [55]argues that radiology requires CAD because of the increasingly becoming unmanageable workload of radiologists.

Computer assisted diagnosis was first, requested for different diagnostic tasks: GoogLeNet was used for identification of location of breast cancer [56] and Esteva et al. in [57] was used for classification of skin cancer.

These two well-designed networks, along with many others, showed that convolutional neural networks can be used very well not only for classifying natural images, but also for classifying and segmenting medical images. Rajkomar etal.[58], in the field of chest x-ray (CXR) classification tasks, used GoogLeNet along with image augmentation and pre-training on ImageNet to classify and identify CXR images as either front or side with 100 percent accuracy. While this is not directly clinically applicable, it is an important proof of the principle of the use of deep learning in CXR photos. They noticed that the 5-layer convolutional neural network was far more efficient than the image descriptor-based similarity. Such a network may be used to help physicians quickly search past cases and help educate their current or potential diagnosis. In [59] used a CNN to identify and assign disease labels for particular diseases in CXR images. The annotated disease was then identified with an RNN based on the CNN and patient metadata.

In this work they were only able to achieve the validation accuracy of 0.698 for this ambitious mission. This performance may be due to their relatively small data set size of 7470 images, the complexities of multi-class classification and the inclusion of textual data from patient records. The paper in [19] successfully designed a CNN to identify and recognize high-precision lung nodules in CXR images to diagnose particular diseases.

While these methods serve as very good proof of perception and are useful for assisted diagnosis, they are not easily generalizable to various diseases, as new training data and labels must be collected for retraining models for each particular sub-question. The research in [60], the image registration technique was used to locate the heart and lung region and then computed radiographic indexes such as cardiothoracic ratio (CTR), cardiothoracic area ratio (CTAR) to distinguish cardiomegaly from X-ray images.

We strive to identify images into a particular abnormality class and locate particular categories of disease, and we will provide a much more flexible system that can accessible to a much wider patient population.

3.6 Summary

In this chapter, we have revised some of the common approach to segmentation and classification of medical images.

In the field of image processing and computer vision, segmentation is still a challenge in many real-time applications and thus more creative work is required. Various types of segmentation algorithms are discussed and the findings of some common algorithms are presented in this paper. It is factual that there is no universal algorithm for the segmentation of medical image, in which the choice depends on the image modality, the characteristics of the region of interest and the submission.

The revised studies in this chapter have been, abnormality detection and localization in chest x-ray images system uses different approaches: In general, there are many methods that have been proposed to solve these challenging problems on image classification, which can be categorized into two broad types of methods: statistical or traditional method and deep model methods. Traditional methods include color and texture, thresholding, region based, watershed, graph cut, random forests, and support vector machines [27]. Deep model studies to classify and locate tumors or cysts on medical images include [26]-[27].

From the experimental result of the revised papers, deep model and hybrid or combination of deep model approaches showed good performance, deep neural network combined with other deep neural network can be extended to medical images processing and for natural languages processing. In general, deep neural approach with deep layer and optimized unit of neurons can be applied in medical images processing for anomality and anatomy recognition and are expected to provide better result.

Chapter 4

Design and Implementation

4.1 System Description

This section discussed the system architecture and diagrammatical work flow of automated classification and localization of thoracic diseases and the implementations. The proposed method contains different image processing techniques with the integration of state-of-the-art method of deep learning techniques, CNN algorithm.

The system architecture consists of four major stages namely input data acquisition, preprocessing, segmentation, diseases or abnormalities classification and location identification. In the image preprocessing stage: there are four sub stage that are applied on the input images: first, reshape or resize the input images to reduce memory and shortening the training time. Secondly, the images are normalized to make the image familiar or normal for the senses and to increase contrast, which helps to improve the extraction or segmentation of the image. Thirdly, histogram equalization (evenly distributes the occurrence of pixel intensities) and balance poor illumination. Therefore, the images must be unified within a certain size as required by the proposed network. Fourthly, we use median filter, a nonlinear operation used to reduce artifacts, salt and pepper noises. The output of this preprocessing method is fed into segmentation stage, segmentation aiming to remove all background and retain only the lung region.

The next part is classification, in this stage, a deep neural network is trained, using the preprocessed and segmented data, to be used for classifying the diseases into one of the predefined classes. Finally, from pretrained model of classification stage, identify the approximate location of the recognized diseases.

4.2 System Model

The proposed system includes five basic stages as shown in Figure 4.1 data collection, preprocessing, Segmentation, diseases classification and location detection.

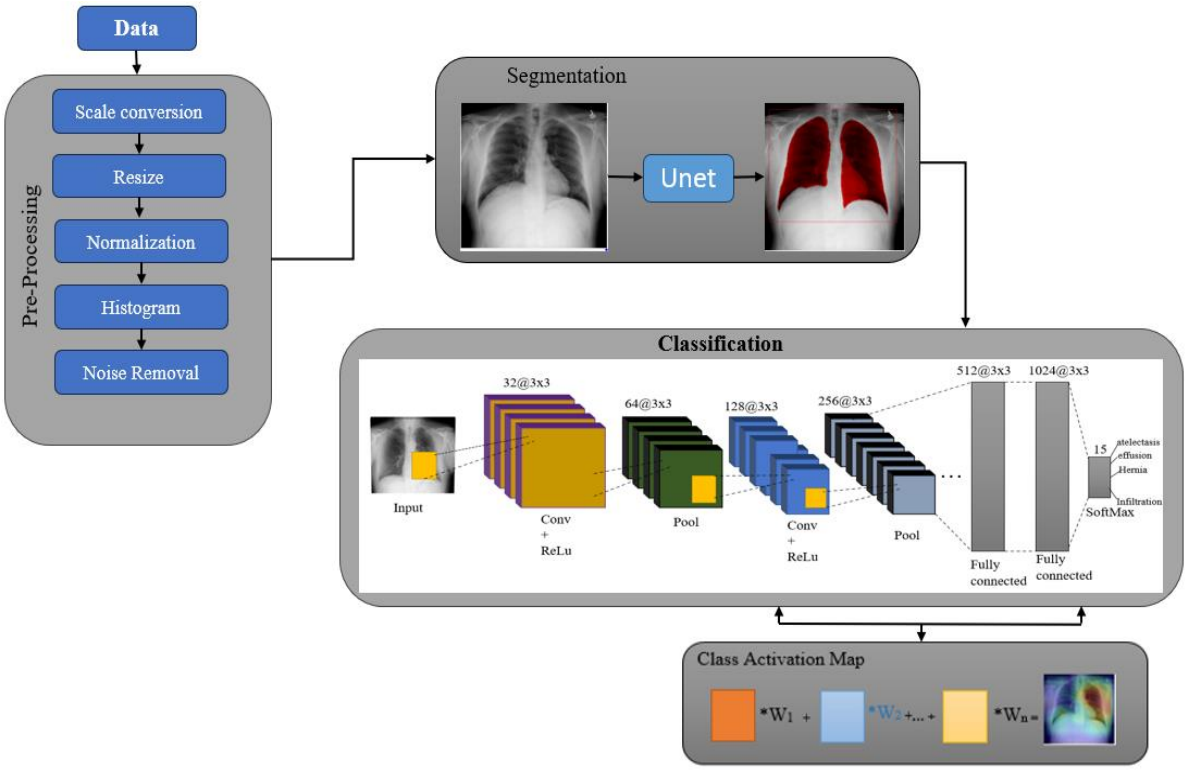


Figure 4. 1 The proposed system's block diagram

Algorithm: Thorax disease Classification and localization Procedure

Input: Image I; Label vector L

Output: Probability score

1. Lung segmentation mask M using U-net
2. Generate bounding box [top, right, bottom, left] using largest connected component on mask M
3. Crop image Ic from I use bounding box and resize I_c to appropriate input size (224, 224)
4. Compute probability score base from I_c using Inception Model
5. Compute Grad-CAM for I

4.3 Data Acquisition

The input image for the proposed system is collected from national institute of health clinical center (NIHCC). The data set consisted of 112,120 x-ray images in frontal view from the accompanying radiological report utilizing natural language processing of 30,805 unique patients with text-mine 14 disease image labels (which might be multiple labels in one image).

4.4 Preprocessing

In this stage, preprocessing is applied on the collected X-ray images to improve the quality of images, to clean unwanted pixels and to cut rib cage areas. There are five steps in this stage: cropping images, gray scale conversion, image resize, normalization, histogram equalization and segmentation.

Gray scale conversion: The input images acquired from national institute of health clinical center, which are in RGB color format are converted in to gray scale format using the luminosity or luma method. It converts RGB data to grayscale values (Luminance) by generating a weighted sum of the Red (R), Green (G), and Blue (B) components using the formula:

$$Luminance = 0.299 * R + 0.587 * G + 0.114 * B \quad (4.1)$$

The steps of the algorithm are as follows:

1. Reade or load the input image
2. Compute the source color image's brightness and chrominance values.
3. Calculate the sum of luminance and chrominance
4. The RGB values are approximated using RGB components
5. Calculate the average of the four values (luminance, chrominance, the sum of luminance and chrominance and RGB values)
6. Finally, we get (I_1) the resulted gray color image

Resize (reshape): input image resizing changes the image dimensions of height and width to make the input image suitable or fit for the convolutional neural network. The original dataset dimension was 1024x1024. Image size has impact on network model training and resource consumption. In this work, the dataset changed to 32x32 to optimize the memory of the system and to shorten the time taken to train the neural network. In neural networks, big integer inputs can disrupt or slow down the learning process, whereas tiny weight inputs can process the learning. As a result, resizing the input image is a smart technique to make the neural network model's learning process faster.

Normalization: The need of image normalization is to make pixel distribution uniform. Image normalization is often used to increase contrast, which helps to improve the extraction or

segmentation of the image. It also helpful to remove high frequency noise and very low noise from the image. Input data normalization can be computed using equation (2.1).

The detailed steps of the algorithm are discussed below:

1. Load the image
2. Convert the image into the NumPy array
3. Show the data type of the NumPy array
4. Show the minimum and maximum pixel values across all the three channels
5. Convert the array's data type to float.
6. Normalize to a range of [0,1]
7. Confirm the normalization

Histogram equalization: distributes pixel intensities equally so that the whole range of intensities is considered. This approach often improves the global contrast of images, particularly when the image's useful data is represented by near contrast values. This modification can better disperse the intensity of the histogram. This permits areas with poor local contrast to gain contrast. Probability density function (pdf) is calculated for the histogram[8]. The histograms equalization can be calculated using equation (2.2).

The steps for histogram equalization algorithm are as follows:

1. Read the image
2. Examining the image's shape (size). This means our image has channels (RGB), height (number of columns) and width (number of rows).
3. Finally, we loop through the rows and columns and print out the different pixel values (intensities) at each row or column pair.

Noise removal using median filter: Median filter is a nonlinear operation used to reduce artifacts, salt and pepper noises [8]. Noise filtering using median filters is easier to eliminate noise and retain digital images' edges and fine information of images. we have used median filter technique, since a median filter it retains sharp edges of images and fine details of digital images in the X - ray under certain conditions while removing noise [29]. The median filter value is calculated using equation (2.3).

The steps for median filter algorithm are as follows:

1. Setting window size
2. Sorting all the pixel values from the window into numerical order
3. Then replacing the pixel being considered with the middle (median) pixel value in the neighborhood of that pixel

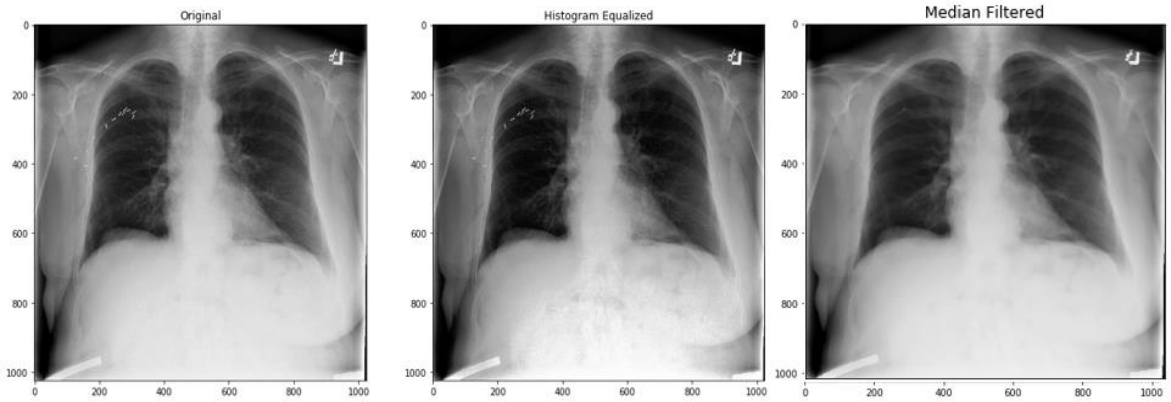


Figure 4.2 preprocessed gray scale images (normalized, histogram equalized and filtered)

4.5 Segmentation

Image segmentation is the process of dividing an image into multiple parts [32]. As mentioned in Section 2.6, segmentation is often used to identify objects or other useful information from images. The output of the preprocessing step is fed into the segmentation stage in the proposed system design.

4.5.1 Lung Segmentation

After preprocessing stage, lung segmentation aiming to remove all background and retain only the lung area. we need to get the region of interest, which is where most pathologies are expected to be available. The purpose of image segmentation is to extract the area or region of interest (ROI) in an image. Lung segmentation or finding region of interest (ROI) make to reduce search space and speed up tuning computation parameters. Thus, to get the most available region of interest it needs to be perform lung segmentation using different segmentation techniques, in this study we used U-Net lung segmentation technique which is used by many researchers [34] now a day and more efficient state of the art technique. Following segmentation, we used a morphological opening of 5 pixels to eliminate tiny bright spots that typically occurred outside the lung area. We also used a 5-pixel morphological dilation to enhance and smooth the

predicted mask boundary. Finally, we also cropped all images to keep only the ROI indicated by the mask.

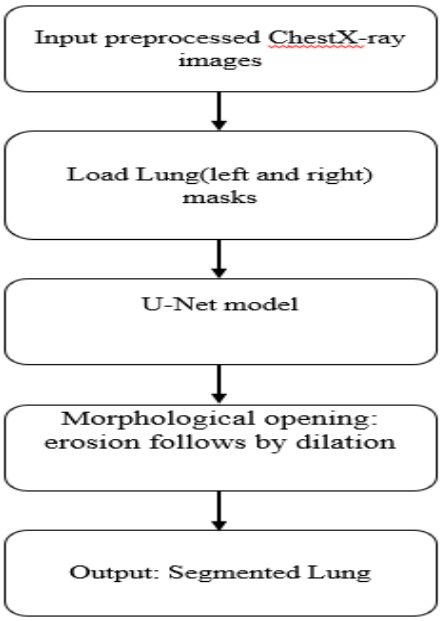


Fig 4.3 Lung Segmentation Block Diagram

Table 4.1 U-Net model

Layer name	Input Size	Convolutional Filter size
Epoch	100	
Optimizer	adam	
Batch size	64	
Conv	3@32× 32	$7 \times 7, 64, s = 2, p = 3$
Max Pool	64@64× 64	$2 \times 2, s = 2$
ResBlock	64@112× 112	$\begin{bmatrix} 1 \times 1, 64, s = 1, p = 0 \\ 3 \times 3, 64, s = 1, p = 1 \\ 1 \times 1, 256, s = 1, p = 0 \end{bmatrix} \times 3$
ResBlock	256@56×56	$\begin{bmatrix} 1 \times 1, 128, s = 1, p = 0 \\ 3 \times 3, 128, s = 1(2), p = 1 \\ 1 \times 1, 512, s = 1, p = 0 \end{bmatrix} \times 4$
ResBlock	512@28×28	$\begin{bmatrix} 1 \times 1, 256, s = 1, p = 0 \\ 3 \times 3, 256, s = 1(2), p = 1 \\ 1 \times 1, 1024, s = 1, p = 0 \end{bmatrix} \times 6$
ResBlock	1024@14×14	$\begin{bmatrix} 1 \times 1, 512, s = 1, p = 0 \\ 3 \times 3, 512, s = 1(2), p = 1 \\ 1 \times 1, 1024, s = 1, p = 0 \end{bmatrix} \times 3$
Adaptive Pool	1024@7×7	$7 \times 7, s = 7$
Fully Connected	1024@1×1	1024×14

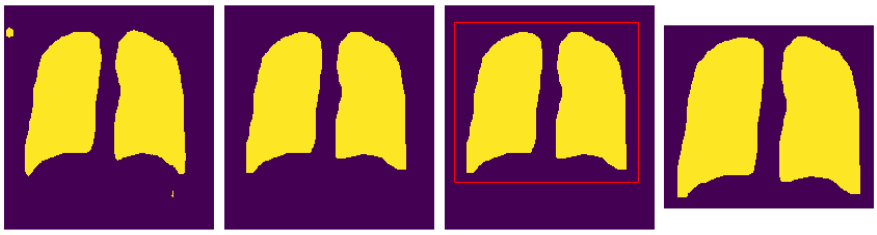


Figure 4.4 Segmented and post processed Lung Region

4.6 Model Development and Training

4.6.1 Classification

In this section, the proposed system of the model development, the model parameters, and the training algorithm of deep neural network used for classification and localization of chest diseases is presented.

Network Architecture: - The proposed network architecture consists of a 1024 input features or neurons each having binary values that are obtained from segmented x-ray images normalized to 32x32 vector size. The number of output nodes or units is determined by the number of unique classes at the flatten layer, in our case the number of classes are 15 fourteen type of thorax diseases class (atelectasis, cardiomegaly, consolidation, edma, effusion, emphysema, fibrosis, hernia, infiltration, mass, nodule, pleural thickening, pneumonia, pneumothorax) and one non pathology (normal) class. As shown in figure 4.5 CNN architecture consists the most important parameters (the learning parameters). **CNN pattern**: $input \rightarrow conv \rightarrow ReLu \rightarrow conv \rightarrow ReLu \rightarrow pool \rightarrow ReLu \rightarrow conv \rightarrow ReLu \rightarrow pool \rightarrow Fully\ connected$

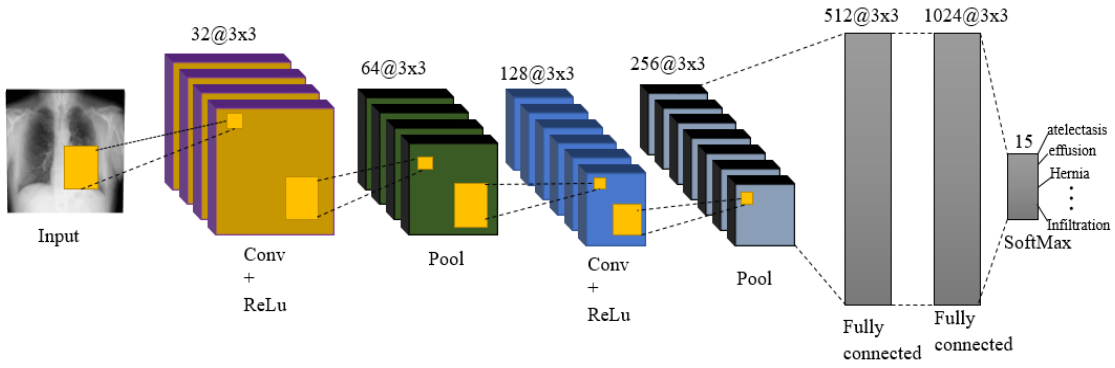


Figure 4. 5 depicts the proposed system's CNN architecture

Learning Parameters: - When designing a neural network, it is important to determine the number of different network parameters. The following learning parameters are required for the proposed network given below:

Batch size: the number of training instance per batch. The typical value depends on the training data.

Learning rate: training parameter that controls the size of weight and bias changes in learning of the training algorithm. it is used to control how fast the network learns the training examples. The value of this parameter varies between 0 to 1[0,1]. The step size with which network weights are changed during training is determined by the learning rate. If the value set for the learning rate is too high, learning is completed in fewer epochs (the network only memorizing the training data) the network weights have not been properly tuned. This situation called as over-fitting. If the value set for the learning rate is too low, then the network runs insufficient learning of the training data by the time the set number of maximum epochs is reached. It makes the learning process much slower, and the network may not converge to the required set MSE.

Number of hidden neurons: the hidden layers are the layers at which the feature learnings are started**.** If the number of hidden neurons is too few, then the network has not enough knowledge about the features in the training samples. In contrast, if the number of neurons employed is too large, the network will create a much more complicated model to the training data than is necessary, resulting in over-fitting.

Momentum: Momentum is just a fraction m of the previous weight update added to the current one [32]. The parameter momentum is used to avoid the convergence of the system to a local minimum or saddle point. Increased system convergence may also result from a high momentum value. Setting the momentum parameter too high, on the other hand, increases the danger of overshooting the minimum, which may lead the system to become unstable. An extremely low momentum coefficient does not reliably avoid local minima and can also decrease machine training. Recommended value: [0, 1].

Training epoch: epoch is a measure of the number of times that weights are modified for all training vectors. For batch training every sample passes the learning algorithm at a time before weights are modified. when this value is zero it means train by epoch, and when the value is one means train by minimum error. Recommended value: [0, 1]

Cost or error function: a cost function relating the difference between the network's actual response and the targeted response. The objective of neural network training is to minimize the cost function. The network training stopped, When the network reaches this specified value for the MSE goal.

Activation function: The total potential is generated by taking the weighted sum of a neuron's inputs and passing it through an activation function.

CNN layers: the details of CNN model layers are as shown in section 2.7.3.

Convolutional layer: Convolution is the first layer that extracts the features from the input image. Convolution retains the relationship between pixels by using of small squares input data for learning image features.

Batch normalization layer: Each input channel is normalized over a mini-batch using a batch normalization layer. used to accelerate convolutional neural network training and minimize sensitivity to network initialization. Between convolutional layers and ReLU layers, batch normalization layers are used.

Strides: The number of pixels that shift through the input matrix is referred to as the stride. When the stride is 1, the filters are moved one pixel at a time. When the stride is 2, the filters are moved two pixels at a time, and so on.

Padding size: Sometimes the filter may not properly match the input image. We have two choices:

- Pad the image with zeros (zero padding) so that it fits.
- Drop the part of the picture where the filter did not fit. This is known as valid padding, because it only maintains a valid portion of the image.

Non-Linearity (ReLU): (ReLU) is an abbreviation for Rectified Linear Unit and used for a non-linear operation. The result is:

$$f(x) = max(0,x) \tag{4.2}$$

The purpose of ReLU is to introduce non-linearity in our convNet. Because real-world data would prefer to learn non-negative linear values. Other nonlinear functions, such as tanh and

sigmoid, can be used in place of ReLU. Many data analysts prefer ReLU because it outperforms the other non-linear functions.

Pooling layer: these layers are used to reduce the number of parameters when the images are too large. It is also known as subsampling or down sampling since it reduces the size of each map while preserving important features. There is different type of polling layers:

- Max pooling
- Average pooling
- Sum pooling

Fully-Connected layer: this layer integrates all of the features (local information) learnt by the preceding layers throughout the image in order to detect larger patterns. The last fully connected layer in a classification issue combines the features to categorize the images.

Table 4.2 Inception Model Parameters

Number of inception Block	3
Filter size	3x3, 5x5, 7x7
Stride	2x2 with padding same
Subsampling layer	3x3 Max-pool
Subsampling layer	Global average pooling
Optimizer and learning rate	Adam with 0.001
Regularizer	dropout (0.5)
Batch-size	64
Activation function	ReLU
Loss	Categorical cross entropy

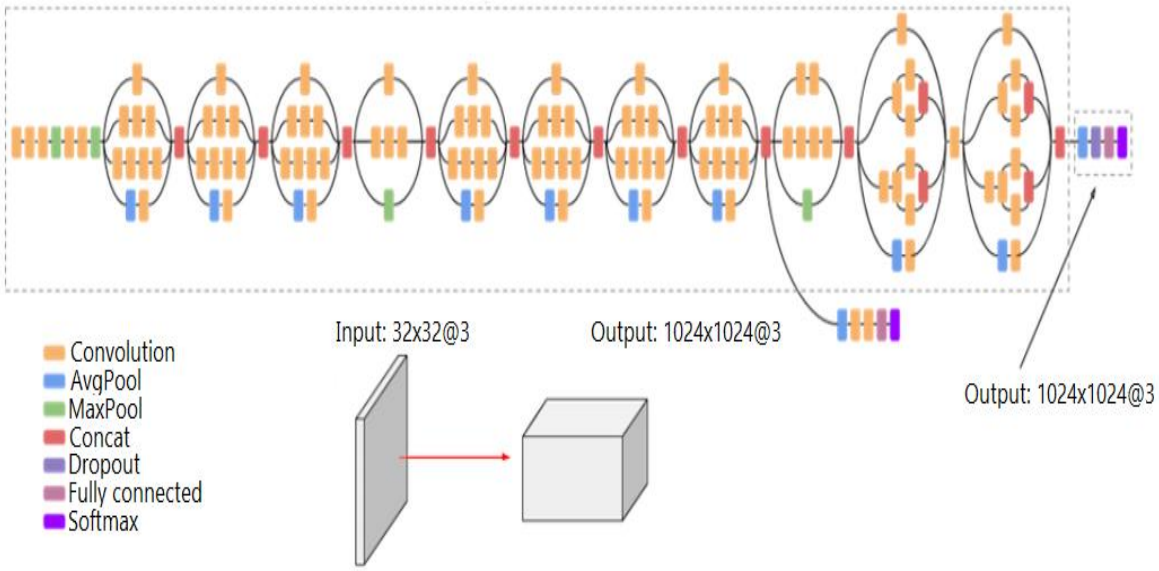


Figure 4.6 Inception Model

Training of proposed system: the following steps are used to train our procedures.

Forward pass

Step 1: Initialize network parameter

First step is to initialize weights and biases using **rand ()** function.

Step 2: calculate net input at hidden layer nodes

The Net Input is calculated by multiplying input with weight and adding bias. Using the sum-product () function, we can obtain the resulting matrix by multiplying the input vector with the weight vector.

Step 3: pass the net inputs through the activation function (sigmoid)

pass the output from Step 2 as input to the activation function at each neuron of hidden layer which can easily be done by $f(x) = \frac{1}{1+e^{(-x)}}$.

step 4: calculate net input at output node

Now the outputs from Step 3 will act as inputs to the output node. Net input at output become the product of the output (from hidden layer) and initial weights plus bias at output.

Step 5: obtain final output of neural network

The result from step 4 pass through the activation function as f(x) can be calculated using $f(x) = \frac{1}{1+e^{(-x)}}$, thus resulting in the final output of neural network.

Back propagation

After obtaining the Feed-forward output, the next step is to evaluate the network output by comparing it to the desired outcome. The challenge here is to find the optimal weights and biases that can minimize the sum of square error (MSE):

$$E = \frac{1}{2}\sum(\text{Network output} - \text{Target output})^2 \tag{4.3}$$

Where E is error function received by the network. We must examine the error produced by each of these weights and biases separately and then keep updating them to minimize the error. This process will be iterated till the convergence. The network will be called a trained network once an optimal is reached.

Step 1: updating weights at hidden layer

The calculated derivative of error function **E** with respect to weights using chain rule, after simplification is equal to:

[Network output – target output] x [derivative of sigmoid at output node] x [$(input)_j$ to output node] (For j = 1,2…).

were, Derivative of sigmoid function $f(x) = [f(x) * (1 - f(x))]$

The new updated weights will be:

$$new_{weights} = [initial\ weights] - \left[\{(learning\ rate) * \frac{d(E)}{d(wh_i)}\}\right], (for\ i = 1,2 \dots) \tag{4.4}$$

Step 2: update bias at output node

The computed derivative of error function E with respect to bias, is equal to:

[network output – target output] x derivative of sigmoid at output node, $\frac{d(E)}{d(b_i)}$, $(for\ i = 1,2,3 \dots)$.

$$new_{bias(i)} = [initial\ bias] - \left[learning\ rate * \left\{\frac{d(E)}{d(b_i)}\right\}\right], (for\ i = 1,2,3 \dots) \tag{4.5}$$

Step 3: update weights at input

Update the weights at the input to the hidden layer, and compute the derivative of the error E with respect to the weights:

[Network output – target output] x [derivative of sigmoid at output node] x [$(\text{weights})_j$ to output node] x derivative of sigmoid at hidden layer node x [$(\text{input})_I$ to hidden layer nodes (for j = 1,2,3 and i = 1,2,3,4)

$$new_{updatedweights} = [initial\ weights] - \left[learning\ rate * \frac{d(E)}{d(w_{ij})}\right] \quad (4.6)$$

Step 4: update the bias at hidden node

The computed derivative of error E with respect to the biases at the hidden node, is equal to:

[Network output – target output] x [derivative of sigmoid at output node] x [$(\text{weights})_j$ to output node] x [derivative sigmoid at hidden node].

$$Newupdatedbiases = [initial\ bias] - \left[learning\ rate * \frac{d(E)}{d(BH_I)}\right] \quad (4.7)$$

The calculation repeated until the error optimized.

4.7 Implementation

This section describes the tools, languages and libraries used to implement the proposed system. The dataset preparation for the system input and the experimental set up are also presented in detail.

4.7.1 Implementing Tools

The proposed system to classify and localize thorax diseases is implemented using the following tools and module libraries.

1. Programing language and tools:

- cuda-ubuntu1604-amd64.deb used to install GPU driver for ubuntu
- python 3.7 (anaconda3 3.7) for IDE
- Jupyter Notebook used for code editing
- NumPy, Matplotlib, Pandas, SimpleITK, etc. used for data manipulation.

2. Experimental setup:

A laptop computer with the following specifications was used to conduct the experiments.

Table 4. 3 Hardware specifications for experimental setup

Processor	Intel® core™ i5-6300HQ CPU@2.30, 2.30 GHz processor
RAM	8GB
HDD	1TB
Operating system	Ubuntu1604
Gpu	Nvidia GeForce GTX 950M 4 GB

Chapter 5

Experimental Results and Discussion

5.1 Introduction

This chapter presents the implementation details and experimental results of the proposed models for chest diseases classification and localization. We perform a comprehensive set of experiments to evaluate the performance of the proposed system. We planned to compare the outcomes of different learning rate, patch sizes and epoch size on classification efficiency and also to overcome overfitting of the designed model. In the first step, we hypothesize that the accuracy of the model will increase and the training will converge faster by using a suitable learning rate to the network, since Learning rate is the step size the model takes for each optimization step. In the second step, we hypothesize that the network sensitivity increases when we use large patch sizes, since large patch sizes can cover a large number of nodules with many background details. In the third case, we hypothesize that using longer iteration will increase training accuracy, since training longer gives the model the necessary time to converge. Section 5.2 presents the datasets used to training, testing and validating our models. Section 5.3 deliberate about the implementation and experimental results. Section 5.4 deals about the discussion.

5.1 Datasets

The datasets used to conduct these experiments are collected from national institute of health clinical center (NIHCC) Chest X-ray dataset consists of 112,120 frontal-view X-ray images of 30,805 unique patients with text-mind disease image labels (each image might have several labels), from the corresponding radiological reports extracted by using natural language processing. Atelectasis, Cardiomegaly, Consolidation, Infiltration, Pneumothorax, Edema, Emphysema, Fibrosis, Effusion, Pneumonia, Pleural thickening, Nodule, Mass, and Hernia are among the fourteen frequent thoracic pathologies, which are an extension of the eight basic disease patterns described in [22]. We split the dataset into training (78468 images, 21528 patients), validation (11219 images, 3090 patients) and testing (22433 images, 6187 patients) with no patients overlap between each set. We resize the original 3-channel images from 1024 x 1024 to 32 x 32 pixels for fast processing. The pixel values in each channel are normalized to [-

1,1]. We used only disease labels in our experiments as ground-truth in training and testing the model in the classification of diseases. We have used bounding boxes only for a visual assessment of the localization of the disease area on X-rays.

Contents:

1. A total of 112,120 frontal chest X-ray PNG images with a size of 1024*1024 (under images folder)
2. Meta data for all images (Data_Entry_2017.csv): Image Index, Finding Labels, Follow-up #, Patient ID, Patient Age, Patient Gender, View Position, Original Image Size and Original Image Pixel Spacing.
3. Bounding boxes for ~1000 images (BBox_List_2017.csv): Image Index, Finding Label, Bbox [x, y, w, h]. [x y] are coordinates of each box's top left corner. [w h] represent the width and height of each box.
4. Two data split files (train_val_list.txt and test_list.txt) are provided. Images in the Chest x-ray dataset are divided into these two sets on the patient level. All studies from the same patient will only appear in either training/validation or testing set.

The .csv files: Data_Entry_2017.csv consists of Image Index, Finding Labels, Follow-up #, Patient ID, Patient Age, Patient Gender, View Position, Original Image [Width Height], OriginalImagePixelSpacing [x y]. The sample shown as below:

	Image Index	Finding Labels	Follow-up #	Patient ID	Patient Age	Patient Gender	View Position	OriginalIma
0	00000001_000.png	Cardiomegaly	0	1	58	M	PA	2682
1	00000001_001.png	Cardiomegaly|Emphysema	1	1	58	M	PA	2894
2	00000001_002.png	Cardiomegaly|Effusion	2	1	58	M	PA	2500
3	00000002_000.png	No Finding	0	2	81	M	PA	2500
4	00000003_000.png	Hernia	0	3	81	F	PA	2582

Figure 5. 1 Sample Data_Entry_2017.csv file

BBox_List_2017.csv consists of Image Index, Finding Labels, Bbox [x y w h] used to get the location map. The sample shown as figure below:

	Image Index	Finding Label	Bbox [x	y	w	h]
0	00013118_008.png	Atelectasis	225.084746	547.019217	86.779661	79.186441
1	00014716_007.png	Atelectasis	686.101695	131.543498	185.491525	313.491525
2	00029817_009.png	Atelectasis	221.830508	317.053115	155.118644	216.949153
3	00014687_001.png	Atelectasis	726.237288	494.951420	141.016949	55.322034
4	00017877_001.png	Atelectasis	660.067797	569.780787	200.677966	78.101695

Figure 5. 2 Sample of BBox_List_2017.csv file

The dataset used to conduct this experiment is summary as table 5.1.

Table 5. 1 Summary of the Dataset

Data type	Resolution	Total number of images		Input vector size 32x32x1
PNG	1024x1024	112,120		
CSV		Bbox	Meta data	
		1000	112,020	

5.2 Implementation and Results

Hyperparameter tuning: We need to set some hyperparameter values before conducting the training process in order to achieve better classification results. The Deep Learning (DL) training model relies heavily on hyperparameter tuning. We performed several experiments to get the best set of hyperparameters for model training. The experiment we conducted was separated into three categories:

- Evaluation of tuning parameters on the classification performance
- Evaluation of different convolutional models
- Evaluation of preprocessing step to enhance classification performance

5.2.1 Evaluating Tuning Parameters

In this experiment we evaluate the following tuning parameters: Optimization algorithms, different learning rate, batch size, iteration size, different convolutional layers, preprocessing and Segmentation.

5.2.1.1 Evaluating Optimization Algorithms

There are different learning rate schedules and adaptive learning rate optimization methods while training deep neural networks. Learning rate plans tend to modify the training rate by lowering the learning rate on a predefined timeline. The challenge with utilizing learning rates is that we must set our hyperparameters early on and rely largely on the type of model and task. Another issue is that the learning rate for all parameter changes is the same. When there is a scarcity of data, we may want to alter the parameters. Adagrad, Adadelta, RMSprop, and Adam are adaptive gradient-descending algorithms that can be used instead of traditional SGD. These per parameter learning rate approaches provide an optimization technique without the time-consuming and costly task of manually setting hyperparameters for the learning rate schedule. Adam has low memory requirements and is invariant to the diagonal rescaling of the gradients. The Adam optimizer also has an efficient regret bound on the convergence rate. It is also well suited for problems involving large size of data. The following figure 5.3 below shows the comparation of Adaptive gradient-descending algorithms.

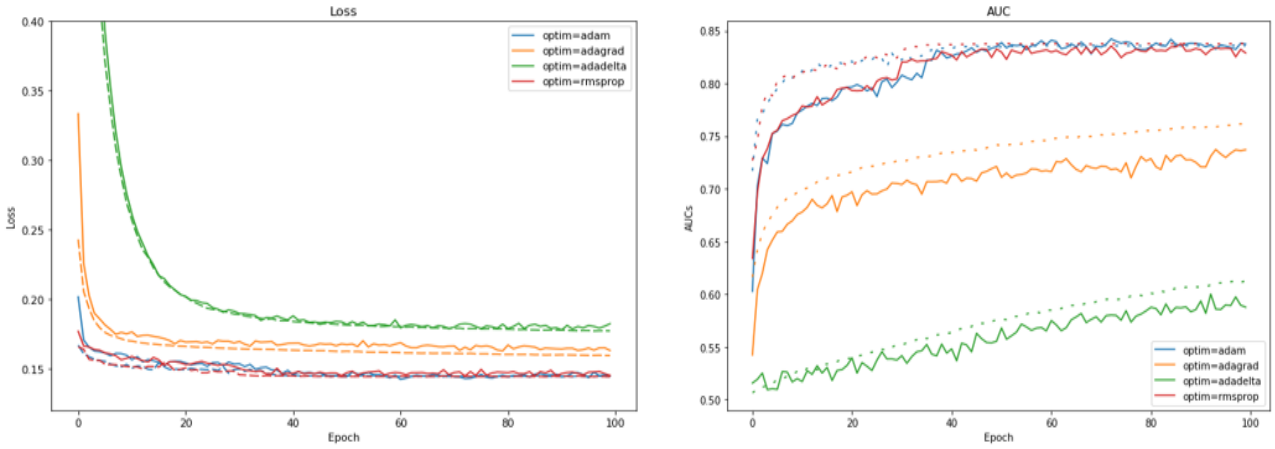


Figure 5. 3 Optimization algorithm comparison

As shown from the above experiment of evaluation of common optimizing algorithms adaptive momentum optimization (Adam) algorithm provide best result and root mean square probability (RMSprop) optimization algorithm give next best result.

5.2.1.2 Evaluating Learning Rate

The learning rate is the step size of the model for each step of optimization. High learning rates lead to unpredictable and possibly divergent learning. Low learning rates lead to longer converging time. We experimented with different learning rates ranging from 0.000001 to 0.01 ($1e^{-6}$ to $1e^{-2}$) with exponential step size. As shown In Figure 5.4, when the learning rate is 1e-6, learning has progressed very slowly. When the learning rate is 1e-2, learning progress cannot be converging. So, as we can see from the figure, we selected the learning rate to be 1e-4 in this experiment.

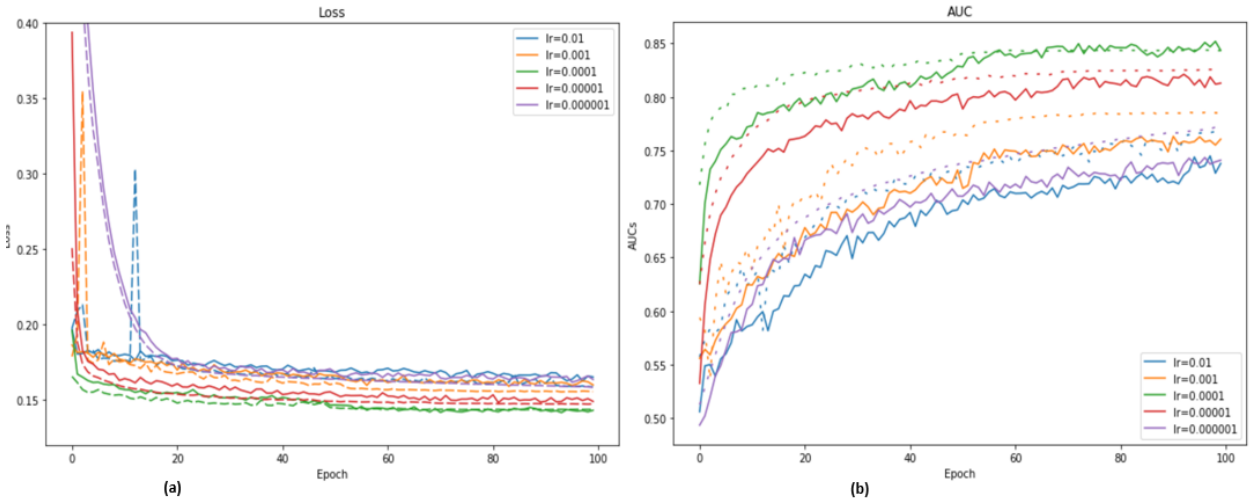


Figure 5. 4 Effect of Learning rate on model performance

(a) loss vs epoch for different learning rates **(b)** accuracy (AUC) vs epoch for different learning rates.

5.2.1.3 Evaluating Batch Size

Batch size is the number of image samples from the training dataset used in the error gradient estimation for each optimization step. It is a significant hyperparameter that affects the dynamics of the learning algorithm. The impact of batch size on the learning process is as follows:

- Batch size controls the accuracy of the estimate of the error gradient during training neural networks (measures the precision of the error level estimation).
- It also controls speed and stability.

Depending on this reason, we experimented with different batch sizes ranging from 16 to 128 with an exponential step size.

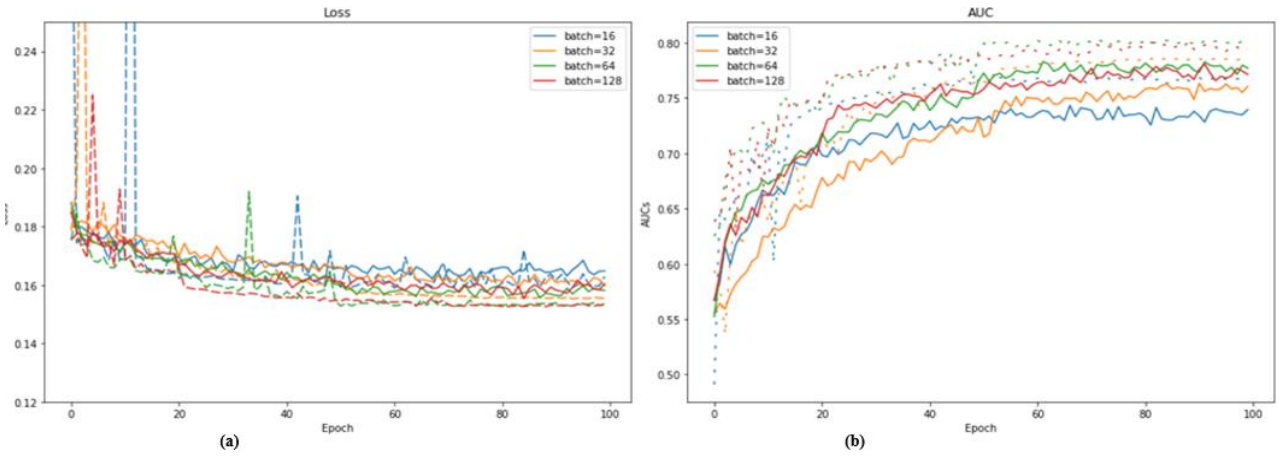


Figure 5. 5 Effect of batch size on model performance

(a) loss vs epochs for different batch size **(b)** accuracy (AUC) vs epochs for different batch size

As we can justified from the experiment graph (figure 5.5), when the patch size increases the accuracy (AUC) of the model also increase up-to some batch size, since large patch sizes can include a large number of possibilities of pathology as well as basic information. However, when the batch size excessed 64, there is no more benefit for bigger batch size. Therefore, exceeding batch size more than some typical size does not have important benefit for the learning process but it has a high impact for memory consumption. To process large batch size (to upload large batch size image) it requires larger memory capacity. Due to this, we have chosen the best batch size to be 64 in this experiment.

5.2.1.4 Evaluating Iteration

When the complete dataset is distributed forward and backward across the neural network only once, this is referred to as an Epoch. The epoch is when we iterate over the whole training set, the longer epoch gives the model the necessary time to converge. We have conducted experiment with different epoch ranging from 100 to 500 with a step size of 100.

The model loss and performance stable after 100 epochs, as shown in Figure 5.6. Based on this experiment, we determined that the epoch should be 100.

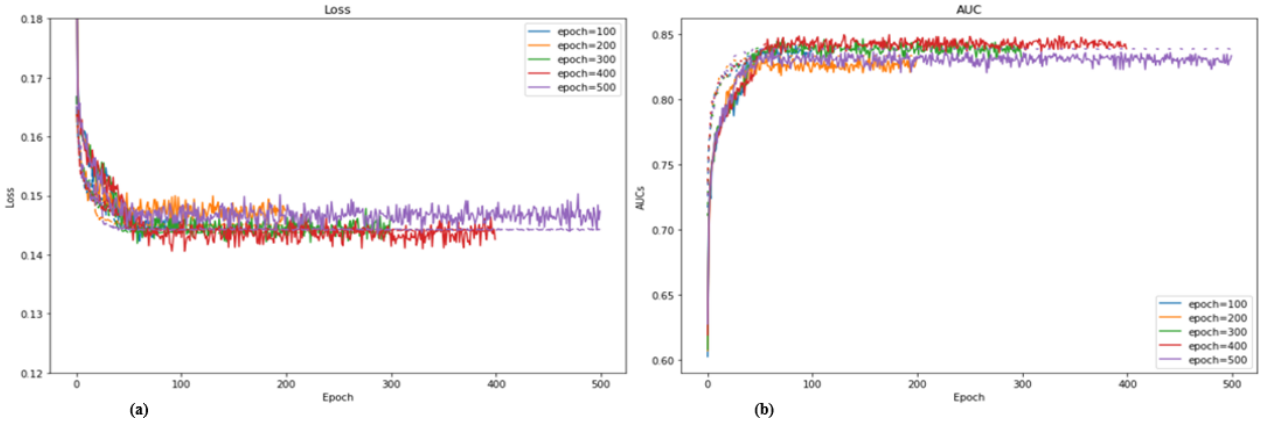


Figure 5. 6 Iteration evaluation on model performance

(a) loss vs epoch for different epochs **(b)** accuracy (AUC) for different epochs

5.2.2 Evaluating Different State of the Art CNN Models

Inception network basically acts as multiple convolution filters, that are applied to the same input, with some pooling. This allows the model to take advantage of multi-level feature extraction. For example, it extracts both global (5x5) and local (1x1) features simultaneously. Using multiple features from different filters improves network performance. Aside from that, there is another factor that makes the inception architecture better than others. By performing the 1x1 convolution, the inception block is doing cross-channel correlations, ignoring the spatial dimensions. The 3x3 and 5x5 filters are then used to perform cross-spatial and cross-channel correlations. Convolutions, average pooling, max pooling, dropouts, and fully connected layers are among the symmetric and asymmetric building elements used in the model. Batch normalization is performed to activation inputs and is utilized extensively throughout the model. SoftMax is used to compute loss. Inception model Work with Factorizing Convolutions. Factorizing Convolutions used to reduce the number of connections and parameters to learn. This will increase the speed and gives a good performance. As shown from the experiment in figure 5.7 inception model give better performance than densnet and resnet model.

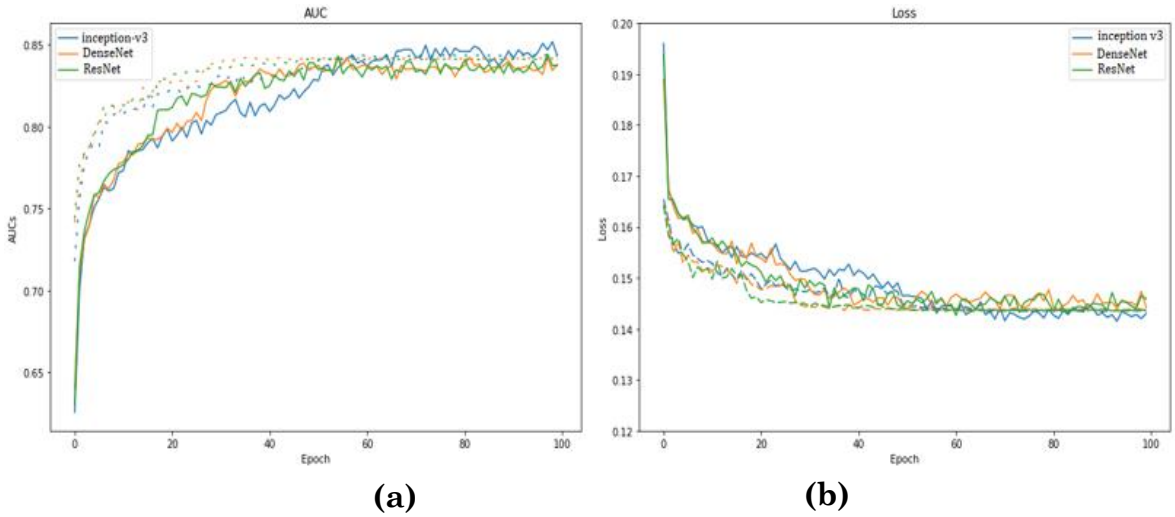
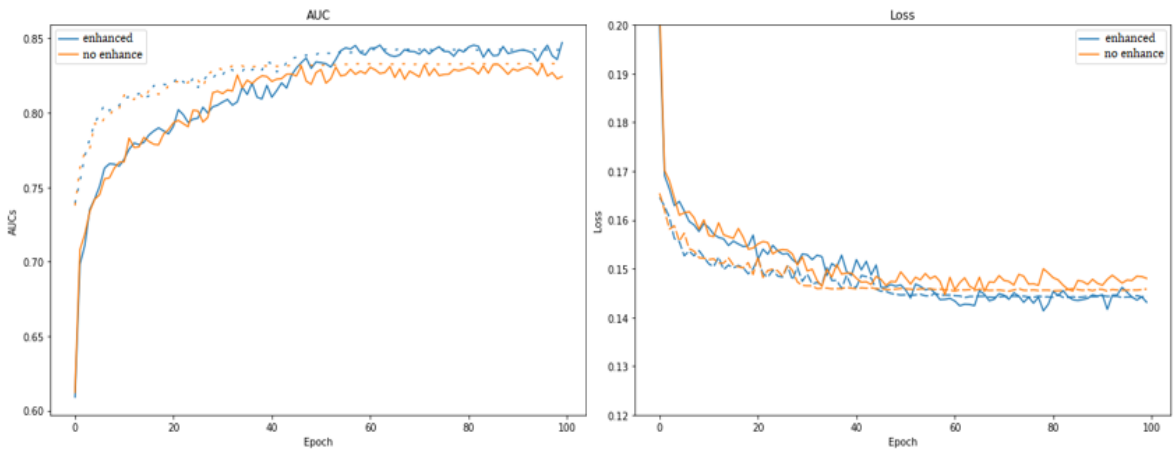


Figure 5. 7 Different CNN models evaluation

(a) Accuracy (Auc) vs epochs (b) loss vs epochs for different CNN models

5.2.3 Evaluating Preprocessing

In this section, a separate preliminary experiment was performed to demonstrate the importance of image pre-processing, before conducting CNN training, to enhance the classification accuracy of the model. First, this could be done by conducting experiment without applying image preprocessing techniques such as image resizing or reshape, augmentation, histogram equalization, pixel normalization. Second, the experiment was conducted using pre preprocessing techniques and we gate a performance enhancement in the accuracy (AUC) classification. In this experiment as a whole, we have observed that image reshaping plays important role in dimension reduction which take less time for training and reduce high memory consumption, and data augmentation play a significant role to balance the imbalance classes using data generator function.



(a) **(b)**

Figure 5. 8 Accuracy vs epoch and loss vs epoch

(a) Accuracy vs epochs (b) loss vs epochs using preprocessing and without preprocessing

5.2.4 Effect of Region Segmentation

The purpose of image segmentation is to extract the area or region of interest (ROI) in an image. Lung segmentation or finding region of interest (ROI) make to reduce search space and speed up tuning computation parameters. Thus, to get the most available region of interest we used U-Net lung segmentation technique which is used by many researchers [34] now a day and more efficient state of the art technique. The experiment that we have done with segmentation and without segmentation of lung image using U-Net segmentation technique is summarized in the table as shown below.

Table 5. 2 Comparison of Before and After Segmentation

Lung image	Dataset	Optimizer	Regularizer	Epoch	Learning rate (Lr)	Batch size	Training time(hrs)
Before segmentation	Chestx-ray14	Adam	Dropout (0.5)	100	0.0001	64	35.25hr
After segmentation	Chestx-ray14	Adam	Dropout (0.5)	100	0.0001	64	31.18hr

5.2.5 Classification Result

The classification result of each pathologies using the state of the art deep neural network model of Inception model is shown in the figure as follows.

Table 5.3 Classification Result

Chest Diseases	AUROC
Atelectasis	0.8283
Cardiomegaly	0.9177
Effusion	0.8852
Infiltration	0.7112
Mass	0.8659
Nodule	0.7886
Pneumonia	0.774
Pneumothorax	0.8875
Consolidation	0.8162
Edema	0.8971
Emphysema	0.9271
Fibrosis	0.8408
Pleural Thickening	0.7873
Hernia	0.9572
Average	0.8489

5.2.6 Localization Result of Abnormality Regions

Using the classification as pre-trained model for multi-label disease classification, we can calculate the disease heatmaps using the activations of the transition layer and the weights from the prediction layer, and even generate the attention guide line for each pathology candidate.

Multi-label classification framework produces the heatmap which indicates the approximate spatial location of particular thoracic disease class each time. This helps to visualize the most indicative areas used by the model during classification. The heatmaps are created by pooling along the channel dimension and then averaging the class-wise features. The heatmaps show that the network that is trained on this architecture generalizes well and demonstrates good interpretation ability in localizing the areas of interest. In this study, we used the gradient Class

Activation Map (Grad-CAM) method proposed by [39] to generate an interesting heatmap and to help the doctor assess the model performance. Grad-CAM is a CAM version that takes into account not only the weights but also the gradients going into the final convolution layer. Gradients have the advantage of being able to be applied to any layer of the network. Still, the last one is especially relevant for the localization of the parts of the image that contribute most to the final prediction. Furthermore, the layer used as input for the prediction can be followed by any module and not only by a fully connected layer. In Figure 5.9 we produce heat map of top-1 diseases to visualize the most indicative pathology areas on X-rays from evaluation subset.

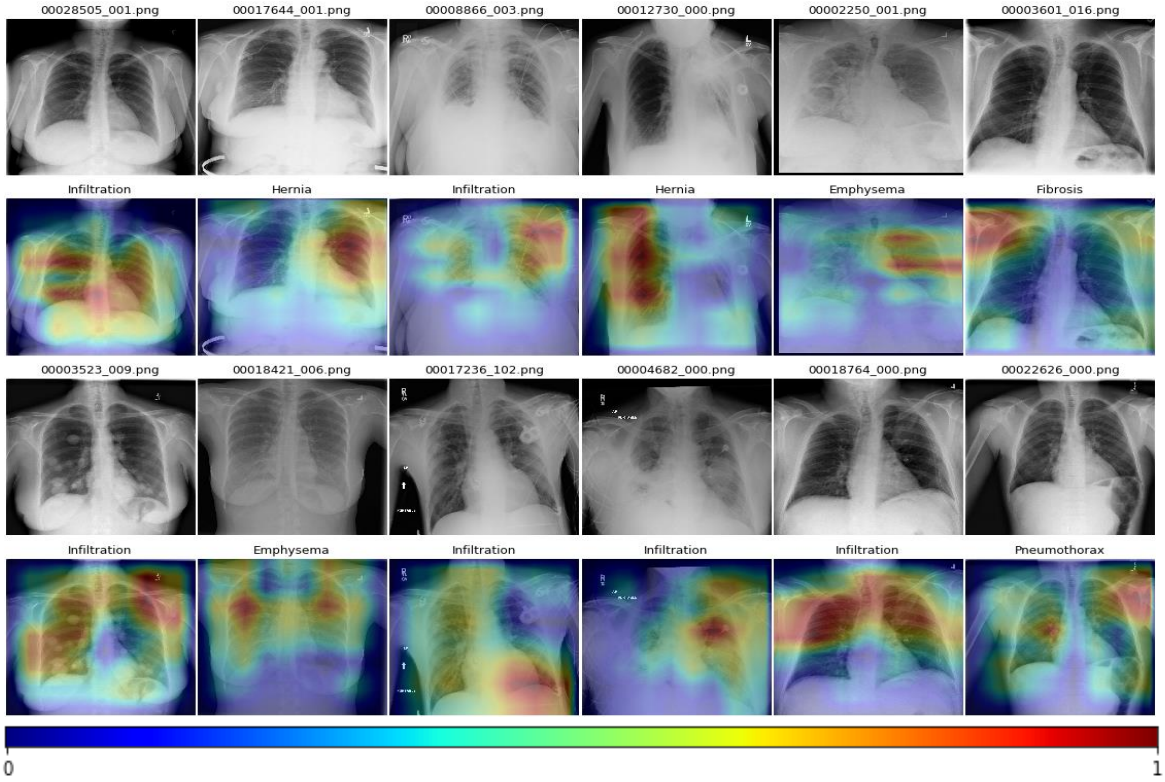


Figure 5.9 Localization (CAM) result of top-1 diseases

5.3 Discussions

In this paper we have proposed thorax diseases classification and localization using convolutional neural network. We have got interested result in accuracy, sensitivity and specificity, based on the results of this proposed system. Our model 's performance depends heavily on the size of the dataset, quality of the dataset, model parameters and algorithm architecture used. As we have seen from the test result, it indicates that inception v3 model are a promising method in thorax diseases classification and localization. The performance of the

proposed architecture can classify into four viewpoints. First, we perform pre-processing on our dataset, abnormalities can be effectively classified, due to the ability of this method for successful preprocessing and segmenting of abnormalities from the input data provided, and the ability of CNN model to automatically extract deep features of the input data from low-level features which reflect the relevant input information. Second, we compared the performance of different learning rate schedules and optimization algorithms. Third, we performed the performance examination using different convolutional layers and filter size. Fourth, we employed class activation map (CAM) to get the approximate location of each abnormality.

In this thesis instead of developing thorax abnormalities classification architecture as normal or abnormal, we carefully emphasize on multi-labeled classification i.e., each patient may have more than one diagnosis and may have another abnormality.

In terms of computational cost, each experiment was performed on Intel® core™ i5-6300HQ CPU@2.30, 2.30 GHz processor, GPU: Nvidia GeForce GTX 950M 4 GB dedicated graphics memory that had a RAM of 8 GB and 1TB HDD using the Kera's deep learning library and Python was used as the programming language.

Table 5. 4 Comparison of related resent studies

Author	Wang et al [22]	Rajpukar et al [23]	Hongyuwang [25]	Somnath et al [28]	Li et al [62]	Ours
Dataset	Chestx-ray8	Chestx-ray14	Chestx-ray14	Chestx-ray14	Chestx-ray14	Chestx-ray14
Method	AlexNet, GoogleNet, VGG&ResNet-50	ChexNet	ChestNet	ResNet-18	ResNet-V2-50	Iception-v3
Resources	GPUs :4 Titan X	N/A	N/A	**Ram**: 256 GB **GPU**: NVIDIA GeForce GTX 1070	GPU: 5 Nvidia P100	Ram: 8 GB GPU: 4 GB Nvidia GeForce 950
Atelectasis	0.733	0.716	0.821	0.832	0.800	0.8283
Cardiomegaly	0.856	0.807	0.905	0.915	0.870	0.9177
Consolidation	0.711	0.708	0.794	0.814	0.800	0.8162
Edema	0.806	0.835	0.893	0.899	0.880	0.8971
Effusion	0.806	0.784	0.883	0.889	0.870	0.8852
Emphysema	0.842	0.815	0.926	0.935	0.910	0.9271
Fibrosis	0.743	0.769	0.804	0.848	0.780	0.8408
Hernia	0.775	0.767	0.939	0.939	0.770	0.9572
Infiltration	0.695	0.609	0.735	0.714	0.700	0.7112
Mass	0.718	0.706	0.862	0.861	0.830	0.8659

Nodule	0.777	0.670	0.777	0.790	0.750	0.7886
No-finding	Nan	Nan	Nan	Nan	Nan	0.872
Pleural_thickening	0.724	0.708	0.814	0.791	0.790	0.7873
Pneumonia	0.684	0.633	0.763	0.779	0.670	0.7740
Pneumothorax	0.805	0.806	0.893	0.887	0.870	0.8875
Mean	0.761	0.738	0.841	0.849	0.881	0.8489

Table 5.4 shows the results of our proposed method compare with the existing works such as, Wang et al. [22]provided the dataset and a baseline model, although this method mainly focuses on Chest X-ray Database and Benchmarks on 8 Common Thorax Diseases, Rajpukar et al[23] used ChexNet algorithm for pneumonia detection and compare the radiologists and ChexNet algorithm on F1 score metrics. Hongyuwang [25] used the classification branch which is the ResNet-152 model for label prediction and the attention branch is used to exploit the correlation between class labels and the regions of pathological abnormalities, Somnath et al [28] used ResNet18 model for feature extraction and classification is performed using a fully connected network that give good performance of 0.850 mean AUC, but it has large memory consumption. Li et al[62] used ResNet as features extractor, use multi-instance learning assumption to predicts both labels and class-specific localizations. Trained on 5 Nvidia P100 high performance machine and it give best result. Although our model still needs some improvement to boost the overall accuracy, it is still comparable to the existing CNN architectures in the thoracic disease classification and localization tasks.

The Research questions that have been answered by this thesis are:

1. Does the Inception-v3 model we proposed better than the baseline approach of weakly supervised classification and localization of common thorax diseases?

2. what is the impact of lung segmentation to enhance the computational speed of the system?

3. Do the preprocessing techniques enhance classification performance of our approach?

Based on the experiments we have conducted our contributions with regard to these questions can be summarized as:

Inception model has multiple convolution filters with some pooling, that are applied to the same input, this allows the model to take advantage of multi-level feature extraction. For instance, it extracts general (5x5) and local (1x1) features at the same time. Using multiple features from

multiple filters improve the performance of the classification network. Convolutions, average pooling, max pooling, dropouts, and fully connected layers are among the symmetric and asymmetric building elements used in the model. Batch normalization is performed to activation inputs and is utilized extensively throughout the model. Inception model Work with Factorizing Convolutions. Factorizing Convolutions used to reduce the number of connections and parameters to learn. This will increase the speed and gives a good performance. As shown from the experiment in figure 5.7 inception model give better performance than DensNet and ResNet model.

Adam has low memory requirements and is invariant to the diagonal rescaling of the gradients. The Adam optimizer also has an efficient regret bound on the convergence rate. It is also well suited for problems involving large size of data. The figure 5.3 shows the comparation of Adaptive gradient-descending algorithms and adaptive momentum optimization (Adam) algorithm provide best result.

The other issue that we have been justified in this thesis is the effect of the preprocessing techniques on the performance of our model. This could be done by experimenting first without the application of image preprocessing, which means feeding the raw image files in their current form and, secondly, by using image preprocessing techniques such as resizing, normalization, histogram equalization and augmentation techniques to the chest x-ray images and feeding to our model. Even if our model (inception - v3 model) is using multiple filters to extract multiple features to improve the performance of the classification, still using preprocessing techniques add enhancement to our classification model as we seen in figure 5.8.

Chapter 6

Conclusions and Future Work

6.1 Conclusion

Automated thorax diseases classification and localization plays an important role in the diagnosis of many different lung and heart diseases. Due to the existing of a large number of clinical cases and the treatment given by radiologists is not yet supported by CAD methods for their diagnosis the classification and localization of multilabel pathology has become very time consuming, tedious process, not accurate and costly, reliability of the manual classification depends on knowledge and skill of the radiologists. Hence, it is very important to build a computer-aided detection (CAD) which can help the radiologists to make informed decisions more quickly. Our study aimed to create a prediction model that would be effective in identifying chest diseases and find the approximate location of the diseases from thoracic X-ray images.

The proposed Automated Thorax disease Classification and localization (ATDCL) has three main stages, the preprocessing stage, the thorax classification stage and the localization stage. In the preprocessing stage, image resizing, normalization, histogram equalization and median filtering was used to improve the quality and remove the noise of the input images. In the thorax classification stage, we used a novel deep learning model named inception model. The proposed model is then run on the Chest X-ray 14 datasets with 14 different lung and heart abnormal labels and one normal label. We have achieved the Average accuracy value (AUC) of 0.8489 and are also developing a Class Activation Map (CAM) that can significantly improve the understanding of radiologists about the approximate location of the existed pathology.

There are several concerns about the ChestX-ray-14 dataset. First, we found that labeling accuracy of the medical data is questionable, Since the medical dataset is the text-mind the visual labeling and the meaning of the label's failure. Second, some labels such as "consolidation" and "pneumonia" have a similar visual appearance that cannot be easily distinguished from a raw chest x-ray. Third, some of the labels ignored the clinically relevant sub-labels, such as a lethal pneumothorax and a treatable pneumothorax. This can lead the model to learn something unimportant or even dangerous.

The following general points can be noted from this work:

1. Preprocessing techniques are important techniques to make input data suit to the model and the model can easily learn the features from the input data.
2. It is also useful to reduce the learning rate while training deep neural networks as the training progresses. This can be accomplished by using pre-defined learning rates schedules or adaptive learning rate methods. Adaptive learning rate approaches show better performance than learning rate schedules and require far less effort in hyperparameter settings.
3. Another point is, deep learning requires a huge number of training data and it have numerous layers of learning algorithms each providing a different interpretation to the data it feeds on, which enables to learn complex features. However, a deep learning algorithm is an efficient algorithm for acquiring complex features, it has high memory consumption as hidden layer increase network parameter is also increase which need more memory consumption.so it is better to limit the layers as input data complexity and available memory space.

6.2 Future Work

We have implemented different techniques and built our CNN architecture for classification and localization of thorax abnormalities to achieve an encourageable result. However, we believe that incorporation of the following ideas as a future works, would achieve improved result:

- Implementing hybrid deep learning models such as UNet and ResNet together.
- Using powerful background and foreground segmentation algorithm because segmentation adds an extra layer of complexity to the system.
- Perform the experiment using different datasets like Chest X-ray 14 dataset vs Japanese Society of Radiological Technology (JSRT) dataset.

References

[1] A. Cruz-Roa, J. C. Caicedo, and F. A. González, "Visual pattern mining in histology image collections using bag of features," *Artificial intelligence in medicine,* vol. 52, no. 2, pp. 91-106, 2011.

[2] A. B. Aldasoro CCR, "Volumetric texture segmentation by discriminant feature selection and multiresolution classification," *IEEE Trans Medical Imaging,* no. 26, pp. 1-14, 2007.

[3] T. S. Lee NR, Andrew FL, "Interactive segmentation for geographic atrophy in retinal fundus images," *Conference Record/Asilomar Conference Signals,* 2008.

[4] M. A. Grau V, Alcaniz M, Ron K, Simon KW, "Improved watershed transform for medical image segmentation using prior information," *IEEE Transactions Medical Imaging* vol. 26, pp. 1-5, 2004.

[5] A. J. Mondal T, Sardana HK, "Automatic craniofacial structure detection on cephalometric images," *EEE Transactions on Image Processing,* vol. 20, pp. 2606-2614, 2011.

[6] A. K. Kubota T, Maneesh D, Marcos S, Arun K, "Segmentation of pulmonary nodules of various densities with morphological approaches and convexity models," *Medical image analysis,* vol. 15, pp. 133-154, 2011.

[7] J. M. Boykov YY, "Interactive graph cuts for optimal boundary & region segmentation of objects in ND images," in *Internation Conference on Computer Vision*, Vancouver, Canada, 2001.

[8] W. R. Russ JC, *The image processing handbook*. usa: CRC Press Addison-Wesley, 1993.

[9] B. K. A. Esteva, R. A. Novoa et al, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature,* vol. 542, pp. 115–118, 2017.

[10] S. O. G. E. Hinton, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation,* vol. 18, p. 7, 2006.

[11] I. S. A. Krizhevsky, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM,* vol. 60, p. 6, 2012.

[12] K. S. a. A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *ICLR*, Oxford, 2015.

[13] M. R. C. Barata, M. Francisco, T. Mendonça, and J. S. Marques, "Two systems for the detection of melanomas in dermoscopy images using texture and color features," *IEEE Systems Journal,* vol. 8, pp. 965–979, 2014.

[14] Z. D. Y. Zhang, A. Liu et al., "Magnetic resonance brain image classification via stationary wavelet transform and generalized eigenvalue proximal support vector machine," *Journal of Medical Imaging and Health Informati,* vol. 5, pp. 1395–1403, 2015.

[15] W. C. Q. Li, X. Wang, Y. Zhou, D. D. Feng, and M. Chen, "Medical image classification with convolutional neural," in *IEEE 13th International Conference on Control, Automation, Robotics & Vision*, Singapor, 2014.

[16] H. R. R. H.-C. Shin, M. Gao et al., "Deep convolutional neural networks for computer-aided detection: CNN architectures,dataset characteristics and transfer learning," *IEEE Transactions,* vol. 35, pp. 1285–1298, 2016.

[17] M. A. Binit Topiwala, Hari Prasad, "Detecting Thoracic Diseases from Chest X-Ray Images," 2019.

[18] C. J. a. J. D. R. Golan, "Lung nodule detection in CT images using deep convolutional neural networks," *IEEE,International Joint Conference on Neural Networks (IJCNN),* pp. 243-250, 2016.

[19] A. a. A. A. Tartar, "Ensemble learning approaches to classification of pulmonary nodules," in *International Conference on IEEE*, Malatya, Turkey 2016.

[20] W. M. Samrawit Solomon, "Diagnosis and Risk Factors of Advanced Cancers in Ethiopia," *Journal of Cancer Prevention,* vol. 24, pp. 161-172, 2019.

[21] W. L. Jackson, "Diagnostic Imaging," ed: In Radiology, Turnaround Time is King, 2019.

[22] Y. P. Xiaosong Wang, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, Ronald M. Summers. (2017, Dec 5) ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseas. *Department of Radiology and Imaging Sciences, Clinical Center*.

[23] J. I. Pranav Rajpurkar, "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning," *arXiv:1711.05225v3 [cs.CV],* Dec 25 2017.

[24] K. A. A. A. S. S. Sertan Kaymak, "Classification of Diseases on Chest X-Rays Using Deep Learning," *Springer Nature Switzerland,* vol. 1, pp. 516–523, 2019.

[25] Y. X. HongyuWang, "A Deep Neural Network for Classification of Thoracic Diseases on Chest Radiography," *Computer Vision and Pattern Recognition,* pp. 1-8, 2018.

[26] A. MohammadHavaei, DavidWarde-Farley,AntoineBiard,AaronCourville,YoshuaBengio,ChrisPal,Pierre-MarcJodoin,HugoLarochelle, "Brain tumor Segmentation with Deep Neural Networks," *Medical Image Analysis,* vol. 35, pp. 18-31, 2017.

[27] Z. L. a. H. Deng, "Medical Image Classification Based on Deep Features Extracted by Deep Model and Statistic Feature Fusion with Multilayer Perceptron," *Computational Intelligence and Neuroscience,* vol. 2018, p. 3, 2018.

[28] I. S. Somnath Rakshit, Michal Wlasnowolski,Ujjwal Maulik and Dariusz Plewczynski, "Deep Learning for Detection and Localization of Thoracic Diseases Using Chest X-Ray Imagery," *Springer Nature Switzerland,* pp. 271–282, 2019.

[29] R. M. a. H. Aggarwal, "A Comprehensive Review of Image Enhancement Techniques " *Journal of computing* vol. 2, no. 3, pp. 8-13, 2010.

[30] A. a. Sheeraz, "Artificial Neural Network Based Classification of Lungs Nodule Using Hybrid Features from Computerized Tomographic Images," *Applied Mathematics and Information Sciences* vol. 9, p. 183, 2015.

[31] S. P. a. T.Velmurugan, "Preprocessing by Contrast Enhancement Techniques for Medical Images," *International Journal of Pure and Applied Mathematics,* vol. 118, no. 18, pp. 681-3688, 2018.

[32] W. Atilaw, "Lung Nodules Detection from Computed Tomography Scans Using Deep Belief Networks," ed: AAU Institutional repository, 2018.

[33] N. Puneet, "Binarization Techniques used for Grey Scale Images," *nternational Journal of Computer Applications,* vol. 71, no. 1, pp. 8-11, 2013.

[34] P. F. Olaf Ronneberger, and Thomas Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," ArXiv150504597 Cs, Germany, 2015.

[35] Wikipedia, ed.

[36] T. I. Kazuto Ashizawa, Heber MacMahon, Carl J. Vyborny, Shigehiko Katsuragawa, Kunio Doi, "Artificial neural networks in Chest Radiography: Application to the Differential Diagnosis of Interstitial Lung Disease " *Academic Radiology,* vol. 11, no. 1, pp. pp. 29–37, 2005.

[37] R. H. A. a. a. M. K. S. Ma'aitah, "Deep Convolutional Neural Networks for Chest Diseases Detection," *Journal of Healthcare Engineering,* p. 1, 2018.

[38] L. Z. QingZeng Song, XingKe Luo, and XueChen Dou, "Using Deep Learning for Classification of Lung Nodules on Computed Tomography Images," *Journal of Healthcare Engineering,* vol. 2017, pp. 1-7, 2017.

[39] M. C. Ramprasaath R. Selvaraju, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, Dhruv Batra, "Grad-CAM: Visual Explanations from Deep Networksvia Gradient-based Localization," *arXiv:1610.02391cs.cv,* vol. 4, pp. 1-23, 2019.

[40] A. K. Rebecca Thornhill, "Texture Analysis: A Review of Neurologic MR Imaging Applications," *American Journal of Neuroradiology,* pp. 809-816, 2010.

[41] S. S. Prakash Hiremath, "Texture Classification using Wavelet Packet Decomposition," *Journal of GVIP* vol. 6, no. 2, pp. 77-80, 2006.

[42] I. Z. P. Karch, "An Experimental Comparison of Modern Methods of Segmentation," *8th IEEE International Symposium on Applied Machine Intelligence and Informatics,* pp. 247-252, 210.

[43] R. S. A. Ashish Thakur, "A Local Statistics Based Region Growing Segmentation Method for Ultrasound Medica Images," *World Academy of Science, Engineering and Technology, International Journal of Medical and Health Sciences,* vol. 1, no. 10, pp. 564-569, 2007.

[44] A. S. Nadeem Mahmood, Ahmad Waqas, Adamu Abubakar,Shafia Kamran, Syed Baqar Zaidi, "Image segmentation methods and edge detection:an application to knee joint articular cartilage edge," *Journal of Theoretical and Applied Information Technology,* vol. 71, no. 1, pp. 87-96, 2015.

[45] M. G. Nihad Mesanovic, Haris Huseinagic , Matija Males , Emir Skejic , Muamer Smajlovic, "Automatic CT Image Segmentation of the Lungs with Region Growing Algorithm " *18th International Conference on Systems, Signals and Image Processing,* pp. 395-400, 2011.

[46] D. B. Thomas M. Lehmann, Christian Thies, and Thomas Seidl, "Segmentation of medical images combining local, regional, global, and Segmentation of medical images combining local, regional, global, and hierarchical distances into a bottom-up region

merging scheme " *Medical Imaging International Society for Optics and Photonics,* pp. 546-555, 2005.

[47] S. Sadek, "Entropic Image Segmentation: A Fuzzy Approach Based on Tsallis Entropy," *International Journal of Computer Vision and Signal Processing,,* vol. 5, pp. 1-7, 2015.

[48] A. B. Asma Kerkeni, Antoine Manzanera, Mohamed Bedoui, "A coronary artery segmentation method based on multiscale analysis and region growing," *omputerized Medical Imagingand Graphics, ,* vol. 48, pp. 49-61, 2016.

[49] M.-P. J. Yuri Y. Boykov, "Interactive Graph Cuts for Optimal Boundary & Region Segmentation of Objects in N-D Images," in *Internation Conference on Computer Vision*, Canada, 2001.

[50] Q. K. Al-Shayea, "Artificial Neural Networks in Medical Diagnosis," *IJCSI International Journal of Computer Science Issues,* vol. 8, no. 2, pp. 150-154, 2011.

[51] M. D. Egmont P, Heinz H, "Image processing with neural networks - a review," *Pattern Recognition,* vol. 35, pp. 2279–2301, 2002.

[52] H. L. Mohammad Havaei, Philippe Poulin, Pierre-Marc Jodoin, "Within-Brain Classification for Brain Tumor Segmentation," in *Arxiv*, 2015.

[53] P. T. B. a. H. T. H. Trong-Ngoc Le, "Liver Tumor Segmentation from MR Images Using 3D Fast Marching Algorithm and Single Hidden Layer Feedforward Neural Network," *Hindawi Publishing Corporation,* vol. 2016, pp. 1-8, 2016.

[54] Y. H. Karima Elmasria, Xin Yang, Xianfang Sun, Rebecca Pettit,William Evans, "Automatic Detection and Quantification of Abdominal Aortic Calcification in Dual Energy X-Ray Absorptiometry," *20th International Conference on Knowledge Based and Intelligent Information and Engineering Systems,* vol. 96, pp. 101-1021, 2016.

[55] B. v. Ginneken, "Fifty years of computer analysis in chest imaging: rule-based,machine learning, deep learning," *Radiological Physics and Technology,* vol. 10, pp. 23-32, 2017.

[56] R. G. Dayong WangAditya Khosla, Humayun Irshad,Andrew H Beck, "Deep Learning for Identifying Metastatic Breast Cancer," *ArXiv e-prints,* pp. 1-6, 2016.

[57] B. K. Andre esteva, Roberto A. Novoa,Justin Ko,Susan M. Swetter,Helen M. blau and Sebastian Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature,* vol. 542, no. 02, pp. 115–118, 2017.

[58] S. L. Alvin Rajkomar, Andrew G. Taylor,Michael Blum, and John Mongan, "High-ThroughputClassification of Radiographs Using Deep Convolu-tional Neural Networks," *Journal of Digital Imaging,* vol. 30, pp. 95–101, 2017.

[59] K. R. Hoo-Chang Shin, Le Lu, Dina Demner-Fushman, Jianhua Yao, Ronald M Summers, "Learning to Read Chest X-Rays:Recurrent Neural Cascade Model for Automated Image Annotation," in *Conference on CVPR*, USA, 2016.

[60] S. J. Sema Candemir, Wilson Lin,Zhiyun Xue, Sameer Antani, George Thoma, "Automatic Heart Localization and Radiographic Index Computation in Chest X-rays," *SPIE Medical Imaging.International Society for Optics and Photonics,* vol. 9785, pp. 978517-1-978517-8, 2016.

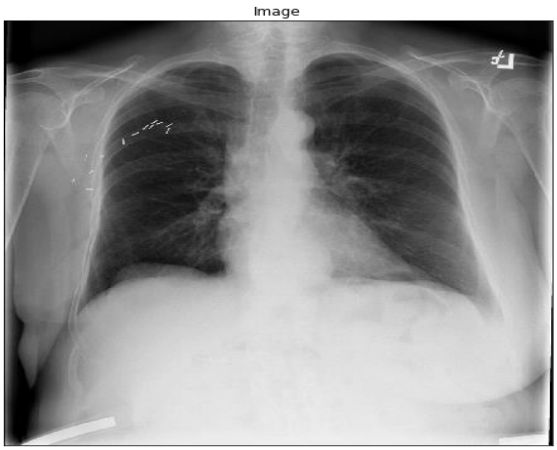
[61] A. M. G. Anil K. Bharodiya, "An Improved Segmentation Algorithm For X-Ray Images Based On Adaptive Thresholding Classification," *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH,* vol. 8, no. 9, 2019.

[62] Z. Li *et al.*, " Thoracic disease identification and localization with limited supervision. ," in *IEEE Conference on Computer Vision and Pattern Recognition*, USA, 2018.

Appendix A

A.1 Image Pre processing sample codes

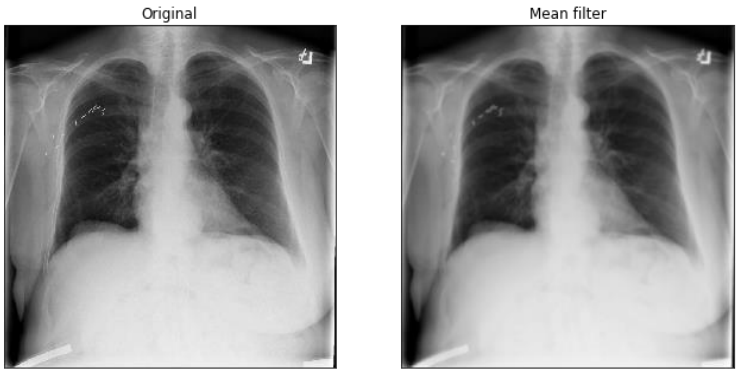
```
#First a few librabries will be imported.
import numpy as np
import cv2
import os
from matplotlib import pyplot as plt
from PIL import Image, ImageFilter
%matplotlib inline
image = cv2.imread('00000003_003.png')
image = cv2.cvtColor(image, cv2.COLOR_BGR2RGB)
plt.figure(figsize=(8,8))
plt.imshow(image)
plt.title('Image')
plt.xticks([])
plt.yticks([])
plt.show()
```




```
#Now a mean filter will be applied to a gray scale image.
# The image will first be converted to grayscale
image2 = cv2.cvtColor(image, cv2.COLOR_HSV2BGR)
image2 = cv2.cvtColor(image2, cv2.COLOR_BGR2GRAY)
figure_size = 9
new_image = cv2.blur(image2,(figure_size, figure_size))

plt.figure(figsize=(11,6))
plt.subplot(121), plt.imshow(image2, cmap='gray'),plt.title('Original')
```

```
plt.xticks([]), plt.yticks([])
plt.subplot(122), plt.imshow(new_image, cmap='gray'),plt.title('Mean filter')
plt.xticks([]), plt.yticks([])
plt.show()
```




```
#Gaussian Filter

#A 9 x 9 Gaussian filter (with sigma = 0) will be applied to the image.

new_image = cv2.GaussianBlur(image, (figure_size, figure_size),0)

plt.figure(figsize=(11,6))

lt.subplot(121), plt.imshow(cv2.cvtColor(image,
cv2.COLOR_HSV2RGB)),plt.title('Original')

plt.xticks([]), plt.yticks([])

plt.subplot(122), plt.imshow(cv2.cvtColor(new_image,
cv2.COLOR_HSV2RGB)),plt.title('Gaussian Filter')

plt.xticks([]), plt.yticks([])

plt.show()
```