Comparative Analysis of Detection and Classification of Lungs Cancer through Deep Neural Networks

MASTER OF SCIENCE
IN
COMPUTER SCIENCE

Submitted by:
Amna Islam
L1F18MSCS0019

DEPARTMENT OF COMPUTER SCIENCE
FACULTYOF INFORMATION TECHNOLOGY
UNIVERSITY OF CENTRAL PUNJAB

August 2020

Comparative Analysis of Detection and Classification of Lungs Cancer through Deep Neural Networks

A Thesis submitted in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE
IN
COMPUTER SCIENCE

Submitted By:
Amna Islam
L1F18MSCS0019

DEPARTMENT OF COMPUTER SCIENCE
FACULTYOF INFORMATION TECHNOLOGY
UNIVERSITY OF CENTRAL PUNJAB

August 2020

ABSTRACT

This study holds a comparative analysis of various techniques implemented with various models in order to detect and classify lungs cancer. This study has been conducted in order to provide a comprehensive view of performances among different techniques and models of CNN. The objective of this study was to explore the potential of CNN as feature extractor as well as classifier. In addition to this, it was also aimed to include Radiomics module to compare its results with deep learning techniques. The study was divided in three experiments. Each experiment was further divided in four modules. The modules include pre-processing, feature extractor, classification and ensemble. For each feature extractor module of each experiment, a separate technique and model of CNN are implemented. These include two deep learning methods and Radiomics. The features extracted from these feature extractor modules are then sent into a classification module. The classification module is same for all the experiments and is a combination of two deep learning techniques and models and one standard machine learning classifier. Three results were obtained for each experiment. Ensemble technique was applied on these results to bring stability and robustness in the model. It was observed that Radiomics should be included, in addition to deep learning techniques for feature extractor as it gives quantitative features specific to ROI, if annotations and segmentations are provided. Deep learning performs exceptionally well even when there are no annotations and segmentations provided. It can be used for extracting qualitative features. SVM gives better and stable results in classification than deep learning. Ensemble removes variance and brings stability in a model. It gives more reliable results. Hence, any model which include deep learning and Radiomics as feature extractor, SVM as classifier and then Ensemble applied on it would have likely chance to provide best and stable results.

DEDICATION

To my late father, Islam ul Haq, who always wished and dreamt for me to excel in academia with excellence, but never could witness my success and endeavors. To my late mother, Dr. Azmat Sultana, whom I lost during the course of this degree, who is the real role model of my life and without whom it would not be possible to get higher education and resilience to pursue research in the most tough time of my life.

In addition to this, I want to thank my sister, Maria Islam, whose help, support, kindness and humor have been a huge help in tough routine to complete my work. I want to thank my pet, Lia, as well for all the stress releasing fun she could provide.

I want to include a note of special gratitude to my Aunt, Dr. Tehmina Afridi, whose unconditional care, love and motivational support has helped me through the highs and lows of this journey.

I want to say a huge thank you to my friends and supports system Abida Bashir, Raiha Akram, Dr. Humaira Ahmad Khan, Ijaz Ali, Aria Khan, Hira Hanif, and last but not least Amina Tanveer.

I would also like to thank my mentors and friends, Dr. Sadia Zulfiqar and Jessica Marie Butnariu, who have inspired me to work hard and excel in academia and research. I want to thank my teachers from all the institutes I have studied.

Thank you to all who prayed, wished me best and helped me during the course of my MS degree.

ACKNOWLEDGEMENTS

I would like to give my special thanks my supervisor, Dr. Shahzad Majeed Tiwana, for his continuous and pedagogic guidance and support throughout the course of my thesis. His constant efforts and motivation for this work has been a great gesture. I would also like to express a note of special thanks to my internal evaluators Dr. Adnan Qureshi and Dr. M. Umair. Without their guidance and critique through the seminars and presentations, I would not have got a clear direction to conduct this study. I am also thankful to my teacher, Dr. Atif Mehdi for inspiring me to dedicate to research and hard work.

I am grateful to Dr. Haider Ali to whom I could present my work for initial stage and who helped me refine my technical approach towards my thesis. I am also very grateful to Dr. Rahul Paul who helped me develop a basic understanding of Bioinformatics, in general and Radiomics, in particular. I pay a very special thanks to Dr. Andrey Fedorov, Assistant Professor at Harvard Medical School, for his quick replies to my queries and number of technical components explained for my research.

I am grateful to the PyRadiomics community for its in depth implementation and guidance for using PyRadiomics library. I am also very much thankful to 3D Slicer Community for guiding me with solution to my problems in using the 3D Slicer software.

In addition, I would like to thank Sheharyar Arif and Sana Khalid for providing me technical suggestions and support through their expertise and knowledge.

DECLARATION

I, Amna Islam, D/O Islam ul Haq, a student of “MS Computer Science”, at “Faculty of Information Technology”, University of Central Punjab, hereby declare that this thesis titled, “Comparative Analysis of Detection and Classification of Lungs Cancer through Deep Neural Networks”

is my own research work and has not been submitted, published, or printed elsewhere in Pakistan or abroad. Additionally, I will not use this thesis for obtaining any degree other than the one stated above. I fully understand that if my statement is found to be incorrect at any stage, including after the award of the degree, the University has the right to revoke my MS/M.Phil. degree.

Signature of Student: ______________________________

Name of Student: ______________________________

Registration Number: ______________________________

Date: ______________________________

PLAGIARISM UNDERTAKING

I solemnly declare that the research work presented in this thesis titled, "Comparative Analysis of Detection and Classification of Lungs Cancer through Deep Neural Networks" is solely my research work, and that the entire thesis has been completed by me, with no significant contribution from any other person or institution. Any small contribution, wherever taken, has been duly acknowledged.

I understand the zero tolerance policy of the HEC and University of Central Punjab towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarized and that every material used from other sources has been properly acknowledged, cited, and referenced.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis, even after the award of MS/MPhil. degree, the University reserves the right to revoke my degree, and that HEC and the University have the right to publish my name on the HEC/University website for submitting a plagiarized thesis.

Signature of Student: ______________________________

Name of Student: ______________________________

CERTIFICATE OF RESEARCH COMPLETION

It is certified that this thesis titled, “Comparative Analysis of Detection and Classification of Lungs Cancer through Deep Neural Networks”, submitted by Amna Islam, Registration No. L1F18MSCS0019, for MS/M.Phil. degree at “Faculty of Information Technology”, University of Central Punjab, is an original research work and contains satisfactory material to be eligible for evaluation by the Examiner(s) for the award of the above stated degree.

Dr. Shahzad Majeed Tiwana
Assistant Professor
FOIT
University of Central Punjab

____

Signature

Co-Supervisor’s Name (If applicable)

Designation

Faculty

University Name

____

Signature

Date: ____________________

CERTIFICATE OF EXAMINERS

It is certified that the research work contained in this thesis titled "Comparative Analysis of Detection and Classification of Lungs Cancer through Deep Neural Networks." is up to the mark for the award of "Masters of Science in Computer Science".

Internal Examiner

Signature:

Name: ________________________

Date: __________________________

External Examiner

Signature:

Name:______________________

Date:_______________________

Dean
Faculty Name
University of Central Punjab

Signature:

Name: ________________________

Date: __________________________

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

LIST OF GRAPHS

LIST OF ABBREVIATIONS AND ACRONYMS

AUC	Area Under Curve
CNN	Convolutional Neural Networks
CT	Computed Tomography
GLCM	Gray-Level Co-Occurrence Matrix
LUNA16	LUng Nodule Annalysis-16
LDCT	Low Dose computed Tomography
LIDC/IDRI	Lung Image Database Consortium/ Image Database Resource Innitiative
NLST	National Lung Screening Trial
NRRD	Newly raw Raster Data
ResNet	RESidual NETwork
SVM	Support Vector Machine
TL	Transfer Learning
TPR	True Positive Rate
XML	Extensible Markup Language

CHAPTER ONE: INTRODUCTION

In the last decade, artificial intelligence has grown tremendously with number of applications in various fields. The aspects of machine learning, deep learning and computer vision has brought remarkable changes and advancements in the course of events. Medical imaging is one of the fields which has got the most attention and benefit from the stability of deep learning. From medical imaging, lungs cancer is among those complex problems which need technologically advanced solutions for its early detection.

1.0 Detection and Diagnosis of Lungs Cancer

1.1 *Subject Importance*

Lungs Cancer is the foremost reason of death, worldwide. It reports to have the highest number of incidence rate as well as mortality rate among all types of cancers. The patients do not exhibit any symptoms prior to the critical stage. Therefore, it is diagnosed only at a later stage. A better approach to detect lungs cancer on early stage will help in preventing the spread of disease. Low-dose Computed Tomography (LDCT) is known to be the best modality of screening the lungs cancer. The rate at which medical imaging is growing is high. A high number of CT scans are being produced for examination by relatively less number of doctors, radiologists and experts.

1.2 Role of Deep Learning

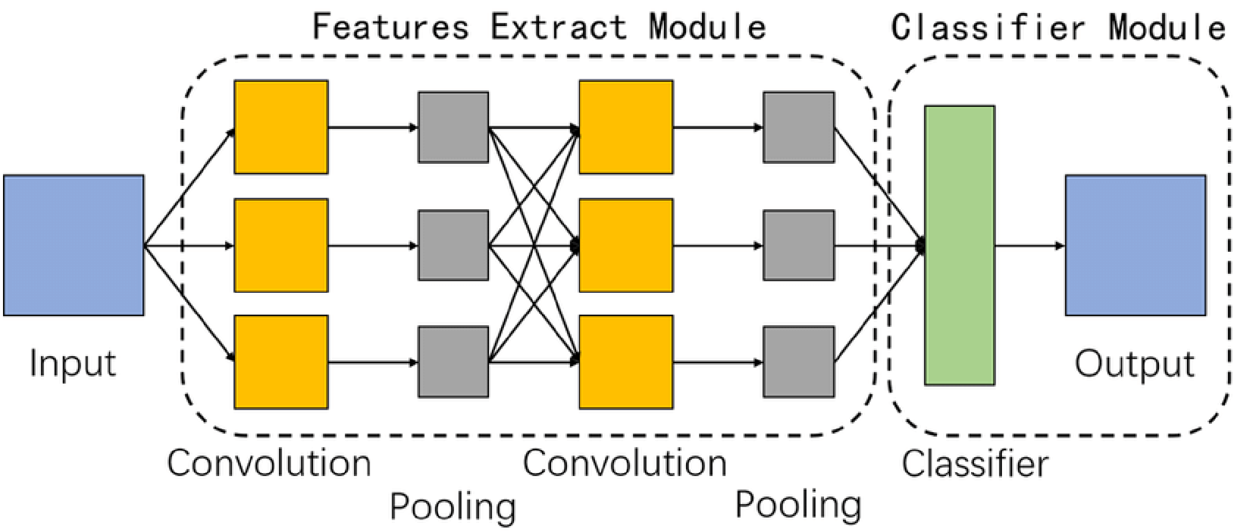
Advances in computer science and particularly, in artificial intelligence are providing new opportunities and solutions to the problems of all domains. Healthcare is also one of those domains which are being benefitted from the advantages of artificial intelligence and machine learning. The improvements and recent trends of deep learning have successfully helped out medical imaging in various aspects. Clinical Decision Support Systems have been revolutionized from qualitative understanding of medical imaging to quantitative mode of research and understanding. Moreover, sophisticated systems and models are still required for the early detection of pulmonary nodules and their classification.

1.2.1 Role of Convolutional Neural Networks

Convolutional Neural Network (CNN) is one of the most famous deep learning technique for images. It has also contributed remarkably in medical imaging. The structure of a CNN consists of stacks of layers. There are two modules of a Convolutional Neural Network:

1- Feature Extraction Module

2- Classifier Module



In solving the problem of lungs cancer with deep learning, the Feature Extraction Module and Classifier module of CNN will map as follows, respectively:

1- Computer Aided Detection (CAD*e*)

2- Computer Aided Diagnosis (CAD*x*)

There are more than one models and techniques, in which CNN can be implemented for the detection and diagnosis of lungs cancer. There is room to conduct a comprehensive comparative study using different models and techniques of CNN among the two modules of CNN itself. Many combinations have been implemented, yet a few more can also be executed. The previous studies and researches have not used different techniques and models of CNN in comparison. This study includes those combinations which have not been applied yet. Moreover, it is a thorough study which experimented with different CNN in different capacities.

1.3 Introduction to Radiomics

In precision oncology, Radiomics is playing a major role. Radiomics is the process of extracting quantitative features from radio-graphical images. The Radiomics features include such characteristics of the disease which are impossible to be detected and calculated by naked eye. Though, field of Radiomics include the many of the same features as hand-crafted features, yet, it provides high throughput and more number of features. Advances in medical imaging and breakthrough in the field of Radiomics has opened the gateway for collaboration between precision oncology, computer science and artificial intelligence. The high dimensional data produced from Radiomics invite number of researches to be conducted with respect to machine learning and data science.

Radiomics include statistical approaches and computational techniques to extract quantitative features from different radio-graphical imaging techniques which include Computer Tomography (CT) imaging, Positron Emission Tomography (PET) and Magnetic Resonance Imaging (MRI). There are mainly four categories of Radiomic features:

1- Shape
2- Intensity
3- Texture
4- Volume

Features of tumor or Region of Interest (ROI) are extracted from various statistical methods including which includes first-order statistics, second-order statistics and higher order statistics. Hence, Radiomics is a very helpful approach and advanced technique in precision oncology and deep learning for making the early detection of lungs cancer possible. Recent developments and implementations in this fields have automated the feature extraction process, while increasing number of features extracted. It is recommended by experts and researchers as well to include Radiomics in the studies for early detection of lungs cancer through deep learning.

1.4 Overview

In this study, there are three modules of experiments. Two modules of experiments are implementation of different techniques and models of CNN. both feature extractor module and classifier module of CNN. The third module consists of a combination of Radiomics and deep learning. Each experiment is further divided in four phases as below:

1- Preprocessing

2- Feature Extraction

3- Classification

Ensemble Averaging*1.5 Aims and Objectives*

It is a comprehensive study which has explored capabilities of different models and techniques of CNN in combination with Radiomics.

1.5.1 Aims

The aim of doing this research is to improve the incidence of early detection thus reducing the mortality rate caused by Lung Cancer through the latest advancements in computer science, in general and machine learning, in particular. It is also aimed to ease the burden of doctors by improving and automating the process of lung cancer detection. Moreover, it is an intention to integrate Radiomics as well.

1.5.2 Presided Objectives of the Research

1- Comparative Analysis of early detection of Pulmonary Nodules through feature extraction.

2- Comparative Analysis of the prediction of pulmonary nodule through classification.

3- Comparing Radiomics Model with deep learning models.

CHAPTER TWO: LITERATURE REVIEW

This chapter includes a thorough review of literature and covers all the aspects of work done related to this study. It includes a background section which highlights the research done about the problem statement and possible solutions of it with techniques and approaches used in this study. This chapter also includes the work done with respect to techniques as well as models used. It covers a complete section of background and literature review on Radiomics technique. In the last, this chapter sums up the research gap.

2.0 Literature Review

2.1 Background

In 2018, the incidence rate and mortality rate of lung cancer were reported to be 11.6% and 18.4%, respectively [1] The reason for these facts is that lungs cancer is usually diagnosed at a very later stage [2]. The 5-year-survival rate is currently very low i.e. 18% and with early detection, this can be lifted to 54% [2],[3]. Worldwide, it is the leading cause of death among all cancers. In USA alone, 228,820 new cases, among both the genders, were reported in 2020 [4]. The incidence rate is high in USA and UK whereas in China, Cancer mortality is quite high. "An estimated 4.3 million new cancer cases and 2.9 million new cancer deaths occurred in China in 2018" [5]. It has been reported that the number of lung cancer cases are not detected efficiently because the doctors are overburdened with professional commitments. To reduce the burden of doctors, the automated detection of pulmonary nodules can improve the diagnosis by a prominent percentage. "The heavy workload of Radiologists results in erroneous diagnosis i.e. 7 - 15% for 20 CT scans per radiologists per day. They have to see one scan every 3-4 second in 8-hour window to meet their workload requirements" [6].

2.2 Work Done in Deep Learning

Computer Aided Diagnosis CAD in Medical imaging is divided in two categories:

1- Computer Aided Detection (CADe) for detecting the lesion. Detection is identifying the hidden lesion through medical imaging (CT, PET CT, MRI, X-RAY).

2- Computer Aided Diagnosis (CADx) for diagnosis of the lesion. Diagnosis is about deciding whether the lesion is benign or malignant [7].

In Machine learning, Computer Aided Detection (CADe) is mapped with Feature Extraction from the CT images. Whereas, Computer Aided Diagnosis (CADx) is mapped with Classification.

Feature Extraction is done in two ways:

1- Hand Crafted Feature Extraction.

2- Data Driven Feature Extraction.

Hand Crafted Feature Extraction techniques are those which are carved out with the help of experts in terms of mathematical models. A few examples of Hand Crafted feature extractors are Histogram of Oriented Gradients (HOG), [8] Gabor, Local Binary Pattern (LBP), [9] and Gray Level Occurrence Matrix (GLCM) [10]. Among Data Driven feature extractor CNN is the best example. For a long time, feature extraction has been done using hand-crafted techniques when deep learning was not popular. However, with the publicly available datasets and advancements in deep learning, focus has been shifted towards machine learning now.

2.2.1 Overview of CNN

Among deep learning, CNN has been used in the detection of lung cancer more than any other neural network. Using 3D DenseNet, Region Proposal Network and 3D U-Net, Qin et al. designed a system which could detect Pulmonary nodules through CT images automatically [11], [12]. The system used LUNA16 dataset and came out with sensitivity of 96.7%. For catering, ten times more number of data than a normal CNN can manage, a system based on 3D Geometric CNN (G-CNN) was proposed by Winkles et al. in 2018 [13].

2.2.2 Literature Review of CNN Models

Nobrega at el used 11 architectures of CNN to extract features of CT images and then used the results of each model with 5 classifiers to differentiate between malignant and benign nodules. The analysis of results declared ResNet50 model with classifier SVM RBF to be the best in terms of TPR. This combination also gave the highest Area under Cover (AUC). They considered True Positive Rate to be the most important evaluation metric by saying that it is more important address the issue of leaving malignant nodule undetected. The highest TPR was 85.64%. However, the highest accuracy was 89.91% and it was achieved with Xception model of CNN and Multilayer Perceptron classifier. [14]

Agile CNN was used by Zhao et al for both feature extraction and classification. The Agile CNN is mainly based on minimum number of convolutional layers. The used hybrid CNN of LeNet and ALexNet in a way that layer settings of LeNet and parameter settings of AlexNet were combined. They used LIDC/IDRI dataset and calculated the accuracy to be 82.2% and AUC to be 87.7%. The CNN learnt its own feature instead of using hand crafted feature extractor [15].

Nishio et al conducted a very interesting and important research. They used Hand Crafted feature extractor, Local Binary Pattern for feature extraction from CT images of Lungs. They used database of Kyoto University Hospital, Japan which had information of 1240 patients. The size of voxel they used was 1x1x1. 3D images were converted to 2D images. Three sizes of images were decided to be used for DCNN. The lengths of sizes were 56, 112 and 224. The model used for DCNN was Vgg-16. DCNN were implemented with and without Transfer Learning from LBP. The feature extraction form LBP was separately classified with SVM as well and gave Accuracy of 55.9%. DCNN with Transfer Learning gave best Accuracy of 68.0% and DCNN without Transfer Learning gave best Accuracy of 62.4% [15].The length of image size which gave the best accuracy was 224. The reason is that Vgg-16 is originally trained on this size. They concluded that it is important to have a valid and good technique for pre-processing because it affects the results significantly.

2.2.3 Literature on the Methods and Techniques

There are six categories in the methodology of CNN for using it in lung cancer research.

a- Advanced Off-The-Shelf CNN

b- CNN with Advanced Implementation

c- CNN+

d- Hybrid Systems

e- Ensemble Learners of Multiple CNNs

f- Transfer Learning Based Systems

For the detection of pulmonary nodules and reduction in false positive cases, four methodologies a, b, c, d among the above mentioned are used. The ratio of number of publications on these methodologies in 2018 has been 4:2:2:3.

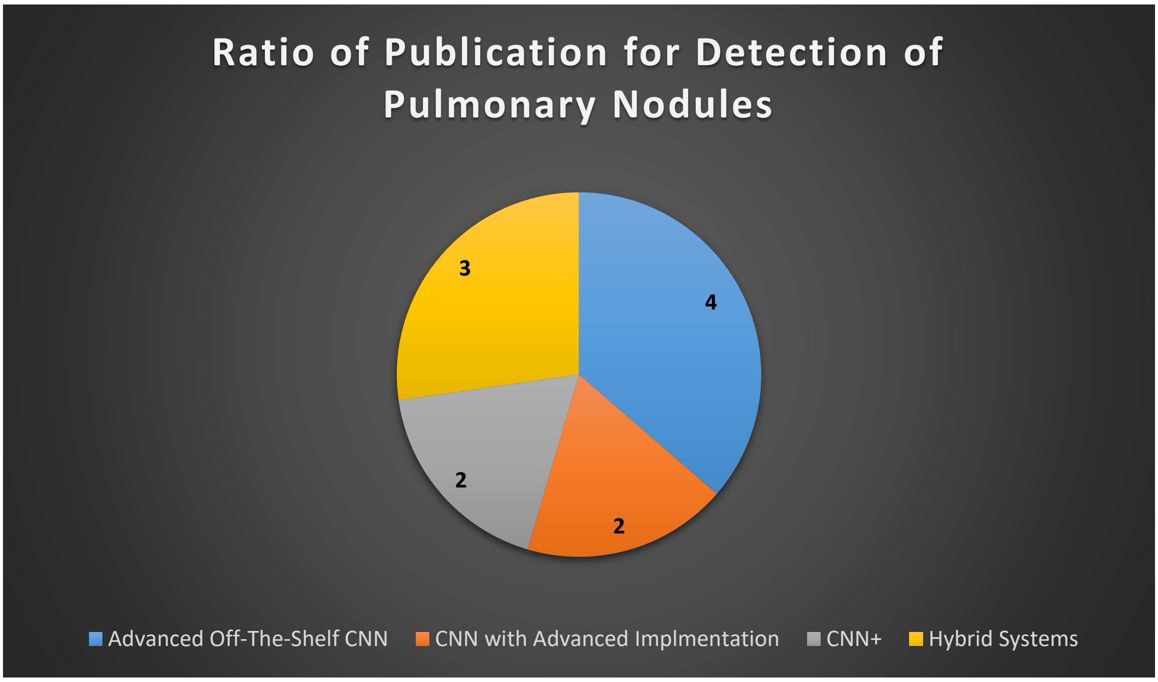


Figure 2: Ratio of Methodologies used in Pulmonary Nodule Detection

Whereas, for the classification of pulmonary nodules, five methodologies b, c, d, e, and f are being used and their respective ratio of number of publication is 6:3:1:1:3 [3].

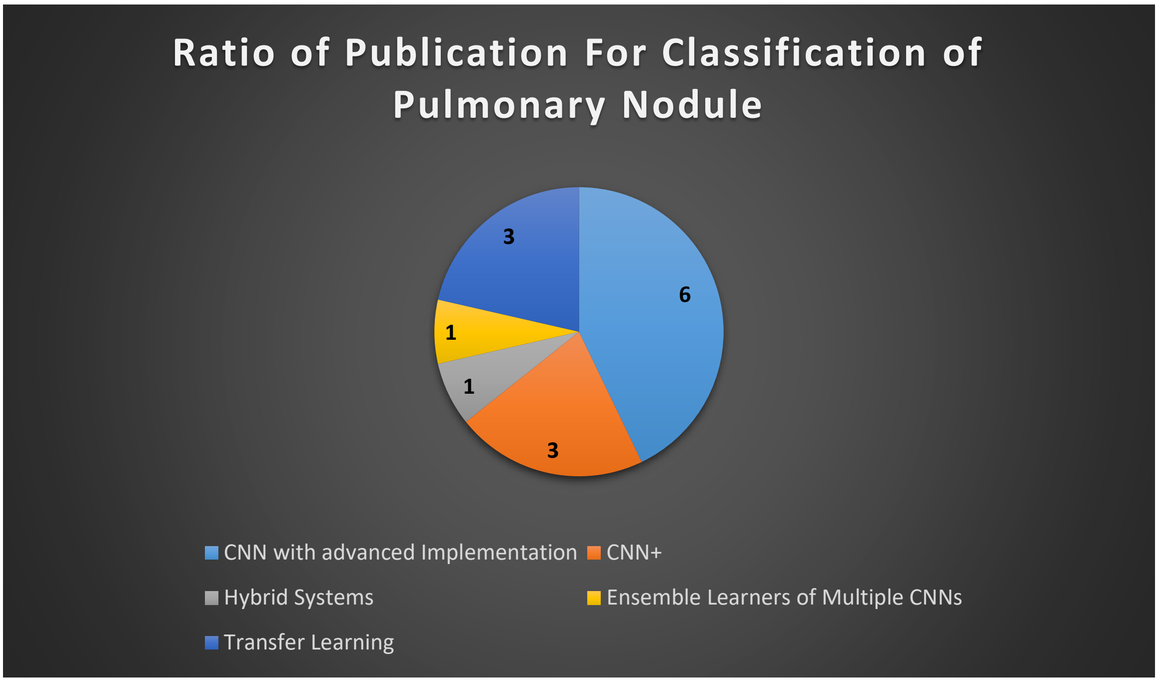


Figure 3: Ratio of Methodologies used in Pulmonary Nodule Classification

For classification, Advanced Off-The-Shelf CNNs were the first to produce results but as the methodology grew rich, trends were shifted to other techniques. The most number of publications in 2018 for the early detection of lung cancer are also being done using this technique. Using this technique, models like region proposal network (RPN), reinforcement learning and Faster Region based CNN were designed for the detection of lung nodule. In contrast, the most number of publication for the classification of benign and malignant lung nodule is done using the CNNs with advanced implementation techniques [16], [17].

CNN has been used with advanced techniques for the detection of Lung Cancer. The advanced techniques include the CNN+ which further has two categories as 2D CNN+ and 2D U-Net+. CNN+ is referred to the system which includes normal CNN and any other technique of image processing. Using 2D CNN+ with linear interpolation, icosahedron based normalization and thresholding, Liu at el. designed a very novel system. They used above mentioned techniques for "preprocessing of raw CT scans, volume of interest detection and generation and selection of views at different scales" [18]. In addition, Hybrid CNN has been very famous for detection and classification of pulmonary nodules as well. It is observed from survey paper of Monkam et al that while using CNN with Advanced Implementation and CNN+, same models of CNN were used in feature Extraction as well as classification. Whereas Hybrid CNNs used two different models of CNNs and gave maximum of 85.0% Accuracy [19] (PATRICE MONKAM 1, 2019). For Detection, Transfer Learning is the new and recent trend and is proven to be very optimal in results as well. Xie et al used Hand crafted features for feature Extraction such as Overall Appearance (OA), Heterogeneity in Voxel Values (HVV) and Heterogeneity in Shapes (HS). ResNet50 model of CNN was learnt with these features through Transfer Learning and as well as Classification. Ensemble Learning was applied on the results of ResNet-50 from classification part. The results of their model were compared with four other models of different scientists who used hand crafted feature but did not used Ensemble Learning. Their model produced the best accuracy which was 93.72% which proved that Ensemble Learning is giving the best results so far for classification and Transfer Learning, for Detection [20].

"Ensemble Learning is an approach to combine several of the same or different types of learned models into one predictive model to enhance prediction performance" [21]. "Ensemble makes the predictive model robust, more stable and less noisy" [22].

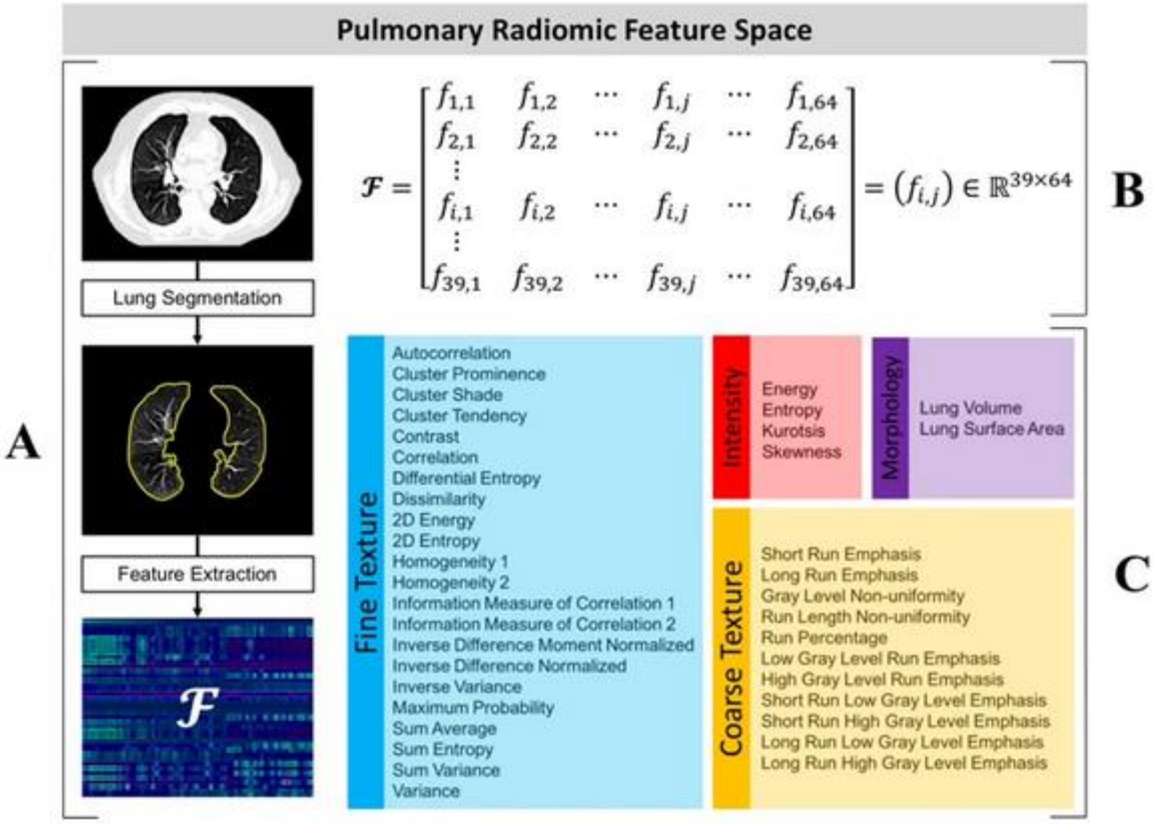
A study using Deep Convolutional Neural Network (DCNN) with and without transfer learning and using conventional (hand-crafted) methods of feature extraction was carried out. The analysis led to the conclusion that DCNN with transfer learning gave the best result among the three mentioned techniques in order to classify between benign and malignant nodules. Along with transfer learning, image augmentation is also used to address the problem of less data. Image Augmentation is used by Rahul Paul at el where they designed three different CNN architectures to carry out ensemble classification through these architectures. The best Area under Curve (AUC) was calculated to be 0.96 which was the best among three subsets [23]. The accuracy of

86.91% was achieved through Ensemble Learning. The research in the field of Lung Cancer has faced many challenges among which few are very prominent. These include insufficient data, absence of radiological features and limited approaches. Francisco Azuaje made a comprehensive study on these challenges and also provided prospective solution. The solutions include introduction of surrogate dataset with the combination of transfer learning. The paper suggests to train the models with different learning parameters. “The resulting models were collectively evaluated, and the highest performing model was selected for further testing” [24]. Surrogate datasets usually have noise more than the original ones. Such surrogate datasets are used for the evaluation of the models which have already been trained.

2.3 Work Done in Radiomics

The solution to the absence of radiological features was provided by Rahul Paul at el. They used Radiomics for feature extraction. The processes of Radiomics are mainly used to extract both qualitative and quantitative information from clinical data and images. “The Radiomic process can be divided into distinct steps with definable inputs and outputs, such as image acquisition and reconstruction, image segmentation, features extraction and qualification, analysis, and model building” [25]. These features were classified with random Forests and gave an Accuracy of 76.79% [24]. Three CNN architectures were made separately and transfer learning from Radiomics was applied on each of these architectures. They used NLST Dataset and 70% was used for training and 30% for testing. CNN Architecture 3 had a cascading design where right and left branches were merged in the mid branch. This architecture gave 81% accuracy. Kyle J. Lafata has conducted a detailed study on Radiomics and how can these quantitative features help in Pulmonary Nodule detection and classification. He classified Radiomic features in four categories. The total number of features were 39 [26]. In contrast, Rahul Paul fetched 219

Radiomic Feature in his study. However, that was because he took experts in the loop and resources from Moffit Cancer Center [24].



$$\mathcal{F} = \begin{bmatrix} f_{1,1} & f_{1,2} & \cdots & f_{1,j} & \cdots & f_{1,64} \\ f_{2,1} & f_{2,2} & \cdots & f_{2,j} & \cdots & f_{2,64} \\ \vdots & & & & & \\ f_{i,1} & f_{i,2} & \cdots & f_{i,j} & \cdots & f_{i,64} \\ \vdots & & & & & \\ f_{39,1} & f_{39,2} & \cdots & f_{39,j} & \cdots & f_{39,64} \end{bmatrix} = (f_{i,j}) \in \mathbb{R}^{39\times 64}$$

Figure 4: Classification of Radiomic Features

2.4 Research Gap

For the detection of lung cancer, mostly used architectures are AlexNet and VGG [19], whereas there are other efficient models as well. The feature extraction is mostly done by hand crafted, whereas, Radiomics can give specific feature related to the nodule and tumor from radio-graphical images. If CNN is used for feature extraction, same model of CNN is used for feature extraction and classification. Secondly, if CNN is used for feature extraction, classifiers like SVM, Random Forests and MLP are used for classification. Hand Crafted Method are used with transfer learning applied on CNN. Hand crafted Methods lacks accuracy and CNNs with transfer learning from Hand Crafted Method could not produce high results in accuracy. The accuracy of hand crafted feature was seen to be 55.9% and through Radiomics, it was 76.97%. Transfer Learning from Hand Crafted features on CNN Vgg-16 gave accuracy of 68.0% and from Radiomics, it was 75.0%. Hence, Radiomics should be used for feature extraction as well as transfer learning for detection. [24], [7], [27]. Different models used for detection and classification enhances the Accuracy, TPR and AUC as well. With hybrid and agile CNN, the accuracy was calculated to be 85.64% whereas accuracy with same models was 62.4% [15]. Ensemble Learning is enhancing the over prediction results and Producing Accuracy up to 89.02% [24], Ensemble Learning also have the least number of publication [19].

CHAPTER THREE: RESEARCH DESIGN

This chapter includes the detailed research design, methodology used and methods selected for this study. It includes the selection of dataset, techniques and models used for experiments. It also describes the importance and reason behind choosing this research design and ensures the validity and reliability of the research design of this study.

3.0 Research Design, Methodology and Methods

3.1 Overview

This study is an inclusive comparative analysis comprising three different sets of experiments. Each set of experiment includes four phases as follows:

1- Pre-processing
2- Feature Extraction
3- Classification
4- Ensemble Learning

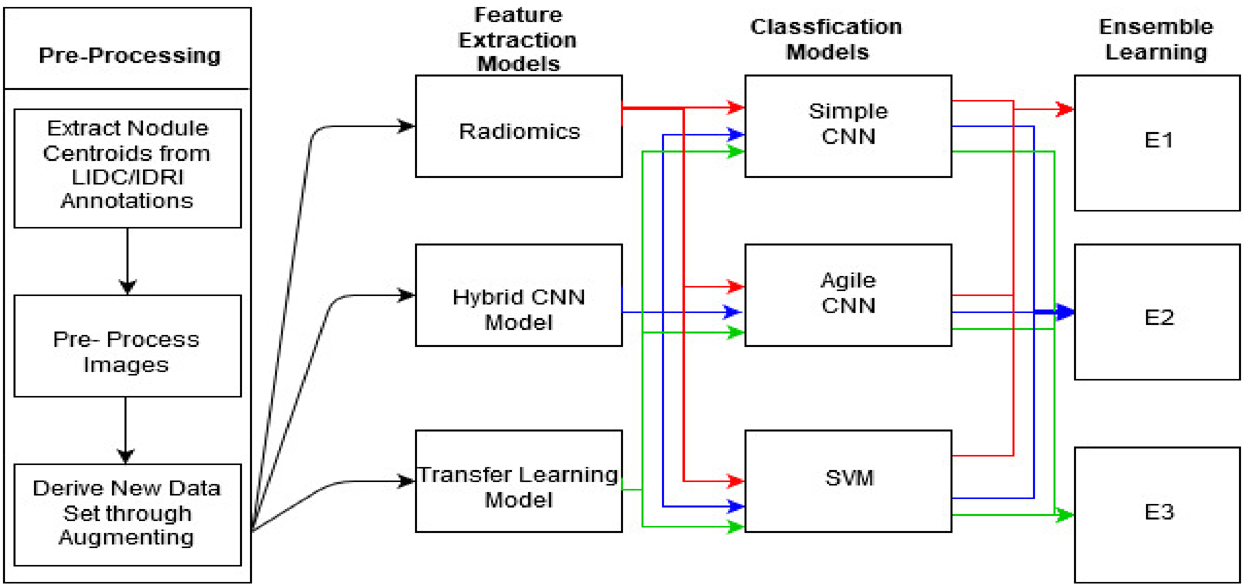


Figure 5: Research Model

Basically, CNN architecture comprises of two modules as discussed in figure 1. The convolutional part is known as feature extractor module. The part containing fully connected layers is known as classifier module. Every study, which includes classification, also needs a feature extractor module. Usually, complete CNN is used for both the module in research studies. In this study, feature extractor module and classifier module of CNN are used as separate modules. Two set of experiments are completely based on deep learning, while the third experiment is a combination of Radiomics and deep learning. For classification, a classic technique for classification, Support Vector Machine (SVM) is also used in comparison with CNN classifiers. Lastly, Ensemble learning is applied with averaging technique to make results more robust.

3.2 Data Set

There were many publicly available datasets provided by The Cancer Imaging Archive (TCIA). For this study, the dataset chosen is commonly known as LIDC/IDRI. LIDC/IDRI stands for Lung Image Database Consortium and Image Database Resource Initiative. The modality of this dataset is CT scans. There are 1018 cases in this dataset. Each case has random number of CT slices and an XML file. The XML file includes annotation of the Region of Interest (ROI) from four different radiologists to reduce the errors. The data is in the shape of DICOM images. DICOM is a special file format for medical imaging. It includes all the CT scans in a uniform size of 512x512 size.

This study used only 26 scans which consists of 1376 original images. The data was augmented later in pre-processing to increase the number of images. LIDC/IDRI is a widely used and most trusted dataset for conducting studies which addresses pulmonary nodule. The reason for its being trusted is that the formation of this dataset took four radiologists in the loop. Each radiologist had to go under two categories of sessions: blinded and unblended. Blinded sessions were where one radiologist did not know about the annotations marked by any of other three radiologists. Contrary to this, unblended sessions included the annotations marked himself as well as by other three radiologists. The difference between blinded and unblended sessions were extremely minor.

Figure 6: Original CT

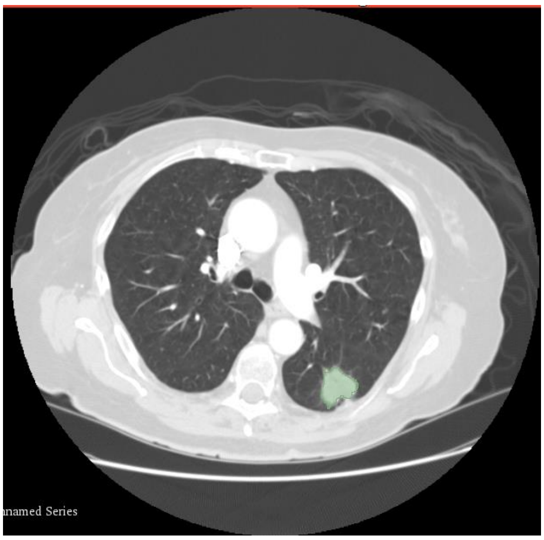


Figure 7: Annotations 1

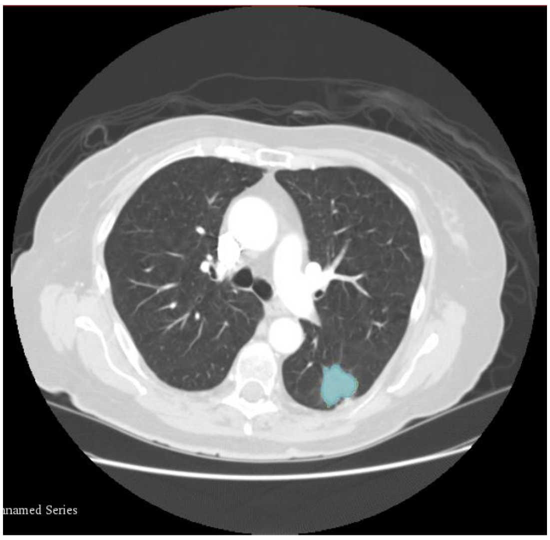


Figure 8: Annotation 2

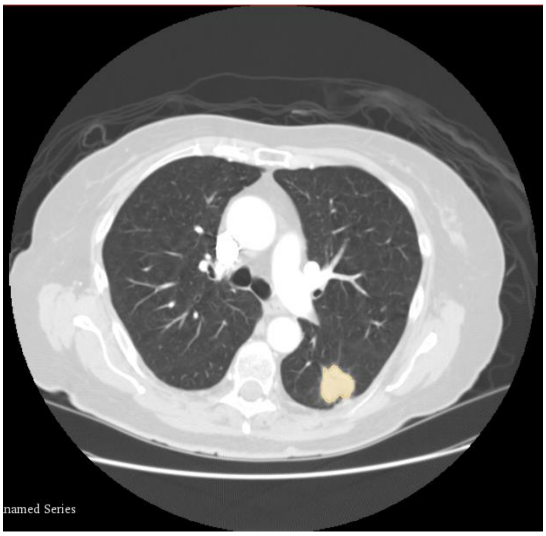


Figure 9: Annotation 3

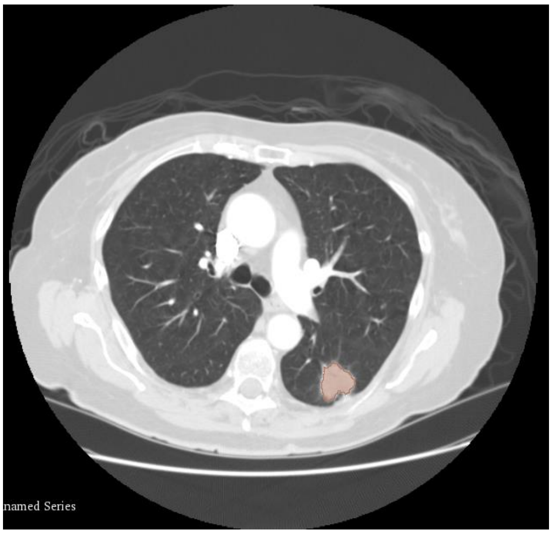


Figure 10: Annotations 4

LIDC/IDRI dataset has its exclusive definition of ROI as well. It classifies nodule in three categories:

1- nodule > or = 3mm
2- nodule < 3 mm
3- non-nodule > or = 3 mm

The difference between nodule and non-nodule is that nodule is a malignant tumor having characteristics like calcification, malignancy, sphericity, texture and more. Also nodules have spherical shape for which set of (x,y) coordinates of ROI are given. Whereas, non-nodule do not have characteristics as nodules. Also, the locus is defined in only one point i.e. 1 value of x-axis and 1 value of y-axis.

3.3 Pre-processing

Pre-processing for each set for experiments was done in a different manner. Each pre-processing phase was carried out according to the demands of the feature extractor phase of that particular experiment. Hence, three different types of pre-processing were carried out done for this study. A common step in all three types of preprocessing was data augmentation.

3.3.1 Pre-processing for Experiment 1

It was not possible to give DICOM files as input. Therefore, DICOM files were converted to Numpy arrays and then passed to the respective network.

3.3.2 Pre-processing for Experiment 2

For experiment 2, the implementation of input module was done with Python Image Library so a JPEG, PNG or similar format was required. For this experiment, DICOM images were first converted into lossless JPEG and then into Numpy arrays with built-in function. These JPEG images were fed to VGG16. A study already supported that VGG16 is resilient to formats and distortion [28].

3.3.3 Pre-processing for Experiment 3

For experiment 3, it was necessary to compile each slice of CT belonging to one patient on top of one another in a stack position. Using Slicer 3D software, this stack of CT images was aligned with respective segmentations. Afterwards, these two files were combined and imported to NRRD format. For experiment three, input could only be provided in this file format.

3.3.4 Data Augmentation

In two set of experiments which included deep learning, data was augmented using Image Generator. For every image, at least 45 versions were made by using parameters such as rotation, zoom, vertical flip, horizontal flip, width shift and height shift.

3.3.5 Initial Classification

To train neural networks upon our data, it was necessary to make categorical folder of positive and negative CT scans. To differentiate between positive and negative classes by hand, XML file was used. Every positive scan had its annotation in the XML file against its UID number. The data was preserved under a specific encoding which involved number of folders within another. It was a thorough job to understand the encoding scheme and automate the planned paradigm. An intensive implementation was carried out from scratch to successfully cover this classification. An example of positive and negative CT scans is given below.

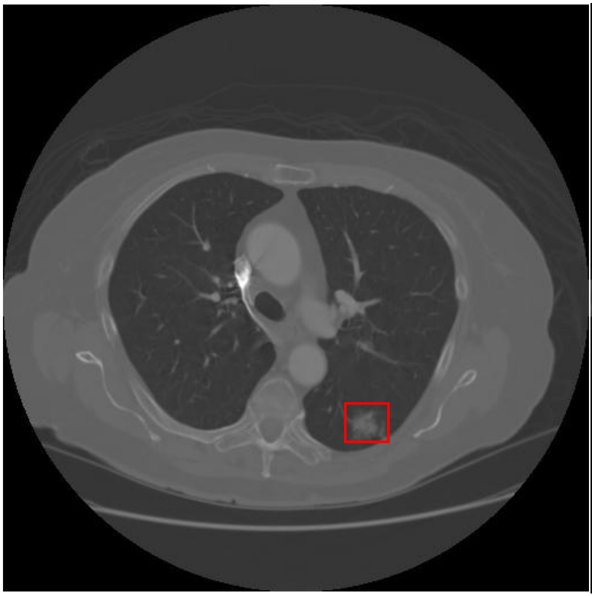


Figure 11: Positive CT Scan of Patient 1

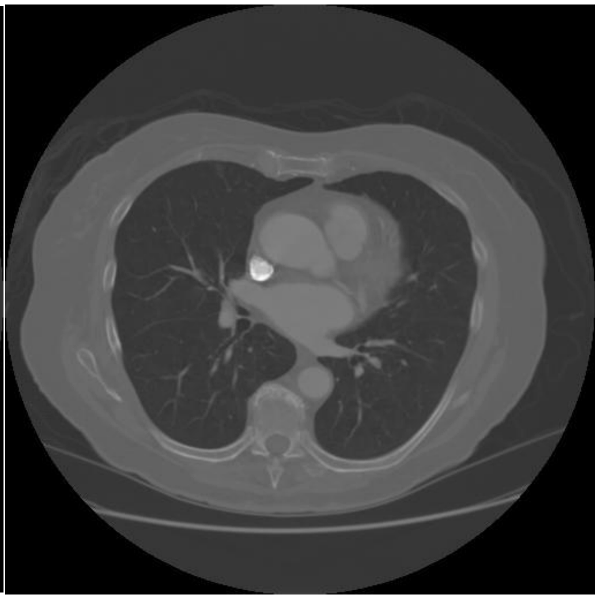


Figure 12: Negative CT Scan of Patient 1

3.4. Feature Extraction

Usually, when CNN is used in a study, it is used as a whole for feature extractor module as well as for classifier module. Recently, a study was conducted in which CNN was used as feature extractor module with standard classifiers for classification such as SVM, MLP, Random Forest and Bayesian Belief Network. [14] Taking inspiration from this work, a complete separate phase for feature extraction was introduced in this study. First two set of experiments drew features using different techniques of CNN while for third set of experiment, Radiomics features were extracted.

3.5. Classification

For classification phase, a combination of three classifiers was designed. It consisted two CNN models as classifiers and one standard classifier, SVM. This design remained same for all three set of experiments. It was designed to observe the strength of CNN as classifiers with two CNN models and compare the results with a classic classifier model.

3.6 Ensemble Learning

Ensemble Learning with Average Technique is a classification method which states to combine many results from one or more models of one or more types. The combining of results involves various approaches such as averaging, median, maximum and voting techniques. It helps researchers making one predictive model by using Ensemble Learning. For this study, averaging technique has been used.

3.7 Methodology

The idea of this study was to observe a comparative analysis among different techniques and models of CNN using the same experimental environment and variables. The techniques used in deep learning units of experiments for feature extractor module are as follows:

1- Transfer Learning

2- Hybrid CNNs

For each technique, models used are as follows, respectively:

1- ResNet50

2- VGG16 and VGG19

For classification module, three classifiers are used against each feature extraction module. Two of them are deep learning techniques and one is traditional classification technique as follows:

1- Simple CNN

2- Agile CNN

3- SVM

In deep learning classification techniques, following CNN models were used:

1- AlexNet

2- VGG16 and VGG19

Later, Ensemble method with unweighted average technique was applied on results received from each module of classification.

3.7.1. Transfer Learning with ResNet50

Transfer Learning technique suggests to use a pre trained model of neural network on a generic dataset and use it to solve the problems of another domain [29]. In this study, for Transfer Learning, Resnet50 was used. A pre-trained Resnet50 on ImageNet dataset was used to extract features. Features were extracted in a Numpy array in shape of 7*7*2048 which means feature maps containing 100,352 feature. These features were fed into classification module.

3.7.2 Hybrid CNN with VGG16 and VGG19

Hybrid framework refers to the architecture in which layers or parameters are combined from different models of neural networks. Hybrid CNN includes the combination of layers or parameters from different CNN models. For this study, VGG16 and VGG19 are used to make a Hybrid model. The implementation of Hybrid model was not possible using Sequential Model of Keras. Therefore, it was implemented using Functional API. It is a cascading architecture in which output from two models is merged after convolution part. The number of features extracted from each models were 25,088 and combined from both, it was 50,176.

3.7.3 Radiomics

Radiomics features are statistical features. Radiomics features are distributed over four categories of ROI. These four categories are as follows:

1- Shape

2- Volume

3- Texture

4- Intensity

This study includes 16 Radiomics features distributed randomly over four categories mentioned above.

3.7.4 Classification Module

The classification module is a set of three classifiers. The same classification module is used for each feature extraction module.

Simple CNN

The first model in classification module is Simple CNN. Simple CNN refers to any original model of CNN without any transition through any technique such transfer learning or hybrid. For this study, Alexnet was used under Simple CNN. AlexNet has eight layers, 5convolutional and 3 deep layers.

Agile CNN

The second model in classification module is Agile CNN. Agile CNN is a type of Hybrid CNN, which is designed after combining two models of CNN. The difference between hybrid and agile is that in Hybrid, two models are combined in terms of layers or parameters whereas in Agile, one model is constructed with the layer or parameters settings of the other model. The inspiration for Agile was taken from the study published by Xinzhou Zhao *et al*, in which they used AlexNet and LeNet [15]. They constructed LeNet with parameter settings of AlexNet. For this study, VGG16 and VGG19 were chosen as for Hybrid CNN. Since, VGG16 and VGG19 have same parameter settings of hidden layers, convolution layers were kept as VGG16's architecture whereas totally number of layers were increased in fully connected layers by 3. Hence, the number of fully connected layers in our model is 6 which since, this part is actually used for classification. Resultantly, total numbers of layers are 19 similar to VGG19.

SVM

Support Vector Machine is a standard technique among classifiers in machine learning. It refers to the classification through the parameters which escalate the differences between classes. It works best for binary problems. SVM is third classifier for each feature extraction module in this research study.

2.7.5 Ensemble Learning

Ensemble Learning is a method to combine the results from different classifiers into one unit to make the system more stable and robust. Ensemble Learning woks with few techniques which are average probability, voting, median, max and other statistical methods. Rahul Paul *et al* describes the importance of ensemble learning by saying, "Ensemble learning is an approach that combines the predictions of multiple learned models to enhance accuracy" [24]. They have used Ensemble learning to combine three classifiers based on CNN architecture with average probability. In this study, Ensemble learning is applied with unweighted averaging technique.

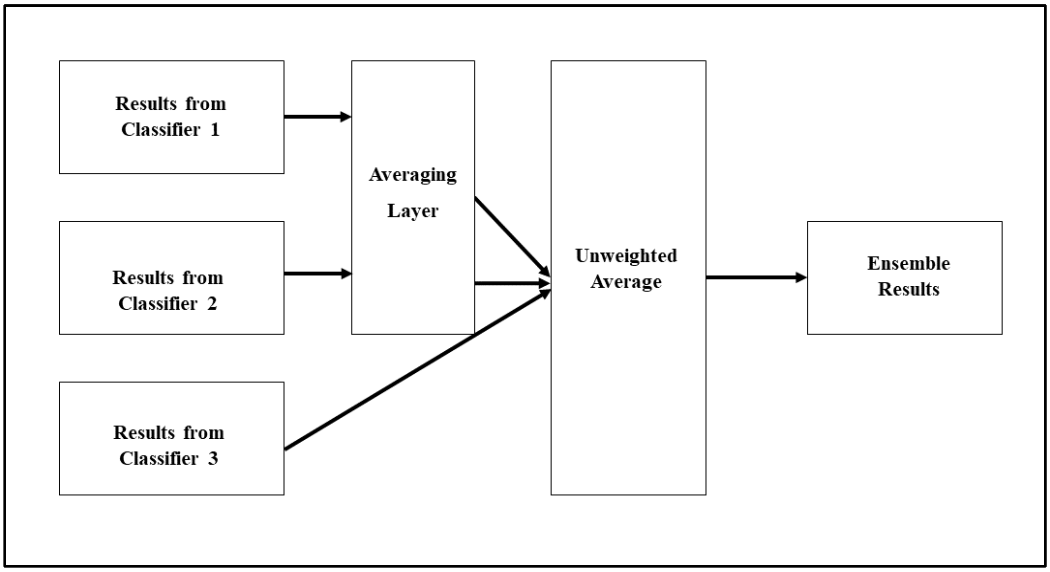


Figure 13: Ensemble Model

3.8 Methods

The experiments for this study, Google Colab was used. GPU runtime provided by Google Colab gave access to random GPUs. For training Hybrid CNN, Tensor Processing Unit – TPU runtime was used from the above mentioned platform. DICOM Viewer software was used for visualizing the data in DICOM format. A software named 3D Slicer was used for visualization of segmentations and annotations, conversion of DICOM file format to NRRD and for stacking up all slices of CT images. For computing these Radiomics features, PyRadiomics package was used. “This is an open-source python package for the extraction of Radiomics features from medical imaging” [30]. It is newly established community which is growing exponentially due to the standing and credibility of the performance. This package is developed by Computational Imaging & Bioinformatics Lab - Harvard Medical School.

CHAPTER FOUR: FINDINGS

This chapters contains the details for experiments and their results. It also contains characteristics of different experiments with respect to features and results. It is a detailed chapter regarding results and empirical findings. It includes sections based upon experiments and parts of experiments.

4.0Experimentation

4.1 Experiment # 1

Results from Experiment 1 can be seen in two splits. The first split is the results from the three classifiers. First split reports the maximum accuracies. Second split gives results from Ensemble Learning. Experiment 1 was initially trained and validated on 100 epochs and later normalized on 30 epochs. The train test split was 70:30. Maximum accuracy is the highest accuracy from all the epochs. Average Accuracy is calculated from the Averaging Layer which gives mean of accuracies from all the epochs. It was necessary step for applying Ensemble. Ensemble of Experiment 1 gives 69.3% accuracy.

Feature Extraction Module	**Classification Module**	**Max. Accuracy**
Transfer Learning (Resnet50)	*Simple CNN: AlexNet*	***60%***
	Agile CNN: Vgg16 and Vgg19	***50%***
	SVM	***97.91%***

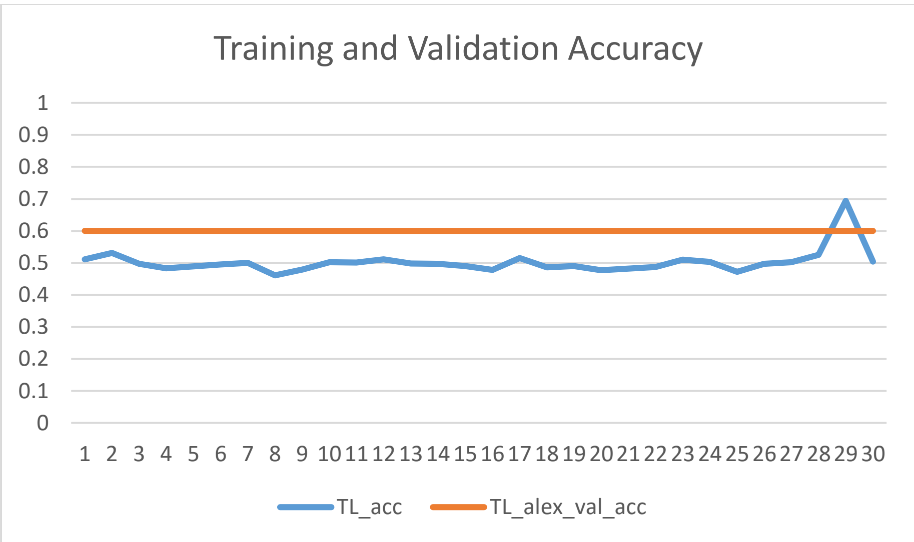
Table 1: Experiment # 1 (without Ensemble)

Feature Extraction Module	**Classification Module**	**Avg. Accuracy**	**Ensemble**
Transfer Learning (Resnet50)	*Simple CNN: AlexNet*	*60%*	***69.3%***
	Agile CNN: Vgg16 and Vgg19	*50%*	
	SVM	*97.91%*	

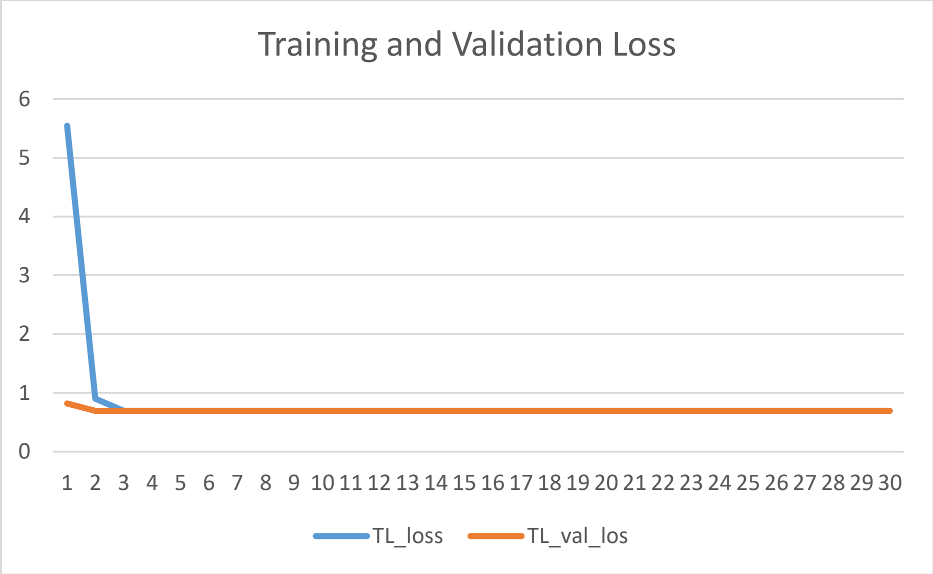
Table 2: Experiment # 1 (with Ensemble)

Following are the graphs plotted for the results of Experiment 1. The graphs are plotted between training and validation accuracies and training and validation losses of deep learning classification.

4.1.1 Transfer Learning – Simple CNN Graphs

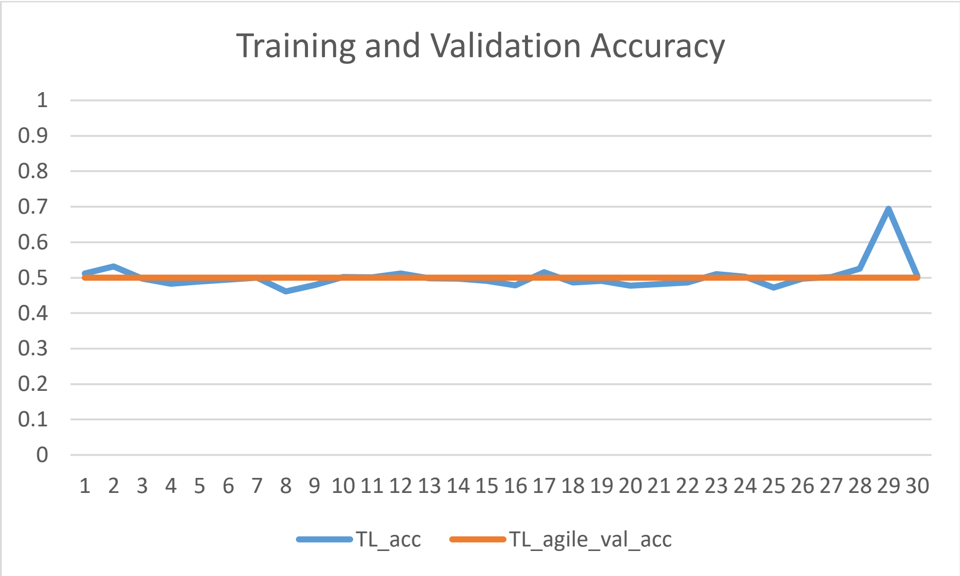


Graph 1: Accuracy Graph of Tl-Simple CNN Module

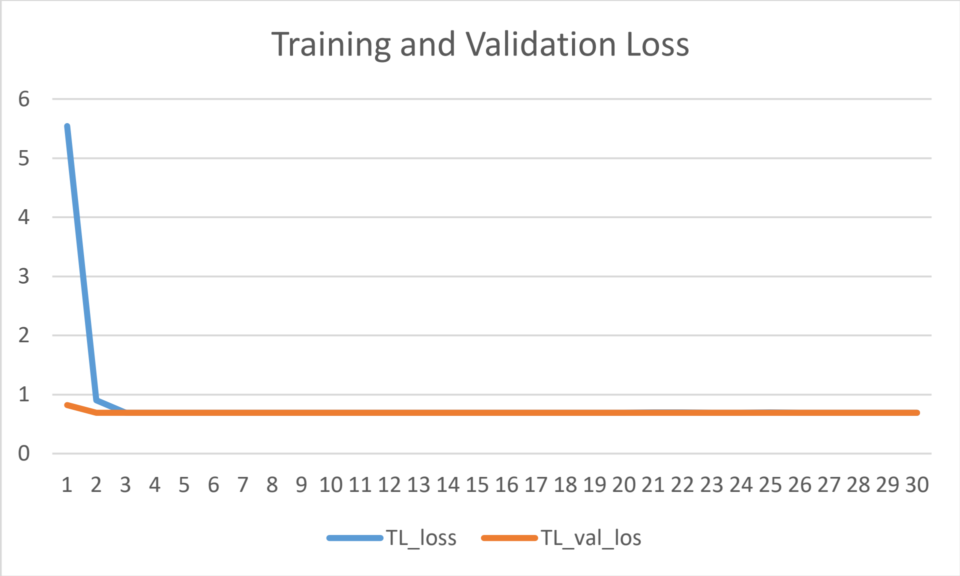


Graph 2: Loss Graph of TL-Simple CNN Module

4.1.2 Transfer Learning – Agile CNN Graphs

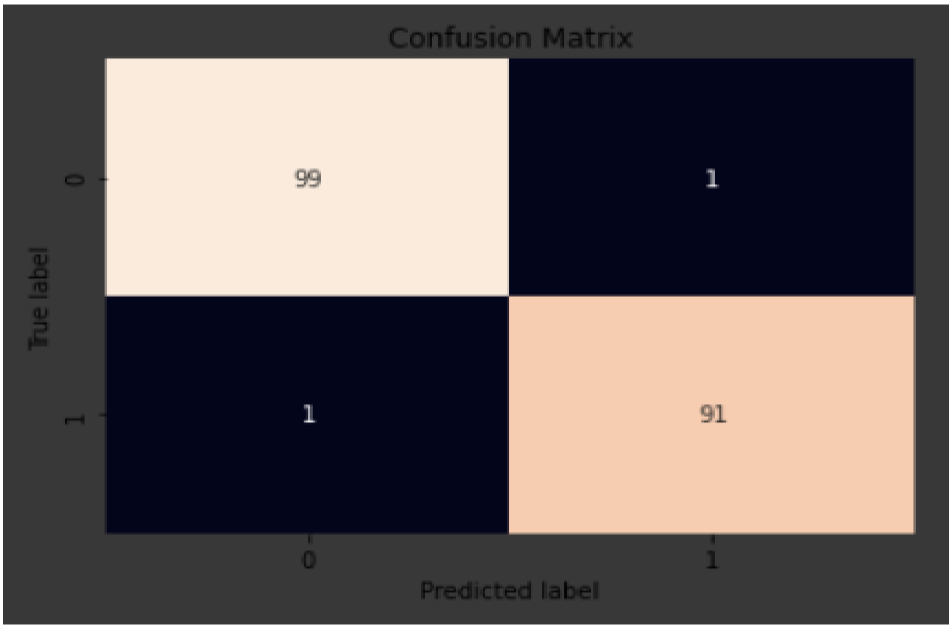


Graph 3: Accuracy Graph of TL-Agile Module



Graph 4:Loss Graph of TL-Agile Module

4.1.3 Confusion Matrix of Transfer Learning –SVM



Graph 5: Confusion Matrix of TL-SVM

4.2 Experiment # 2

Results from Experiment 2 can also be seen in two splits. The first split is the results from the three classifiers. First split reports the maximum accuracies. Second split gives results from Ensemble Learning. Experiment 2 was trained and validated on 30 epochs. The train test split was 70:30. Maximum accuracy is the highest accuracy from all the epochs. Average Accuracy is calculated from the Averaging Layer which gives mean of accuracies from all the epochs. It was necessary step for applying Ensemble. Ensemble of Experiment gives 76.6% accuracy.

Feature Extraction Module	Classification Module	Max. Accuracy
Hybrid CNN (Vgg16 and Vgg19)	*Simple CNN: AlexNet*	***91.26%***
	Agile CNN: Vgg16 and Vgg19	***80.58%***
	SVM	***85.47%***

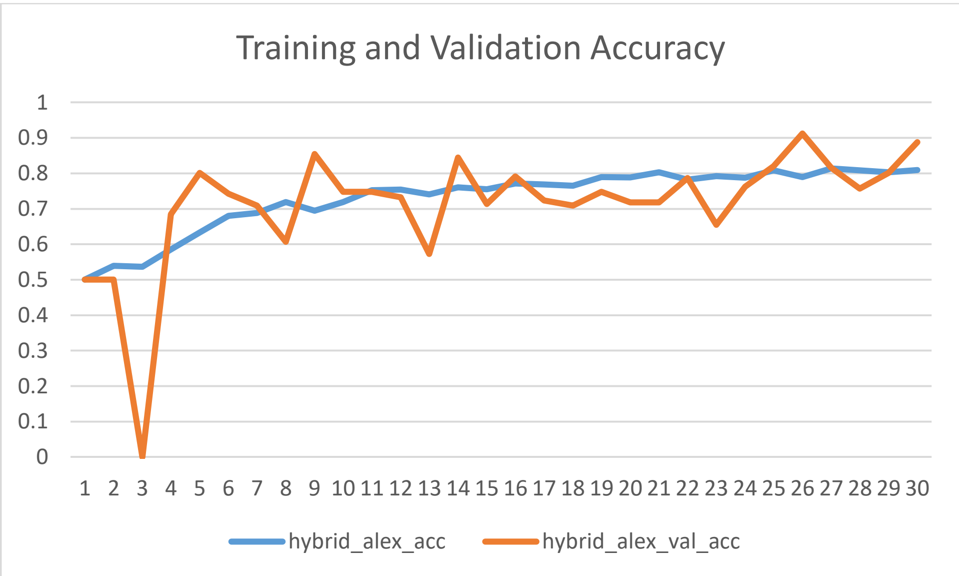
Table 3: Experiment # 2 (without Ensemble)

Feature Extraction Module	Classification Module	Avg. Accuracy	Ensemble
Hybrid CNN (Vgg16 and Vgg19)	*Simple CNN: AlexNet*	*72.88%*	***76.6%***
	Agile CNN: Vgg16 and Vgg19	*71.46%*	
	SVM	*85.47%*	

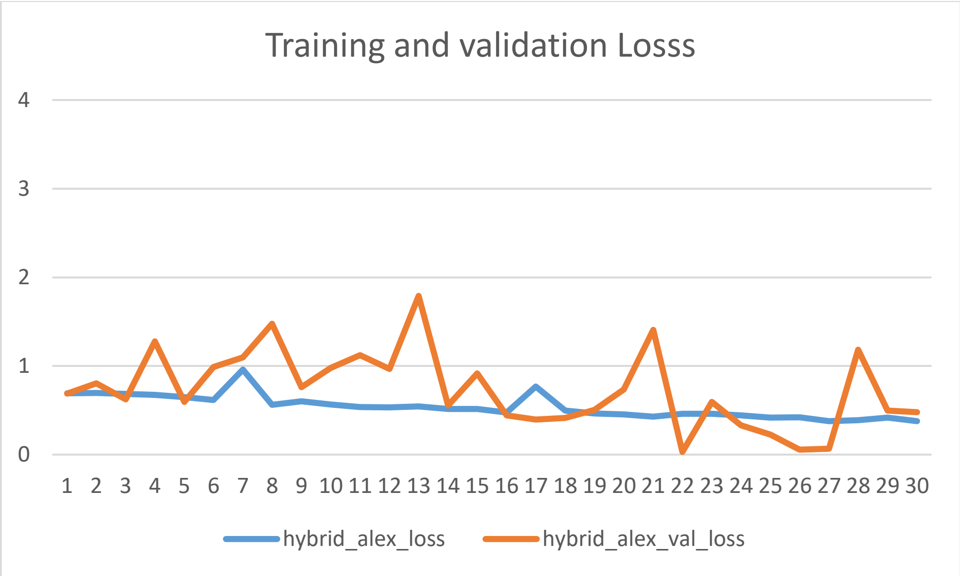
Table 4: Experiment # 2 (with Ensemble)

Following are the graphs plotted for the results of Experiment 2. The graphs are plotted between training and validation accuracies and training and validation losses of deep learning classification.

4.2.1 Hybrid CNN – Simple CNN Graphs

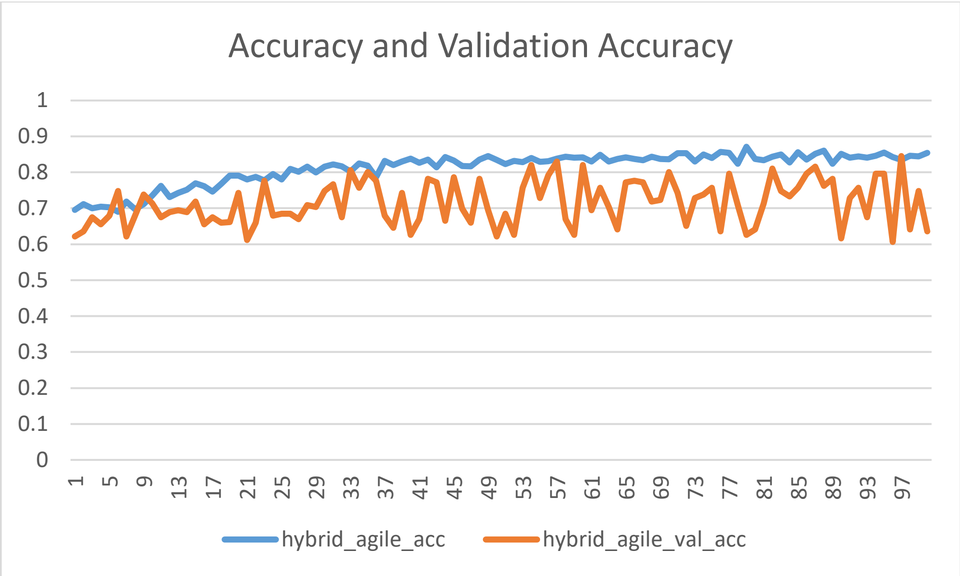


Graph 6: Accuracy Graph of Hybrid-Simple CNN Module

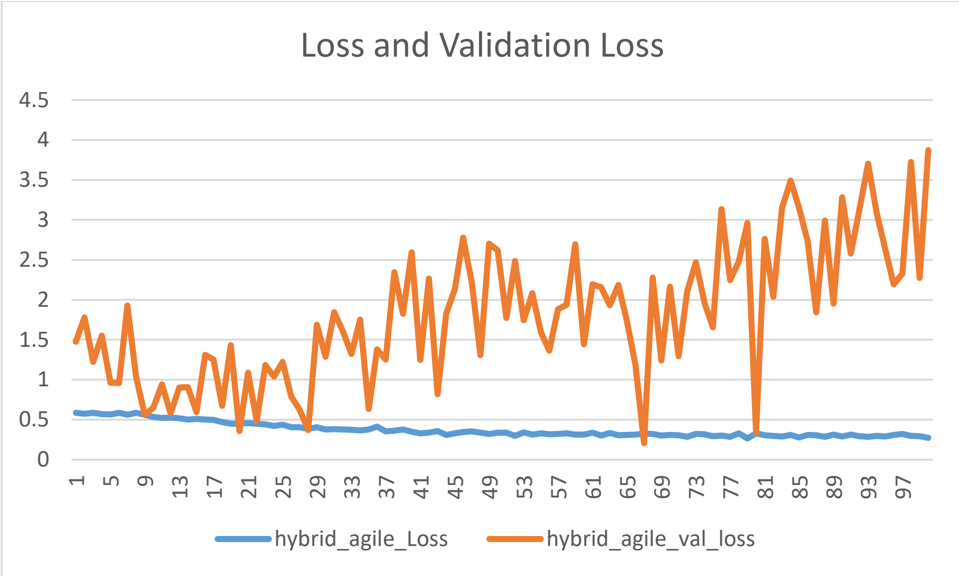


Graph 7: Loss Graph of Hybrid-Simple CNN Module

4.2.2 Hybrid CNN – Agile CNN Graphs

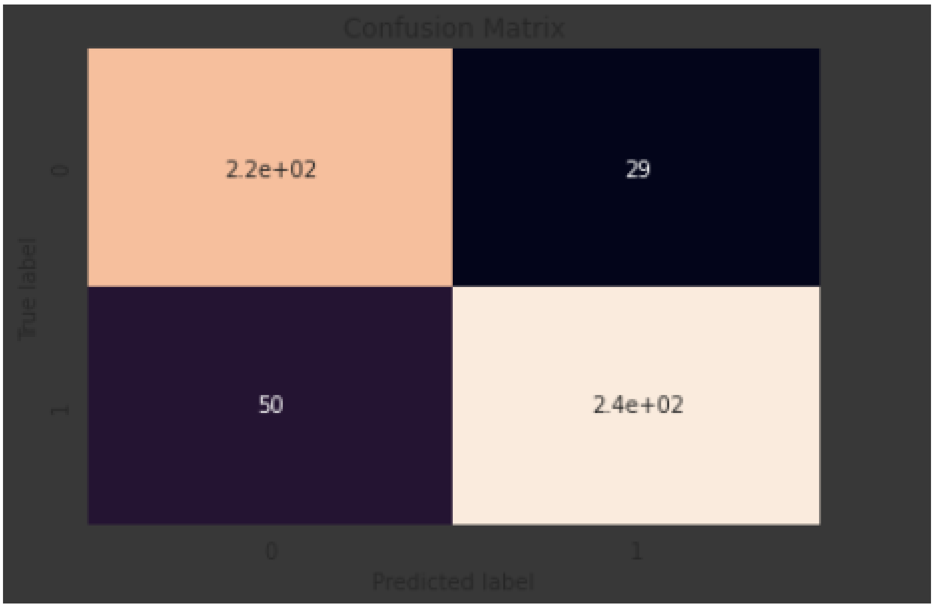


Graph 8: Accuracy Graph of Hybrid-Agile Module



Graph 9: Accuracy Graph of Hybrid-Agile CNN Module

4.2.2 Confusion Matrix of Hybrid CNN – SVM



Graph 10: Confusion Matrix of Hybrid-SVM Module

4.3 Experiment # 3

Results from Experiment 3 can be seen in three splits. The first split is the results from feature extraction. These are just statistical results extracted from Radiomics module. The second split have results from the three classifiers. The third split has results from Ensemble. Experiment 3 was initially trained and validated on 1000 epochs and later normalized on 30 epochs. The train test split was 80:20. Ensemble of Experiment 3 gives 79.65% accuracy.

Feature Extraction Module	Classification Module	Max. Accuracy
Radiomics	*Simple CNN: AlexNet*	***93.33%***
	Agile CNN: Vgg16 and Vgg19	
	SVM	***86.95%***

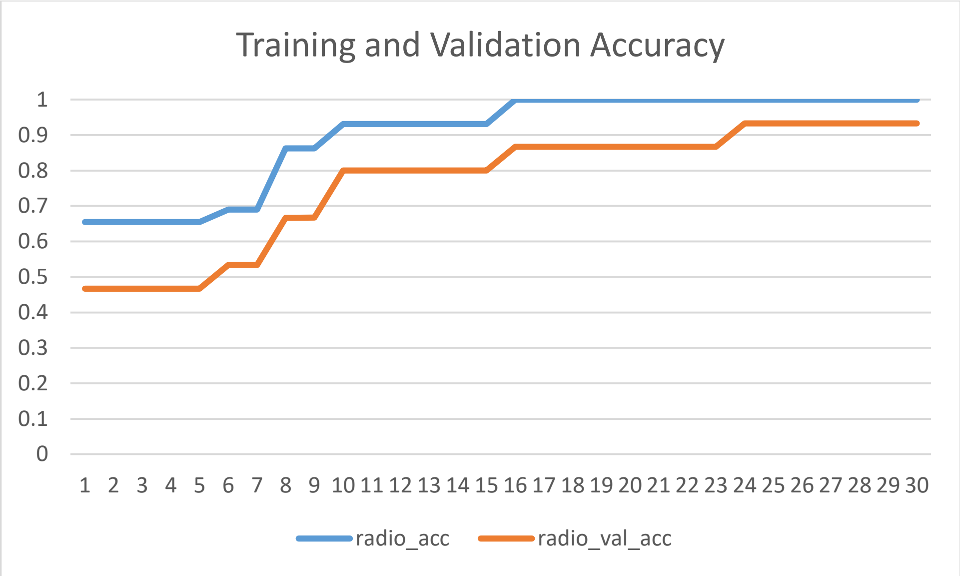
Table 5: Experiment # 3 (without Ensemble)

Feature Extraction Module	**Classification Module**	**Avg. Accuracy**	**Ensemble**
Radiomics	*Simple CNN: AlexNet*	*76%*	***79.65%***
	Agile CNN: Vgg16 and Vgg19	*76%*	
	SVM	*86.95%*	

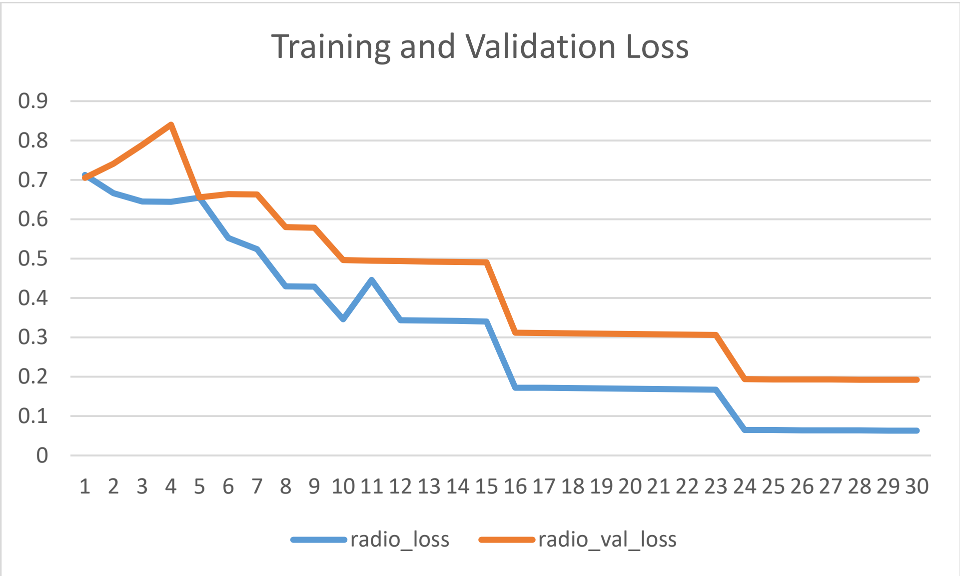
Table 6: Experiment # 3 (with Ensemble)

Following are the graphs plotted for the results of Experiment 3. The graphs are plotted between training and validation accuracies and training and validation losses of deep learning classification.

4.3.1 Radiomics – Simple CNN + Agile CNN Graphs

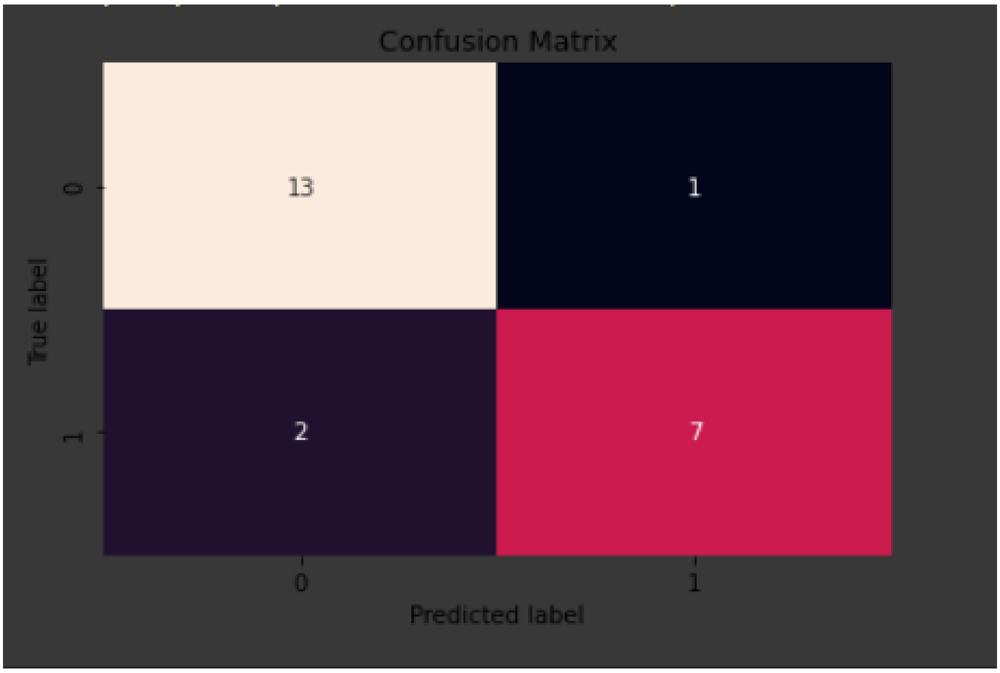


Graph 11: Accuracy Graph of Radiomics-Simple CNN Module



Graph 12: Accuracy Graph of Radiomics-Agile CNN Module

4.3.2 Confusion Matrix of Radiomics - SVM



Graph 13: Confusion Matrix of Radiomics-SVM Module

4.3.3 Radiomics Features

Following is the list of features extracted from Radiomics feature extractor module. Each feature has statistical value against each patient's CT scans.

Sr.	Radiomic Feature Name
1.	original_shape_Sphericity
2.	original_firstorder_Energy
3.	original_firstorder_Entropy
4.	original_firstorder_InterquartileRange
5.	original_firstorder_Kurtosis

7.	original_firstorder_Maximum
8.	original_firstorder_MeanAbsoluteDeviation
9.	original_firstorder_RootMeanSquared
10.	original_firstorder_Skewness
11.	original_firstorder_TotalEnergy
12.	original_firstorder_Uniformity
13.	original_firstorder_Variance
14.	original_glcm_Contrast
15	original_glcm_Correlation
16.	original_glcm_Idm

Table 7: List of Radiomics Features

A brief of these Radiomic features is given below. Let:

- **“X** be a set of *Np* voxels included in the ROI.”
- **“P**(i) be the first order histogram with *Ng* discrete intensity levels, where *Ng* is the number of non-zero bins, equally spaced from 0 with a width defined in the binWidth parameter.”
- *“p*(i) be the normalized first order histogram and equal to **P**(i)/*Np.”*

Malignancy chart is scale of 1-5 where 1 is the lowest and 5 is the highest value. High value of Malignancy means the cancer is spreading.

4.3.4 A few example computations of Radiomics Feature:

1- **Original Shape Sphericity:**

$$\text{Sphericity} = \frac{1}{A}\sqrt[3]{36\pi V^2}$$

Equation 1: Original Shape Sphericity

The range for its value is 0<*sphericity*≤1.

Sample of Values Achieved:

Malignancy	Original Shape Sphericity
1	-0.827970774
3	-0.52239
5	-0.63274703

Table 8: Original Shape Sphericity Table

2- **Original First-Order Energy:**

$$\text{Energy} = \sum_{i=1}^{N_p} (\mathbf{X}(i) + c)^2$$

Equation 2: original first-order energy

Where c is an optional value to shift the values of intensities to avoid negative results.

The larger the energy value, the greater the malignancy is.

Sample of Values Achieved:

Malignancy	Original First-Order Energy
1	1.7E+07
3	2.1E+08
5	1.1E+09

Table 9: Original First-Order Energy

3- **Original First-Order Entropy:**

$$\text{Entropy} = \sum_{i=1}^{N_g} p(i) log_2(p_i + \varepsilon)$$

Equation 3: Original First-Order Entropy

Entropy identifies the uncertainties in an image.

Sample of Values Achieved:

Malignancy	Original First-Order Entropy
1	-5.079448699
3	-4.790213674
5	-4.382795879

Table 10: Original Frist Order Entropy

4- **Original First-Order Interquartile Range:**

$$\textit{Interquartile range} = \mathbf{P}_i - \mathbf{P}_j$$

Equation 4: original first-order interquartile range

Where *i* is the ith image and *j* is the jth image.

Malignancy	Original First-Order Interquartile Range
1	-701.4
3	-165.75
5	-152

Table 11: Original First Order Inter Quartile

5- **Original First-Order Kurtosis:**

$$kurtosis = \frac{\mu_4}{\sigma_4}$$

Equation 5: Original First-Order Kurtosis

where $\mu 4$ is defined as 4th central moment.

"Kurtosis is a measure of the 'peakedness' of the distribution of values in the image ROI. A higher kurtosis implies that the mass of the distribution is concentrated towards the tail(s) rather than towards the mean. A lower kurtosis implies the reverse: that the mass of the distribution is concentrated towards a spike near the Mean value." [33]

6- **Original First-Order Uniformity:**

$$Uniformity = \sum_{i=1}^{N_g} p(i)^2$$

Equation 6: original first-order uniformity

Uniformity is defined as a measure of the sum of squares of each value of intensity. It measures the homogeneity level of the image array. Uniformity is directly proportional to homogeneity.

7- **Original GLCM Contrast:**

$$contrast = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}(i-j)^2 p(i,j)$$

Equation 7: Original GLCM Contrast

Contrast is defined to be a measure of local intensity variation. It favors those values which are away from the diagonal. "A larger value correlates with a greater disparity in intensity values among neighboring voxel" [33].

4.4 Comprehensive Comparison

4.4.1 Comparison among Experiments

Experiment No.	Feature Extractor Module	Classifiers	Max. Accuracy from Classification	Ensemble Accuracies
Experiment No. 1	Transfer Learning (ResNet50)	Simple CNN (AlexNet)	60%	69.3%
		Agile CNN (Vgg16 & Vgg19)	50%	
		SVM	97.91%	
Experiment No. 2	Hybrid CNN (VGG16 & VGG19)	Simple CNN (AlexNet)	91.26%	76.66%
		Agile CNN (VGG16 & VGG19)	80.58%	
		SVM	85.47%	
Experiment No. 3	Radiomics	Simple CNN (AlexNet)	93.33%	79.65%
		Agile CNN (VGG16 & VGG19)		
		SVM	86.95%	

Table 12: Comparison of all Experiments

4.4.2 Comparison among Classifiers (Max. Accuracies)

	Simple CNN	Agile CNN	SVM
Transfer Learning	60%	50%	97.91%
Hybrid CNN	91.26%	80.58%	85.47%
Radiomics	93.33%		89.85%

Table 13: Comparisons Among Classifiers

4.4.2 Comparison among Classifiers (Avg. Accuracies)

	Avg. of Simple CNN	Avg. of Agile CNN	SVM
Transfer Learning	60%	50%	97.91%
Hybrid CNN	72.88%	71.46%	85.47%
Radiomics	76%	76%	86.95%

Table 14: Comparisons among Classifiers (Avg)

4.4.3 Comparisons among Ensemble

Classification Techniques	Ensemble Learning
Transfer Learning	69.39%
Hybrid CNN	76.66%
Radiomics	79.65%

Table 15: Comparisons among Ensemble

CHAPTER FIVE: DISCUSSION

This chapter discusses the expectations behind designing this model and conducting this study, results achieved from the practical implementations and experimentations and the interpretation of these results. This chapters also includes pertinent suggestions with respect to the application and practical implementation.

5.0 Overview of Comprehension

This comprehensive study has produced results which can be interpreted in three dimensions.

First aspect to observe the results is to discuss the performance of feature extractor module. Second way to see the results is to compare the classification module. Third and last dimension is to discuss the use of Ensemble module.

5.1 Feature Extractors

In order to compare the performance of feature extractors used in this study, it is important to understand the differences between them. Experiment 1 and 2 used feature extractors which were built with two different techniques of CNN using separate CNN models. Contrary to this, feature extractor of Experiment 3 is Radiomics module. As per knowledge gathered from literature, CNN was the considered best feature extractor for images. This best behavior of CNN was due to its convolutional layers in which number of patterns were learnt. It was performing exceptionally well for medical imaging as well. This study was designed to compare different techniques implemented with different models of CNN. The feature extractor of Experiment 1 and 2 reported features in terms of Numpy array. Whereas, Radiomics, feature extractor of Experiment 3 produced statistical results.

For a study which includes detection of pulmonary nodules, deep learning performs good as feature extractor. The reason is that it can learn the pattern of an abnormality in a CT scan easily.

However, this study is explicitly about detection as well as classification of pulmonary nodule. Therefore, the focus is altogether on the characteristics of ROI. Any feature extractor which will give specific information about ROI would be considered better. Radiomics gives multiple quantitative features of ROI as described in Table # 7. Radiomics feature extractor provides much more insight about ROI than deep learning feature extractor. The reason is that deep learning feature extractors would only be able to learn the features such as volume, size and other qualitative features. In contrast, Radiomics is specifically designed to calculate many hidden features moreover, quantitative features of ROI. Hence, for any study which include extracting features with quantitative results in medical imaging, Radiomics is suggested to be used.

Contrary to this, if there is a dataset which does not have annotations and segmentation is not possible, then feature extraction through deep learning is the best option. The reason is that Radiomics feature extractor is dependent on ROI. Whereas, for deep learning only labelling a dataset can result into feature extraction. Moreover, for early detection of pulmonary nodule, deep learning would be the perfect option as it can detect small scaled abnormalities as well any lesion or node in pleural or vascular cavities.

Hence, the capacity of CNN as feature extractor should not be undermined. Rather, it should be understood that CNN can be best option provided that there is no segmentation or annotation in the dataset. However, Radiomics should be used in parallel for better understanding of features. Between both techniques of CNN as feature extractor, Transfer Learning is a better approach than Hybrid CNN. The reason to state this analysis is that Transfer learning with Resnet50 could compute twice feature maps than hybrid CNN which was a combination of two CNNs, VGG16 and 19. Secondly, Transfer Learning is implemented with Sequential Model which is way reliable than Functional API in which Hybird model was implemented. No doubt, Functional

API is flexible but its implementation follows a cascading architecture whereas Sequential model is in series.

5.2 Classification

Classification module of this study is one interesting dimension to explore. Three classification techniques with separate models were implemented. A brief comparison of performance is given Table # 13. The classification module of this study is comparison between two techniques of deep learning and one standard machine learning classifier, SVM. The implementation of deep learning as classifiers was done to explore the capacity of CNN models with various techniques. The training and validation of deep learning techniques was carried upon different numbers of epochs. Usually, the maximum accuracy is reported from the training and validation accuracies. The results show that SVM has shown a consistent performance within each experiment. This implies that traditional machine learning algorithm outperforms deep learning in terms of stable results in classification. Deep learning models implemented for classification have given mixed results with varying highs and lows. Hence, for any experimental study which needs to decide between deep learning and SVM for classification, SVM is recommended. Whereas, if the architecture of any study prefers to take the highest accuracy among all epochs, without considering the variation, then deep learning modules can be chosen for classification. This is suggested because highest accuracy reported from deep learning techniques implemented in classification module surpasses the results of SVM.

5.3 Ensemble Learning

Ensemble learning is applied on the results achieved from classification module against each experiments. A total of three ensembles are received from this study. As mentioned above, usually, maximum accuracy is taken from all the epochs implemented in an experiment. However, to make the results stable, all the results from classification was passed from an averaging layer. This gave an average of accuracies from all the epochs. This procedure of averaging of all accuracies brought a significant stability and balance in the model. Ensemble Learning was applied upon these results with unweighted average technique. The purpose of including Ensemble in this study was to make it more robust in terms of authenticity and practicality. The results achieved shows that the inferenced viewpoint of including Ensemble has been achieved. Applying ensemble is recommended in studies and architectures where results are achieved using different models and techniques.

5.4 General Comprehension of Experiments

It was interpreted while designing the architecture of this study that data driven features extractors are better in performance than hand-crafted feature extractors [9]. It has been proved from the performance of deep learning as feature extractor. Without any prior knowledge, except labels, fed into the system, the deep learning techniques were able to extract number of qualitative features. It was believed, that including Radiomics module would be an interesting aspect of comparison. This has also come out as a successful inference as the comparison brought clarity in terms of best conditions for both types of feature extractors to be used.

While initiating this study, it was a notion best believed through literature review that no comprehensive comparison has been done among deep learning techniques as feature extractor and classifiers [19]. This study has reported useful results in this regard as well. Two famous techniques, Hybrid CNN and Transfer Learning, were implemented with famous models VGG16 and 19, and Resnet50. Prior to this study, same CNN model was used for feature extraction and classification. If in any study, a separate classification technique was used, it was chosen from traditional classifiers. This study not only compare two different techniques of deep learning as classifiers but also their performance in comparison to SVM. The combination of Resnet50-SVM was reported to have 86.98% accuracy in [14] on the same dataset. This study has surpassed the results from this combination by a major percentage. This study reports 97.91% accuracy from this combination. A major reason can be the difference in pre-processing the dataset and less amount of data used in this study.

Though, it seems as if applying Ensemble has reduced the accuracy in terms of percentage. Yet, it is a misinterpretation. The fact is that the Ensemble works due to two reasons:

1- Variance Reduction
2- Bias Reduction

"For simple models, average of models has much greater capacities than single model. Averaging models can reduce bias substantially by increasing capacities, and control variance by fitting one component at a time" [31].

The use of unweighted average in Ensemble is the most common technique. It is the most recommended approach for neural networks. "Due to the high capacity of deep neural networks, simple unweighted averaging improves the performance substantively. Taking the average of multiple networks reduces the variance" [33].

5.5 Limitations

This study achieves all the prospects promised by this architecture. However, there has been many limitations to the course of this study. The limited computation power was the major problem. There was no local machine with GPU. Hence, Google Colab was used. Google Colab gave limited time to perform the experiments with the use of GPU as per availability. Due to this, it was not possible to re-run the experiments again and again while changing the parameters for better observations. With continuous disruption and slow speed of internet, it was not possible to use more data from the original dataset than used in this study. The reason is that in order to perform experiments on Google Colab, data had to be used from Google Drive. With the speed of internet available, it was not possible to invest more time in just data uploading.

The results of classification from Transfer Learning - Simple CNN module and Transfer Learning - Agile CNN module are 60% and 50%, respectively. The reason is that for Transfer Learning, pre-trained network was used, the weights of which were trained on ImageNet Data set. ImageNet is a dataset which contain natural images. Moreover, after transfer learning, Simple CNN and Agile CNN were used for classification. For both of these modules, convolutional layers were applied again on the extracted features from respected models. For this experiment to work properly, the parameters dependent upon fully connected layer of respected models had to be removed. Instead, settings of Transfer Learning had to be fixed with the dense layers of Simple and Agile CNN.

Radiomics features extractor gave statistical results. These results had to be fed in Simple CNN, Agile CNN and SVM for classification. For simple and Agile CNN, it was not possible to process these statistical results in convolutional layers. Hence, only dense layers were kept in operation. Since, the number of dense layers in both the models were same, the experiment was conducted only once and was considered as twice with the same results. Due to limited time and access to the resources, only one parameter could be assessed i.e. accuracy.

While applying ensemble learning, many limitations were faced. Since, it is a new technique to enhance the performance of a predictive model, not many publications were available. However, with the available resources, it was only possible in limited time to carry out Ensemble with unweighted Averaging technique. With more time and computation power, it is possible to compare these results with results obtained from averaging probabilities, voting and max and median approaches as well. There could be statistical and mathematical calculations to prove the strength, Ensemble brings in a model. But due to limited time, it was not possible to include proofs and calculations.

CHAPTER SIX: CONCLUSION

This chapter is based on three parts. The first part includes conclusion of this study. The second part is comprised of contributions and the final part contains future work.

6.0 Conclusion and Future Work

6.1 Conclusion

This study is the comparative analysis between different techniques and models of deep learning for detection and classification of Lungs Cancer. This four-part model of this research work gives three comparative results divided among three experiments and four modules. Among the module of feature extractor, it is concluded from this work deep learning techniques should be used as feature extractor module where there is no annotations or segmentation provided and only difference given is negative CT scan and positive CT scan. Transfer learning is a better approach than hybrid CNN in extracting features. Contrary to this, where segmentations and annotations are provided, Radiomics should also be included in parallel to deep learning feature extractor for Radiomics extracts quantitative features. Among the module of classification, SVM shows consistent and stable results. For Ensemble Module, it is concluded that Ensemble should be considered in researches which compensates multiple techniques and models in their architectures. It brings stability and credibility in the results.

6.2 Contribution

The dataset LIDC/IDRI follows a naming conventional which was not explained or documented. With much dedication and investment of time, an implementation was carried out which could fetch the data from XML files with correct results. Moreover, the annotations were recorded in a tree structure in XML format. The data from XML files was unable to be fetched due to the inclusion of same namespaces. It was identified through thorough study and experimentation.

6.3 Future Work

In future, the scope of this study can be expanded by compensating the points discussed in this part. The size of dataset can be increased in future research. With more computation power, experiments can be re-run again and again in order to observe the learning rates and compare the results among different parameters. More features can be extracted from Radiomics module. Further techniques of classification from machine learning can be implemented and compared. Ensemble method should be implemented with other techniques as averaging probabilities, median and voting. More accuracy parameters such AUC, Sensitivity, Specificity and TPR can be included. Different kinds of loss parameters and optimizers can be used and compared.

REFERENCES

1- F.Bray, J. I. (2018). Global Cancer Statistics 2018: GLOBOCAN estimatesof incidence and mortality worldwide for 36 cancers in185 countries. *Cancer J. Clinicians*, 394-424.

2- Medicinet. (2020, August 20). *Lung Cancer*. Retrieved from Medicinet: https://www.medicinenet.com/lung_cancer/article.htm

3- Henschke, C. (2006). Survival of Patients with Stage I Lung Cancer Detected On CT Screening. *New England J. Med. Vol.335*, 1763-1771.

4- Association, A. C. (2020, August 23). *Key Statistics for Lung Cancer*. Retrieved from Canecr.org: https://www.cancer.org/cancer/lung-cancer/about/key-statistics.html

5- Rui-Mei Feng, Y.-N. Z.-M.-H. (2019). Current cancer situation in China: good or bad news from the 2018 Global Cancer Statistics? *Cancer Communications 39*. GLOBCAN.

6- McDonald RJ, S. K. (2015). The effects of changes in utilization and technological advancements of cross-sectional imaging on radiologist workload. . *Acad Radiol.*, 1191-1198.

7- M. Nishio, O. S. (2018). Computer-aided diagnosis of lung nodule classification between benign nodule, primary lung cancer, and metastatic lung cancer. *PLoS ONE, vol. 13, no. 7*.

8- W.Shen. (2017). Multi-crop Convolutional Neural Netwok for Lung Nodule Malignancy Suspiciousness Classification. *Pattern Recognition Vol. 61*, 663-673.

9- F.Han, H. W. (2015). Texture Feature Analysis for Computer-aided Diagnosis on Pulmonary Nodules. *Journal of Digital Imaging. vol 28*, 99-115.

10- A.K.Dhara, S. M. (2016). A combination ofshape and teture features for classification of pulmonary nodules inlung ct images. *Journalof Digital Imaging vol. 29,no 4*, 446-475.

11- Y.Qin, H. Z.-M. (2018). Simultaneous accurate detection of pulmonary nodules and false positive reduction using 3D CNNs. *IEEE Int. Conf. Acoustic, Speech Signal Process. (ICASSP)* (pp. 1005-1009). Calgary: IEEE

12- H. Tang, D. R. (2018). Automated pulmonary nodule detection using 3D deep convolutional neural networks. *15th Int. Symp. Biomed. Imag. (ISBI)* (pp. 523-526). Washington D.C.: IEEE

13- Cohen, M. W. (2019, November 29). *3D G-CNNs for pulmonary*. Retrieved from Cornell Archive: https://arxiv.org/abs/1804.04656

14- R. V. M. d. Nóbrega, S. A. (2018). ‘Lung nodule classification via deep transfer learning in CT lung images. *31st Int. Symp. Comput.-Based Med. Syst.* (pp. 244-249). Brazil: IEEE.

15- X. Zhao, L. L. (2018). Agile convolutional neural network for pulmonary nodule classification using CT images. *Int. J. Comput. Assist. Radiol. Surg., vol. 13, no. 4,*, 585-595.

16- Y. Xie, Y. X. (2019). Knowledge-based collaborative deep learning for benign-malignant lung. *Trans. Med. Imag., vol. 38, no. 4*, 991-1004.

17- I. Ali, G. R. (2018). Lung nodule detection via deep reinforcement learning. *Frontiers Oncol., vol. 8*, 108.

18- X. Liu, F. H. (2018). Multi-view multi-scale CNNs for lung. *Pattern Recognit., vol. 77,*, 262-275.

19- PATRICE MONKAM 1, S. Q. (2019). *Detection and Classification of Pulmonary Using Convolutional Neural.* El Paso: IEEEAccess.

20- Silva. (2017). Ensemble Learning based Classification for BCI Applications. *BioEngineering (ENBENG), 5th Portugeese Meeting* (pp. 1-4). IEEE.
21- Cheng Ju, A. B. (2017). The Relative Performance of Ensemble Methods with Deep Convolutional Neural Networks for Image Classification. *Cornell Univeristy* (p. 20). Cornell.
22- Paul R, H. L. (2018). Predicting Nodule Malignancy using a CNN Ensemble Approach. . *Proc. Int Jt Conf Neural Netw.* (p. 8). PubMed.
23- Paul R, H. S. (2018). Predicting malignant nodules by fusing deep features with classical radiomics features. . *SPIE. J Med Imaging (Bellingham)*, 011021.
24- F., A. (2019). Artificial intelligence for precision oncology: beyond patient stratification. *NPJ Precision Oncology*.
25- Rizzo S, B. F. (2018). Radiomics: the facts and the challenges of image analysis. *Eurr Radiol Exp. 2018;2(1):36*, 8.
26- Lafata, K. Z. (2019). n Exploratory Radiomics Approach to Quantifying Pulmonary Function in CT Images. *Sci Rep 9,* , 11509 .
27- Paul R, H. S. (2019). Explaining Deep Features Using Raidologist Defined Semantoc Features and Traditional Quantitative Features. *TOMOGRAPHY Vol.5 No.1*, 192-200.
28- S.Dodge. (2016). Understanding How Image Quality Affects Deep. *Eighth International Conference on Quality of Multimedia Experience (QoMEX)* (pp. 1-6). Lisbon: IEEE.
29- Yang, S. J. (2010). A survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering, vol 22, no. 10*, 1345-1359.
30- Griethuysen, J. J.-T.-R. (2017). Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Research, 77(21)*, 104–107.
31- Toronto.edu. (2020, August 25). *Introduction to Machine Learning* . Retrieved from Toronto.edu: http://www.cs.toronto.edu/~urtasun/courses/CSC2515/CSC2515_Winter15.html
32- Laan, C. J. (2017). The Relative Performance of Ensemble Methods with Deep Convolutional Neural Networks for Image. *Cornell*, 20
33- Pyradiomics, C. (2020, August 29). *Radiomics Features*. Retrieved from Pyradiomics: https://pyradiomics.readthedocs.io/en/latest/features.html#radiomics.firstorder.RadiomicsFirstOrder