Unmasking Malignancy of Lung nodule using a modernized ConvNet toward the design of a Vision Transformer

by

Jamil Ur Reza
18101693
Khadiza Siddique Tannee
18101260
Sanjana Maruf Orpa
18101476
Hasan Mahmud Fahim
18101318
Riyad Foysal
18101559

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

Department of Computer Science and Engineering
Brac University
May 2022

Declaration

It is hereby declared that

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

Student's Full Name & Signature:

Jamil Ur Reza
18101693

Khadiza Siddique Tannee
18101260

Sanjana

Sanjana Maruf Orpa
18101476

Fahim

Hasan Mahmud Fahim
18101318

Riyad

Riyad Foysal
18101559

Approval

The thesis/project titled "Unmasking Malignancy of Lung nodule using a modernized ConvNet toward the design of a Vision Transformer" submitted by

1. Jamil Ur Reza(18101693)
2. Khadiza Siddique Tannee(18101260)
3. Sanjana Maruf Orpa(18101476)
4. Hasan Mahmud Fahim(18101318)
5. Riyad Foysal(18101559)

Of Spring, 2022 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on May 24, 2022.

Examining Committee:

Supervisor:
(Member)

1.6.22

Arif Shakil
Lecturer
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

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

Abstract

One of the most devastating cancers in the world is lung cancer. It is estimated that nearly a third of the world's cancer fatalities are due to lung cancer. Diagnosis and treatment of primary and metastatic cancers depend heavily on the ability to identify and characterize malignant cells. On the other hand, early detection of lung cancer is crucial for a patient's survival and significantly increases the survival rate. Malignant lung nodules may be detected early by oncologists using a variety of diagnostic methods such as needle prick biopsy and other types of imaging tests such as CT and PET scanning, as well as clinical examinations and other types of imaging tests. It's important to note that these treatments and biopsies are risky. A higher proportion of people are being infected with the disease, on the other hand. CT scans are commonly performed in the early stages of cancer detection. Lung cancer may be detected with a 2.6 to 10-fold higher CT detection rate than analog radiography, according to Awai [1]. As the slices get thinner, so does their ability to recognize objects accurately. To evaluate one slice, radiologists need an average of two to three minutes. The burden of cancer patients is increasing as the number of those diagnosed grows. CT imaging may be used to detect malignancy and cancerous nodules in a patient. When cancer nodules (stage I) are discovered, treatment may begin, and the danger of cancer spreading can be minimized. 70 percent to 92 percent of people diagnosed with stage 1 non-small cell lung cancer (NSCLC) should expect to live for at least five years following their diagnosis, according to existing statistics[30] . Considering the fact that a large number of early detection methods are already available, further research is needed to improve the accuracy of these methods and, as a consequence, the overall survival rate. Using ConvNeXt, we believe we can work more efficiently and precisely. Radiologists will also benefit greatly from this change. The validity of the proposed network was evaluated by comparing its performance to that of the other pre-trained CNNs, such as GoogleNet, AlexNet, and ResNet50, using a simulated dataset of pre-processed CT scan images: the Luna 16 dataset. Since our network outperforms the other networks in terms of classification, accuracy is evident from the results. Aside from pulmonary nodule detection, this proposal's approach may be simply adjusted to conduct classification jobs on any 3D medical diagnostic computed tomography pictures where the classification is very unpredictable and ambiguous, such as any other 3D medical diagnostic CT images.

Keywords: CNN, ConvNeXt, GoogleNet, AlexNet, and ResNet50

Acknowledgement

First and foremost, we are indebted to Allah, who is the creator and owner of the whole world, who is the most distinct and essentially one, all-merciful, and all-powerful, and who has blessed us with the spirit of perseverance as well as bravery, ability,and strength to complete this thesis.

A particular thanks go out to Arif Shakil Sir, who acted as our thesis supervisor and was instrumental in providing us with the necessary encouragement, guidance, and assistance, as well as courteous co-operation throughout the process of putting together our final thesis. His leadership, integration, and involvement have all had an impact on and propelled our study, and they have all acted as a source of inspiration for us throughout the process. We owe a debt of gratitude to the whole BRAC University community, including our professors, librarians, as well as other members of the university's staff, along with all of my lovely friends, for their invaluable advice and suggestions on my thesis project. Additionally, we want to express our appreciation for the help we received from many internet sources, notably the research from other researchers, for that we are very appreciative. All of our students who participated in the thesis project and actually made it a success should be thanked for their efforts. Last but not least, we'd want to express how grateful we are to our parents, who are really kind with their love and financial support, as well as their concern for our well-being.

Table of Contents

List of Figures

List of Tables

Nomenclature

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

$AttentionUNet++$	Attention Nested UNet
CAD	Computer-aided detection
$CLAHE$	Contrast limited adaptive histogram equalization
CNN	Convolutional Neural Network
CPM	The competitive productivity meter
CT	Computed tomography
GA	Genetic algorithm
$GELU$	Gaussian Error Linear Unit
ICV	Inters variance
$JSRT$	Jinterapanese Society of Radiological Technology
ML	Machine learning
NBI	Narrow Band Imaging
$NSCLC$	Non-small cell lung cancer
PET	Positron emission tomography
RAN	Residual Attention Neural Network
$RELU$	Rectified linear unit activation functions
SCC	Small cell lung cancer
SCC	Squamous cell carcinoma
SVM	Support vector machine

Chapter 1

Introduction

1.1 Introduction

Lung cancer is a complex disease with multiple clinically relevant subgroups, each of which has its very own set of characteristics distinct from the others. According to current estimates, more than 17 million people will be at risk of death from lung cancer by 2030 if sufficient medical care and early diagnosis are not given, with the vast majority of fatalities happening in poor nations. NSCLC (non-small cell lung cancer) and small cell lung cancer (SCC) seem to be the two types of lung cancer that may develop in the body. [2] Squamous cell carcinoma (SCC), large cell carcinoma, and lung adenocarcinoma are only a few kinds of non-small cell lung cancer that may occur. A recent study predicts that lung cancer might account for 11.4 percent of all new cancer cases in 2020, leading to 1,796,144 fatalities globally, thus responsible for 18 percent of all the cancer-related fatalities in the globe. [11] Lung cancer has the highest overall mortality rate of any cancer, and men are about twice as likely as women to be diagnosed with it and die from the illness. In most cases, lung cancer is discovered too late in a patient's condition to enable appropriate therapy to be provided. It has been shown that early detection of lung cancer increases the chance of survival by a substantial amount. The early identification of lung cancer may raise the probability of survival by 70 to 92 percent if the disease is discovered while it is in its early stages. Early-stage lung cancer, on the other hand, maybe difficult to detect due to the fact that there are only a few signs and symptoms to search for.

Diagnosis and identification concerns in imaging methods are made feasible via the application of computer-aided detection (CAD) techniques in imaging techniques (CADe). Whenever computerized tomography equipment generates a CT scan image, the resulting technology aids clinicians in understanding the image. To diagnose malignant lung nodules in their initial phases of development, radiologists must signal and identify certain possible focal opacities that may seem to be malignancy on radiographs, which may be highly difficult and time-consuming. Researchers all around the globe have been fascinated by the concept of computer-assisted diagnostics over the last few decades, and it has emerged as one of the most promising areas of scientific exploration in recent years. Extracting information from photos and identifying malignant neoplastic nodules may be accomplished via the use of computerized image segmentation techniques as well as machine learning algorithms, along with many other ways. A new division has been established between CAD techniques: classification methods that depend on hand-crafted features versus deep learning approaches that are based upon automated feature extraction. The categorization methods that are most often used are those that are based on hand-crafted qualities. Several strategies are commonly used to accurately measure radiological parameters such as form, texture, nodule size, as well as location, which can then be used in conjunction with such a classifier to evaluate if a nodule is malignant or nonmalignant. Texture, form, nodule size, and location are all examples of radiological parameters that are frequently quantified. Following this, these processes are more sensitive to measurement errors than other ways as a consequence of the ease with which a suitable set of options for lesion diagnosis can be obtained and chosen. Increasing the number of people who have been affected results in a directly proportional decline in the overall quality of the job. They also have a rising quantity of work on a daily basis, which makes it difficult to keep up with everything. As reported by the American Lung Association, usually, it takes an average of more than 30 days for a single patient to be diagnosed with lung cancer following the completion of a CT imaging examination. As a result, it is possible that the surgery would need to be redone at a certain time in the near future. A large amount of essential time will then be wasted in the process of identifying malignancy in the first place, which will be ensured by the patient if he or she is later diagnosed with cancer.

It was the goal of the very first computer-aided detection systems (CAD systems), enabling nodule identification and localization to develop in order to minimize the number of time radiologists spent doing examinations on patients while performing tests on patients themselves. In both categories, the current generation of computer-aided design (CAD) systems has consistently outperformed professional radiologists, regardless of the fact that professional radiologists really are the gold thing when it comes to nodule recognition and localization tasks, which are still performed by computer-aided design (CAD) systems. In light of recent advances in deep neural networks, particularly in the realm of visual interpretation, some researchers have suggested that chest radiographs be subjected to computer-based analysis as a matter of course. While using deep learning-based algorithmic tools, it is possible to investigate the patterns for pulmonary nodules in a highly time-efficient way while maintaining high accuracy. According to industry analysts, the usage of computer-aided diagnosis (CAD) will grow more accurate and speedier in the next years, and it will eventually replace all use of human radiological analysis over the next few years. Computerized diagnostics may assist oncologists in finding cancer more quickly, thereby resulting in a higher percentage of life savings. In the area of machine learning (ML), a lot has already been done, which will help reduce the number of human workers needed in the workplace. By creating artificial intelligence algorithms, it becomes more effective as they are subjected to much more relevant information using machine learning, which is an artificial intelligence approach that mixes statistics and computers. Many systems do not have appropriate detection accuracy, and it is required to create specific approaches in order to obtain the greatest degree of accuracy attainable, which is one hundred percent, in a vast number of systems. With the help of machine learning and image processing tools, researchers can identify and categorize lung cancer. Some features of lung cancer patients, like smoking behaviours, may, on the other hand, be effective in detecting the illness early in its course. Medical researchers started to experiment with machine learning just after the development of artificial intelligence in the hopes of enhancing the precision of disease and condition diagnosis and categorization. The Convolutional Neural Network (CNN) is the deep architectural model that receives the most attention and is also the most widely used among deep architectural models used mostly for classification as well as image segmentation tasks. The images are processed with a layer of Weights and Biases, enabling feature extraction and learning. It's likely that one or even more Fully Connected Layers will then be added just after the Convolution Layer and Max Pooling Layer, and that the Output Layer, which is the most often utilized initial layer in image processing. Each layer must utilize the Activation as well as Dropout functions, and both are available via the Activation function in order to function. The Convolutional Neural Network, in contrast to the previous layer, creates pictures with various colour channels and filter widths than that of the previous layer, giving it the title "Convolutional Neural Network." Regularization function is performed to each layer one after the other, beginning with the first, to ensure that the layers would not overlap. Regularization may be performed using a number of different activation functions, including linear, non-linear, sigmoid, and rectified linear unit activation functions (Relu). The Fully Connected Layers employ a dropout function in combination with an activation function, both of which are incorporated in this layer. It will find the class with the lowest chance of belonging to a certain class, and the dropout function may be beneficial in reducing overfitting difficulties in classifications and regressions.

The most recent research focused on using attention mechanisms to choose items that are far more relevant to the position at hand over other problematic irrelevant information, instead of vice versa, in a range of activities. Based on one example, the U-Net++ attention mechanism has indeed been utilized to autonomously partition the liver in the U-Net++ architecture using solely task-oriented properties at various layers of the encoder-decoder design. As a result, the attention mechanism has the ability to raise the weight of areas of focus while decreasing the weight of background regions irrelevant to the segmentation work at hand, which is exactly how it works.

This article seeks to make image interpretation as well as nodule recognition as precise as possible early in the process to benefit radiologists in their work. In a perfect scenario, we'd like to incorporate approaches from the area of computer vision into our research. Convolutional neural networks are used in the development of a highly accurate classifier for machine learning, and they're very beneficial for classification problems. A powerful lung cancer classifier is anticipated to significantly speed up and decrease the cost of lung cancer screening, allowing for more widespread early detection and increased overall survival in the future.

Convolutional neural networks (ConvNets) are useful for computer vision tasks in a broad variety of areas due to the fact that their inductive biases are highly variable. It is anticipated that ConvNeXt would outperform Vision Transformers' ConvNet in terms of overall performance. Due to the fact that ConvNeXt has the same number of groups as it does channels, depthwise convolution is used by the program. There is a strong connection between the self-attention weighting sum operation and the depthwise convolution. This operation works on a per-channel basis and only combines data in the spatial dimension[8]. ConvNets are seeing a decline in popularity as a direct result of the fact that Transformers outperform ConvNets in a variety of visual tasks because of the better scaling behaviour of Transformers, which includes multi-head self-attention. Another aspect of Transformers' microscale architecture that ConvNeXt inherits is that of layer-by-layer design. This feature may be found in ConvNeXt's microscale architecture. Traditional Convolutional Neural Networks (ConvNets) are comparable to ConvNeXt in terms of their capacity for efficiency as well as their fully convolutional nature when it comes to training and testing. For this, putting it into practice is not difficult at all[8].

1.2 Motivation

Medical imaging is indeed one of the most integral parts of modern healthcare systems and is heavily being used in various kinds of diagnosis. It is particularly being used for diagnosis of cancer, follow-up cancer therapeutic, and cancer protection measures. All the necessary information about the affected cancer cell, its shape, size, texture, current state, and location can be determined and stored through various kinds of modern visual imaging systems. Usage of high-resolution imaging techniques makes it even more helpful for detailed analysis. These visual cues are crucial in a variety of applications, including illness detection and oncologic research. Additionally, digital radio imaging techniques enable fully automated diagnosis with the help of big data-driven systems and computer-aided diagnosis systems(CAD) to a great extent. As a bonus, the increasing amount and quality of medical images allow for the creation of large-scale big data-driven systems for computer-aided diagnosis (CAD). Doctors frequently use diagnostic imaging (CAD) technology like segmentation in the interpretation and decision-making process of medical images. Patients may benefit from the use of automated disease diagnosis and categorization to improve the accuracy of diagnostic processes.

1.3 Problem Statement:

Our research entails the development of a cutting-edge deep learning system capable of detecting cancerous nodules in individual patients. This is a crucial aspect of our investigation. CT scans of the lungs will be used in the research to identify and classify any cancer patients who may have any symptoms. In order to do so, the researchers will use CT pictures of their test patients to construct a deep learning system tailored to them. According to these studies, a small number of nodules form early in the process and grow in size as time goes on. The cancer cells proceeded to proliferate throughout the lung tissue over time, finally destroying it. If these nodules are found early enough, they can be successfully treated, lowering the risk of patients dying from the disease. As a result, great emphasis has been placed on the early detection of malignant nodules.

1.4 Research Objectives:

Numerous studies have been conducted in the past with varying degrees of success in order to diagnose cancer nodules in the lungs. Machine learning's primary goal is to detect lung cancer nodules in both men and women so that the disease can be prevented before it spreads to others in the other parts of the body. There are different neural network techniques and multiple methodologies such as Convolutional Neural Network(CNN), Residual Attention Neural Network(RANN), Attention-based Multi-instance Neural Network, and others to classify and identify cancer cells, malignant lung nodules, and non-cancerous lung nodules. And we would like to use ConvNeXt to classify. The main dataset we decided to use is Luna-16 which contains CT scans that have been gender-categorized and we will use it to conduct our study. Early detection of lung cancer is essential for lowering patient mortality and increasing patient life expectancy, both of which are important goals themselves.

1.5 Report Overview

Chapter 1:

This first and foremost chapter introduces our entire research endeavour and serves as a primer for the remainder of it. This chapter also gives a brief summary of the disease that is studied, which is in our case lung diseases including cancer. It also explains the major goals and key objectives that we want to do.

Chapter 2:

In the second chapter, we go over the background research that we have just done for our investigation. We have talked about theoretical knowledge in regard to the main topic of our investigation. This section also includes a review of a number of earlier works that have been done.

Chapter 3:

The third chapter provides a high-level overview of our workflow before delving into the specifics of each step of the data preprocessing process in greater depth. This chapter elaborates on different neural network techniques such as ConvNeXt design, which was previously described.

Chapter 4:

The project is summarized in this section. Each Neural Network technique was compared against one another to see which one was the most accurate. In this part, we consider the long-term prospects for the project. We are tasked with creating an efficient embedded system. We also want to enhance the system's detection and preprocessing mechanisms.

1.6 Literature Review:

To generate a better, more accurate, and faster computer-aided diagnosis (CAD) system we have researched through numerous reputed research papers that have been published in the last couple of years. Here are a few citations for recent scholarly articles to get you started.

The ImageNet dataset, which was used by T. L. Chaunzwa et al. [25] to train the CNN, included information on gender, age, and smoking status, which proved useful for the VGG-16 architecture. In addition, CNN was utilized to extract characteristics from photos and sort them into categories. A grayscale picture patch of 50mm x 50mm was then sent into each of the VGG-16's three input channels, one for each of the three input channels. A collection of patients with ADC or SCC histology served as the main model A, while a combination of all the three histological types (ADC, SCC, and "Other") served as the second model B. Over the course of 100 iterations, the fine-tuning process was repeated 100 times. When using VGG-16 (Model-A), they were able to get an AUC of 0.709, which equates to 68.6 % efficiency and 82.9 % specificity. At a p-value of 0.018 percent, it is possible to obtain a level of accuracy of 37.5 percent.

To categorize and diagnose laryngeal cancer in its earliest stages, another research employed attention-based multiple instances learning techniques. The final results were positive (MIL). The Zenodo dataset utilized in [23] included both healthy and malignant laryngeal samples of tissue. A total of 33 patients were recorded. It was categorized into four groups based on the condition of the tissue: healthy, leukoplakia-infected, hypertrophic blood vessels, and intraphpapillary capillary loops (blue spots on the skin). Each class had 330 images. As more than just a starting point for the model creation, the pictures were reduced to 100x100 pixels and manually retrieved from 33 laryngoscopic images taken using Narrow Band Imaging (NBI). The endoscopic images were denoised using a Gaussian filter, which also was applied to the data that was noisy. To summarize, the problem was broken down into two binary categories: either 0 indicates that the tissues are normal or 1 indicates that they may be precancerous. Comparing the recommended strategy to transfer learning with CNN without pre-trained weights, the latter was shown to be superior. While other models were accurate up to an accuracy level of 0.98, these models outperformed them in terms of accuracy.

Biomedical image categorization utilizing deep neural networks, mostly based on CT scan pictures, has diagnosed lung cancer, according to the researchers. A three-stream deep neural network model is suggested in [26] for the detection of lung cancer using computed tomography (CT) scans. Two deep learning architectures and one machine learning architecture are being used to categorize photographs in this method. In machine learning, SVMs are given characteristics from spatial and frequency domains, whereas deep learning relies on transfer learning. Eventually, a weighted fusion is used to aggregate all three networks' projected scores it into a single composite score. Researchers utilize the LIDC/IDRI database to classify various types of lung cancer. This collection has a sum of 1018 images. When it comes to this case, 80% of the data set is being used to train and 20% for testing.

The quality of training images is improved by using rotation, scaling, flipping, and translation. 256 by 256-pixel resizing has been applied to all images. Pre-processing photos involves the use of CLAHE (contrast limited adaptive histogram equalization) and Z-score normalization. Once the frequency domain gets translated back to the domain, the picture is returned to its original state. In this instance, there were three separate results reported by the authors:

Without augmentation	Without augmented data	With augmented data
Resnet-50: 86%	Alexnet: 87.5%	Alexnet: 91.5%
Customnet: 89%	Resnet-152: 93%	Resnet-152: 96%
SVM: 85%	VGG-16: 89.5%	VGG-16: 92.8%
Proposed: 96.3%	Proposed: 96.3%	Proposed: 98%

Table 1.1: Comparison different model without augmentation, without augmented data and with augmented data

Approximately 1000 CT scans from the Luna-16 dataset of large and small tumors were analyzed by the authors of [20]. Prior to being scaled to its final size, it was then converted to Grayscale for better visibility. The Adaptive Median Filter was used to increase the quality of the photos and then instruct them to discriminate among malignant and non-cancerous pictures. The model has been able to design a GCPSO with the highest degree of accuracy possible 95.8079 %.

Malignant nodules are difficult to detect in their earliest stages due to their tiny size, which makes early detection difficult. The suggested technique [19] is used to preprocess the raw data in order to improve the quality of the low-dose images. VGG16, VGG19, and Alex networks have been shown to be useful in producing compacted deep learning features. With the help of a genetic algorithm (GA) trained to identify the most essential qualities for early detection, the recovered data collection may be expanded to its maximum. The next step is to examine the precision with which a given categorization system detects pleural nodules.

Watershed segmentation for identification and SVM for classifying were used by the authors of [14] to build a technique for identifying cancerous nodules for lung CT scan images. They are able to categorize nodules as malignant or benign using this approach. It's vital to have accurate detection tools for distinguishing malignant from benign lung tumors. A median filter was employed to remove salt as well as pepper noise from the CT images that had previously existed. A Gaussian filter was used to smooth the image and eliminate speckle noise, resulting in a clearer image. Watershed segmentation is used to properly distinguish between cancer nodules and then another fictional nodule that comes into contact with it. Following that, the data were used to extract characteristics that would be employed for training features within the classifier's future development. The nodule is subsequently classified as malignant or benign using a support vector machine (SVM) classifier. The accuracy and specificity of the model continue to increase, reaching 92% and 50%, respectively.

In order to reduce false positives throughout complex structures with no homogeneity, the authors of [29] used three key strategies: the merging of three different 3D Attention-based CNN architectures, notably MP-ACNN1, MP-ACNN2, and MP-ACNN3, and using an iterative training procedure to deal with the issue of inconsistent classification. The LUNA16 dataset, which is also extracted from the Luna-16 dataset that contains 1018 patient scan images, was used in this study. Despite the absence of nodule detection techniques, the unique attention strategy aided in the exact identification of tiny nodules whenever applied to 3D-CNNs, as well as a competitive productivity meter (CPM) score of 0.931 was achieved utilizing a 10-fold cross-validation method upon that dataset.

UNet++'s attention method was significantly improved in the medical field, as previously reported. The attention-aware segmentation network (Attention UNet++) discussed in the study [21] was developed and utilized for liver segmentation, as detailed in the paper. The suggested technique includes modified dense skip connectivity and a deep supervised encoder-decoder structure. The LiTS dataset includes 131 training CT scans as well as 70 test CT scans. The approach may raise the weight of the target area while decreasing the value of the background of the image, which is unrelated to the segmentation job, in order to improve segmentation performance. Attention Nested UNet (Attention UNet++) surpasses the competition for liver CT image segmentation in testing.

Bustamam, Abdillah, and Sarwinda (2016) [8], and in this paper they conducted their investigation using a novel attention-based image processing method known as Marker controlled watershed algorithm, and also the Gabor filter and region expansion. A closer look at the procedure from numerous perspectives was achieved by the use of 250 black as well as white CT scan photographs of 50 different individuals obtained from various angles and published in the publication. A cancer cell can be found in any location of the nodule using this technique, which recognizes it and makes a rough choice that can be used to more accurately measure the edge region of the nodule. The image is split up into two sorts of pixel values and binarized to fully understand the participation of cancerous cells in any location of the nodule using this technique.

The best technique for detecting early lung cancer is created by combining image processing, deep learning, as well as metaheuristics [28]. An autonomous technique based on deep neural networks was created to provide the finest potential diagnosis of CT-based lung imaging. The proposed solution uses a metaheuristic methodology known as the marine predator's algorithm to get the best possible architecture and network efficiency. In this study, the Marine Predator Algorithm is implemented to the RIDER dataset, and the results are compared to pre-trained deep networks such as CNN GoogLeNet, ResNet-18, VGG-19, as well as AlexNet to determine which is superior. During the pre-processing stage, the photographs are treated to noise reduction and image-level balancing. A convolutional neural network (CNN) can aid in the development of filters and criteria if appropriately trained. By employing a sequential training technique and extracting high-level features, the recommended model outperforms previous strategies. Regardless of the fact that now the training and also the test data are unrelated, 80% of the photographs inside the dataset are chosen randomly for training, also with the remaining 20% are used for testing. When compared to other state-of-the-art approaches, the suggested MPA-based strategy delivers the highest accuracy, sensitivity, and specificity, and also the lowest error rate. The suggested strategy yields 93.4 %, 98.4 sensitivity, and 97.1 % specificity with the least amount of error.

It is advised in [18] to have a chest CT scan in order to diagnose lung cancer although the anatomical location of the suspicious tumor is uncertain. An attention map and anatomical data for categorizing are two ways this network might learn more about human behaviour. Diagnostic imaging is used to distinguish between malignant or benign nodules during these screening and non-screening operations. Using 3D CT image volumes with clinical data, we can detect and classify lung cancer using a machine learning-based classification pipeline. As a starting point, we used information gleaned from a previously examined thoroughly lung scan. The winning team used a 3D CNN U-Net nodule detection network to create a lung nodule mask as part of their Kaggle Data Science Bowl 2017 application. This show is sure to please the audience. The CNN algorithm was being used to train and evaluate the 3D CT scans after they were obtained. This previously lacking functionality has been added to the 2D network: a 3D soft attention barrier (SAG). A total of four Nvidia Titan XP GPUs were used in this experiment. Using both the desired network as well as attention maps, it is possible to determine performance-related variables. The attention network had an accuracy of 0.687, whereas the clinical demographics only had an accuracy of 0.635, indicating that the attention network was much more effective than the demographics. This pipeline's combination of clinical and imaging features resulted in an Accuracy of 0.787 upon that testing dataset, which was a remarkable achievement.

As stated in [16], lung cancer is the biggest cause of cancer-related deaths globally. In CT scan images, convolutional neural networks(CNN) can differentiate between healthy and cancerous tissue, allowing for more accurate diagnosis and therapy. Pre-trained convolutional neural networks (CNNs), such as GoogleNet, have already been utilized to develop deep neural networks in other research. A decrease in the number of lung cancer deaths could be achieved with the use of low-dose spiral computed tomography (LD Spiral CT) (LDCT). For cancer detection, a CNN is used in CT scan pictures to discriminate between cancerous and non-cancerous tissue. These kinds of scenarios neural network design employ pre-trained CNNs. 60% of the neurons were found on the dropout layers. To see which network performed better, it was compared to other well-trained networks including AlexNet, GoogleNet, and ResNet-50. In terms of classification, both networks performed significantly better. It was built using the Luna-16 dataset, which includes 1018 CT scans of individuals. A 64-bit Windows 10 PC was used for each test. The graphics processing unit (GPU) of the computer is used in the tests (GPU). During ProposedNet's validation, precision values of 98 to 100% were achieved. All three neural networks were tested with about the same number of training examples.

Following the findings of this study [4], it is vital to highlight the importance of lung cancer early diagnosis in order to protect lives. With the help of the appearance, noncancerous lung nodules can be distinguished from cancerous lung nodules. Deep learning architecture frameworks are currently being developed here. CNN is very useful for the ability to recognize photographs in a variety of ways. It also works admirably in biological photo categorization tasks, as evidenced by the excellent results obtained in the research endeavor. A 3D CNN architecture created by Google was used to categorize and organize all of the information collected by the researchers. Although the photos in the data set are of low quality, the experimental findings demonstrate that the method is effective in this particular situation. The system's effectiveness may contribute to the future if the dataset is increased in size as well as the architecture is implemented more successfully than it currently is.

An encoder-decoder that uses MixNet to detect and learn about lung nodule traits is described by the author in [17]. GBM using combination with 3D MixNet is recommended for nodule categorization. According to a statistical evaluation of 1200 LIDC-IDRI images, which included 3250 nodules, the suggested strategy was found to be effective. The LIDC-IDRI comprises lung nodules of both noncancerous and cancerous cells. The novel strategy outperforms earlier approaches in terms of sensitivity 94%, specificity 90%, and accuracy 90%. MixNet, 2D R-CNN, as well as U-net encoder-decoders could be used to detect and learn about lung nodules. On a CT scan of the lungs, 3D Faster R-CNN discovered nodules. On the back end, feature extraction is performed using the MixNet framework. The limiting, nonlinear ReLu, and convolution layers are all created by the inner and outer link modules. To locate nodules, we used the R-CNN with U-Net architecture and MixNet for pixel-by-pixel feature labelling. Deep network features are generated using GBM, as well as data categorization with an efficiency of 87.21 %, once the network has been created.

In [27], the author used a deep learning-based approach to segment lung regions in chest X-rays. The self-awareness component of the suggested strategy is a first of its kind. With the use of the proposed attention modules, U-Net is being used to split lung areas. An experimental study found it outperformed conventional medical image segmentation networks. The recommended attention module's effectiveness was tested on U-Net in a series of trials. ResNet101 was employed as that of the backbone network in this experiment. Tests were carried out while the U-attention Net's module's position was altered. PyTorch was used to build both U-Net and also the attention module. According to the findings, X- and Y-attention modules can be used to enhance lung segmentation upon chest X-ray images. U-attention Net's module was tested in numerous different setups. Segmentation in the U-Net +X(1)+X(2)+Y(1)+Y(2) structure is superior to the other topologies. Here, Subject IDs, ages, genders, and findings are 138 unique in the Montgomery dataset. Pulmonary nodules are shown in 154 pictures, 93 pictures of healthy people are also included in the set compiled by the Japanese Society of Radiological Technology (JSRT). For Study ID - gender - age - results, there are 326 normal images and 336 aberrant images in the Shenzhen dataset.

In[3], the study shows the morphological technique for segmenting lung nodules from CT images termed an effective fuzzy auto-seed cluster. Initial cluster values were established via the average of the least and greatest pixel values at every row of an image. Kernel-based SVM classifier classification Fuzzy clustering is used to segment worrisome lesions in CT scans. Threshold segmentation is ineffective and unreliable.

The author employed a morphology-based auto seed regional grow method to split the nodules. An original fuzzy cluster framework is lightly adapted in this paper This facility saw 56 people with cancer (24 stage I and 32 stages II) and received and analyzed CT scan images from the scan facility (supplied by General Electric, New York, USA). Moreover, FACMM (Fuzzy Auto-seed Cluster Means Morphological) was selected for this study. The 2-D shape feature helps remove unwanted data clusters. A 3-D centroid analysis has been used to exclude blood vessels. Texture features can help remove calcifications. This study's sensitivity, specificity, as well as accuracy were 100%, 93%, and 94 percent, respectively.

The article[32] reexamines design spaces and tests a ConvNet. We progressively "modernize" a normal ResNet into a visual Transformer, revealing several key components along the way. ConvNeXts compete with Transformers in accuracy and scalability, attaining 87.8% ImageNet top-1 efficiency and surpassing Swin Transformers on COCO identification and ADE20K segmentation while maintaining the simplicity as well as efficiency of ordinary ConvNets. ConvNeXts outperforms hierarchical vision Transformers on computer vision benchmarks while keeping the simplicity as well as efficiency of standard ConvNets. Our ConvNeXt model isn't completely new, but our findings are. In the last decade, several design decisions were examined separately.

A deep learning network, termed as ConvNeXt, is used to extract radiomic features, and a malignancy scoring pooling technique is used to pool malignant scores in this paper. In particular, a deep convolutional neural network (CNN) is trained inside this manner of a visual transformer is used: the ConvNeXt network. As an additional perk, it shows how to combine the malignancy ratings of every breast US sequence frame using an efficient pooling strategy based on image-quality statistics. A database of 31 malignant as well as 28 healthy BUS sequences, each matching a patient, was used to build and assess the proposed CAD system. In this paper[31] it t was utilized in this experiments to assess single ultrasonic image CAD solutions with this approach using CNN networks (EfficientNetV2, EfficientNet-B7, MobileNetV3, and ResNet-101) (ViT, ResMLP and Swin transformer). Suggested techniques outperformed all ultrasound image CAD systems tested. This method's accuracy and F1 score were both over 90%. Furthermore, the F1-score of the proposed technique was 4% higher than that of the ConvNeXt-based single ultrasound image CAD system. Furthermore found that the quality of the BUS images affects the accuracy of malignancy prediction models. Malignancy score pooling improved classification accuracy by disregarding low-quality BUS images while generating the final malignancy score, as was indicated[8].

Chapter 2

Methodology

2.1 Working Plan

Machine learning modules are being utilized in this project to demonstrate how CT scans may be identified to detect lung cancer. For this, we need a strong dataset first. LUNA-16, a dataset compiled from 888 CT scans including 1,186 lung nodules, was used to test this new approach of screening. Our model is then tested as quickly as possible once it has been trained by removing any irrelevant features and scaling the pictures to their final size before adding and applying preprocessing filters. Proper training data must always be utilized to enhance the data in order to improve model validity and generalization.In order to increase the model's accuracy and generalization, we will need to supplement the data with additional training data. Different geometric adjustments, such as rotation, zooming, and flipping, were shown to be the most effective method of obtaining the required result. So we decided to divide the data into two pieces so that we can now train as well as test our model in two distinct sessions. As a consequence, our model goes through a process of acquiring information from the data. The network was then evaluated against our testing data, and its accuracy was computed and then compared with those of training data models.

2.2 Data collection:

CT scans are by far the most accurate imaging approach for lung cancer screening because they identify all nodules, regardless of even if they're not suspected of containing cancerous cells. LUNA16 dataset is used, and it contains 1,186 lung nodules that have been found in 888 CT scans. We utilized this dataset to conduct our research. Non- Nodules, nodules that are 3 mm in diameter, and nodules more than 3 mm in diameter seem to be the three kinds of lesions. During the span of a two-phase annotation technique, the lesions are annotated by four highly qualified radiologists. Slices having a thickness of more than 2.5 mm are deemed unsuitable for further examination and testing, which even might be ruled out of further study.

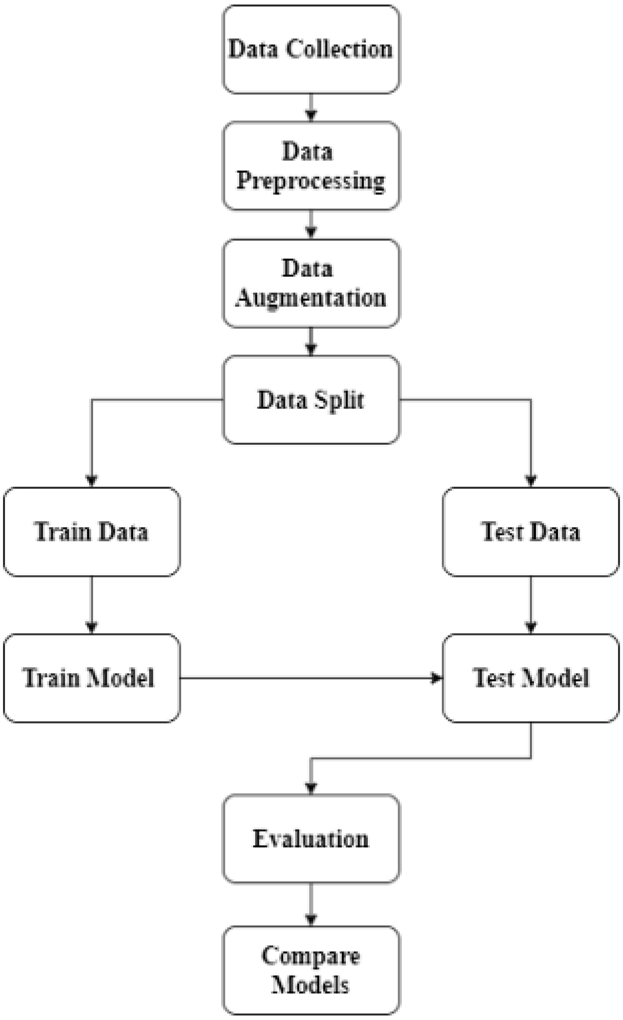

Figure 2.1: A Block Diagram of Working Plan

2.3 Data Analysis and Preprocessing:

On the LUNA16 website, ten subsets of that dataset are accessible. CT images are stored in each subset using the MetaImage format. The pixel data is stored in one—Raw binary file per .mhd. The ”Slicer 4.11” software has been used to analyze the evidence of each patient. We initially picked meaningful segments from 3D volumetric data by using the largest inter-class variance (ICV). To begin, the photos are read as well as resized to 256 x 256 pixels. The images are then gray-scaled once they've been resized. A median filter is also used to reduce noise, accompanied by a Gaussian filter that smooths out and impulses noise. Lastly, for segmentation as well as edge detection, binary thresholding is applied.

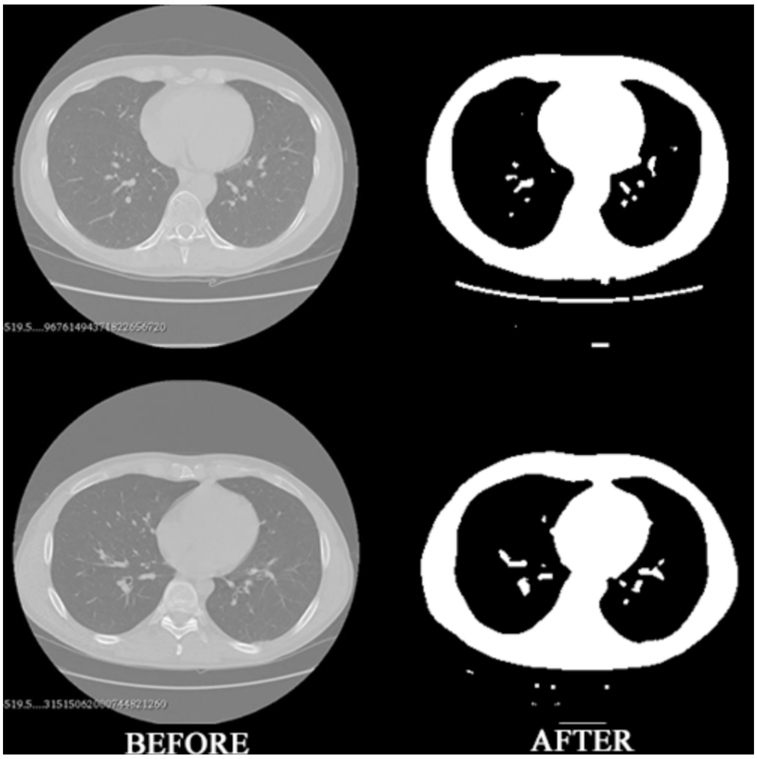

Figure 2.2: Image before and after Preprocessing

Chapter 3

Proposed Methods

3.1 Convolutional Neural Network(CNN):

CNN, or Convolutional Neural Network, is a deep learning strategy supported by the filtering and resizing layer known as the convolutional layer. This layer, also known as the filtering or convolutional layer, is responsible for extracting numerous characteristics from the input images, such as colour, edges, and other characteristics. The filter's dimensions can range from 3x3 to 5x5 to 7x7. A component called a stride, which determines how pixels are processed, also contributes to this unique dimension's formation. Reducing the layer's size through layer pooling or resizing, on the other hand, reduces its computational complexity[5]. Using this layer, the convolved feature map can be compressed in three distinct ways to save computing expenses. For example, Max pooling determines the maximum number of feature points, Min pooling determines the lowest number of feature points, and Average pooling determines the average number of feature points from the neighbourhood. Each layer decides the output by applying a particular function to the input data and remapping the weight or bias of each neuron accordingly. However, since many input pictures may include non-linearity due to dust or other reasons, the Relu layer in CNN augments the input image with non-linear characteristics to make the training more realistic and, as a consequence, provide better long-term outcomes. Finally, the final layer's output images will be flattened in a connected layer that uses normalization to prevent over-fitting while also speeding up training and learning. In most situations, this fully connected layer is used to categorize data at the network's endpoint[5].

Different Layers of CNN:

Input Layer:

The input layer holds all the information from the CNN. The input layer of a neural network is the layer on the left side of the network. The pictures collected from our dataset must be sent into this layer of CNN as inputs. In image processing, it is generally used to represent the pixel matrix of an image in a neural network. The high-level convolutional layer discovers abstract features by incorporating low-level information. The pictures collected from the dataset are sent into this layer of the CNN as inputs. The input layer in the case of greyscale images contains three copies of the same single map, but the input layer for colour images contains three copies of three different maps. For each image, we receive three different dimensions, which are (width x height x depth). The summation of these dimensions is represented by the pixel values. The height and width of the image that was used as an input indicate the size of the image.

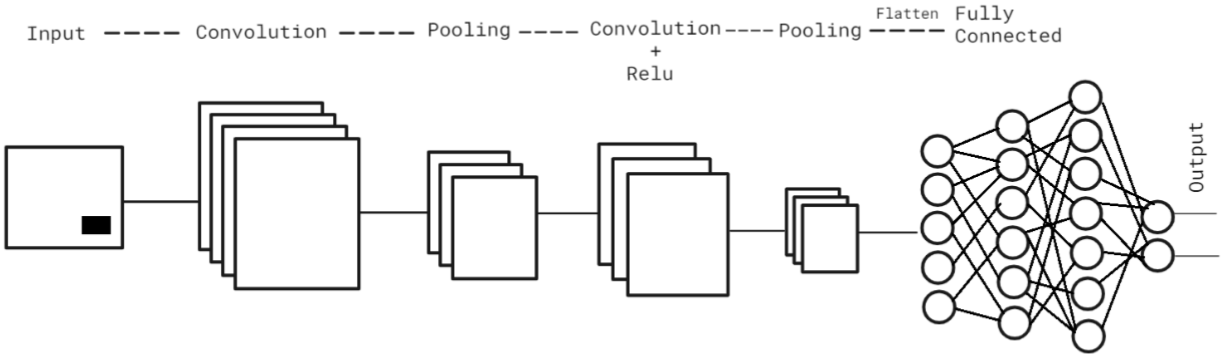

Figure 3.1: Convolutional Neural Network(CNN)

Convolution Layer:

The CNN consists of several layers. Among them, the first one is the convolutional layer. This layer extracts attributes from the input pictures. This layer uses the input image and an MxM filter to perform a mathematical convolution between the two images. It is important to slide the filter over the input picture in an amount corresponding to the filter size in order to acquire the dot product between the filter and the input image proportional to the filter size (MxM). The generated feature map is used to represent the image and provide information about it, such as its corners and edges. This feature map is utilized by later layers to comprehend a range of features about the picture that was used as input. Convolution is a linear procedure that involves applying a kernel to a stack of data as input. The feature map is created by multiplying the result by the output tensor after adding the kernel and input tensor element-wise products to the output tensor. Several kernels are used to build an arbitrary number of feature maps, each of which represents a distinct characteristic of the input tensors; the overall number of feature maps is also arbitrary. The most important parts of convolutional layers are these two. In order to build a convolutional layer in the CNN model, pick and merge the best-performing kernels from a training dataset. Kernel size, number of kernels, padding, and stride are all hyperparameters that must be determined before training the convolution layer. Recent CNN designs construct layers without padding, allowing for the maintenance of in-plane proportions. Kernels must be shared across all picture locations if convolution is to use weight sharing. Furthermore, by dispersing weights among kernels, local feature patterns returned by kernels become translation invariant, enabling them to learn spatial hierarchies of feature patterns through downsampling and pooling, resulting in a larger field of view[5]. Zero padding is widely used in the design of modern CNNs in order to accommodate extra layers while keeping the network's in-plane dimensions. The quantity of data available after convolution would be lower if padding were deleted from the convolution process. Striders are the distance between two kernel points in mathematics, and they are defined as follows. Stride assists CNN in decreasing these detrimental effects by lowering the number of parameters employed. Convolution applies the same weights throughout the image instead of applying distinct weights to each region in the image, which results in more consistent output. Kernels can travel through all image locations and detect localized learned characteristics that have been learned over time thanks to the weight-sharing approach. It is possible to get information on the spatial hierarchy of feature pattern patterns through the use of downsampling and pooling, as well as an increase in kernels. To put it another way, the algorithm's output is a two-dimensional activation map. Stacking the activation maps on top of the depth dimension array yields the output volume[5].

Pooling Layer:

In addition to the network and transport layers, there is a pooling layer that is used for pooling. To reduce the number of parameters and calculations that must be performed in the network, the spatial dimension of the representation must be reduced over time. In a convolutional neural network, a pooling layer is an additional layer that can be added after the convolutional layer to improve performance. Furthermore, feature maps formed by a convolutional layer are recognized as being of interest when a nonlinearity (for example, ReLU) is applied to them. This layer, like the pooling layer, treats each feature map as if it were a separate entity, similar to how the pooling layer does. These are commonly used to reduce the number of dimensions in a network to a bare minimum. A pooling layer is frequently included after the convolutional layer in a convolutional neural network to achieve layer order in a convolutional neural network[5]. It is possible for this pattern to be repeated one or more times inside a model. This is because the pooling layer operates on each feature map separately, resulting in a new set of pooled feature maps that include the same features as the original set of feature maps. Pooling is accomplished through the selection of a pooling operation, which is analogous to the selection of a feature map filter in a feature map. A little amount of space is taken up by the pooling operation or filter, which is frequently 2x2 pixels in size with a stride of 2 pixels and is almost always 2x2 pixels in size[5]. When the pooling layer is used to generate a feature map, the size of the feature map is always reduced by a factor of two; for example, the dimensions of a feature map are always reduced by half, and the number of pixels or values in a feature map is always reduced to one-quarter of its original size. Applying a pooling layer to a feature map with a resolution of 6x6 (36 pixels) will result in a pooled feature map with a resolution of 3x3 (three times the original map size), as seen in the following example (9 pixels). The following are two of the most often used functions in the pooling operation:

Average Pooling:

Using the average Pooling function, calculate the average value for each feature map patch.

Max Pooling:

Calculate the maximum value for each feature map patch using average pooling.

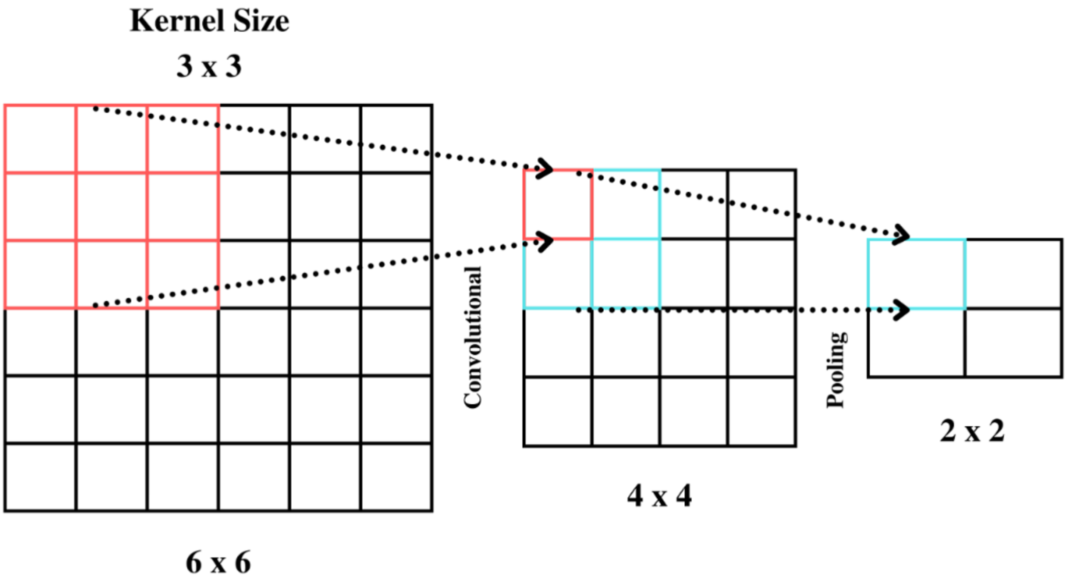

Figure 3.2: Pooling process with 2 x 2 filters

Fully Connected Layer:

Fully Connected Levels are the very last and most critical levels of the network architecture. The output of the final Pooling or Convolutional Layer is fed back into the final Pooling or Convolutional Layer by the entirely connected layer, which is a flattened form of the original output. In most cases, once the output feature maps of the final convolution or pooling layer have been flattened (converted to a one-dimensional (1D) array of numbers (or vector), they are connected to one or more fully connected layers, also known as dense layers, in which each input is connected to each output by a learnable weight[9]. It is common practice in classification tasks to use a subset of fully connected layers to transport features that have been extracted by convolution layers and downscaled by pooling layers to the network's final outputs, which, in classification tasks, are probabilities for each classification class.In the vast majority of cases, the number of output nodes in the final fully connected layer is equal to or more than the number of classes in the final partially connected layer, depending on the situation. Each layer is followed by a nonlinear function that is completely related to the previous layer[9].

3.2 Residual Attention Neural Network:

As a way to train our model, we're exploring employing the Residual Attention Network(RAN), which would be extremely intriguing. It is possible to enhance the network's feature representation by utilizing our Residual Attention Network (RAN). By focusing on CNN, researchers hoped to create a network that could recognize objects and learn about their properties. Stacked residual blocks and an

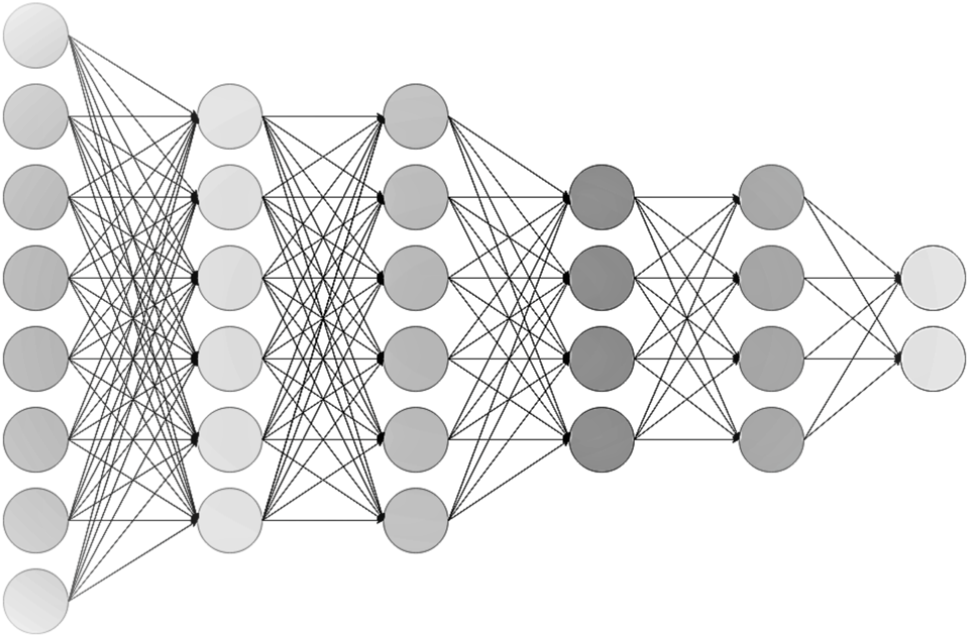

Figure 3.3: Fully Connected layer

attention module allowed for the development of a feed-forward CNN. A Residual Attention Network is formed when multiple attention modules are stacked on top of one other (RAN). In the attention module, there are two branching structures: the soft mask and the trunk. The soft mask branch is positioned on one side of the structure, while the trunk branch is located on the opposite side. Using the fast-feed forward methodology, the attention module combines all of the attention feedback methods it uses into a single feed-forward operation While the soft mask branch serves as a backpropagation gradient update filter in the trunk branch. A bottom-up feed-forward approach is used to create low-resolution feature maps with rich semantic information. A top-down approach is used to create dense features to infer each pixel.
The ability of RAN to distinguish crowded, difficult, and noisy pictures was increased by combining various attention modules.

3.3 Attention-based Multi-instance Neural Network:

Attention-based Multi-instance Neural Network as a means of training a neural network to improve the model even further. It also has an embedding layer and a multi-head attention transformer with the residual connection, as well as a bag-level MIL pooling layer and a series of instance-wise completely connected layers, all of which are part of the same structure and gradually followed by a sigmoid function [22]. Using the multi-head attention transformer in the MIL neural network to acquire the intra-relationship of instances placed in different embedding subspaces of the MIL neural network. This is especially useful in the medical field because, despite the fact that symptoms in distinct body parts or organ systems are commonly related, each individual body component or organ system can be considered as an

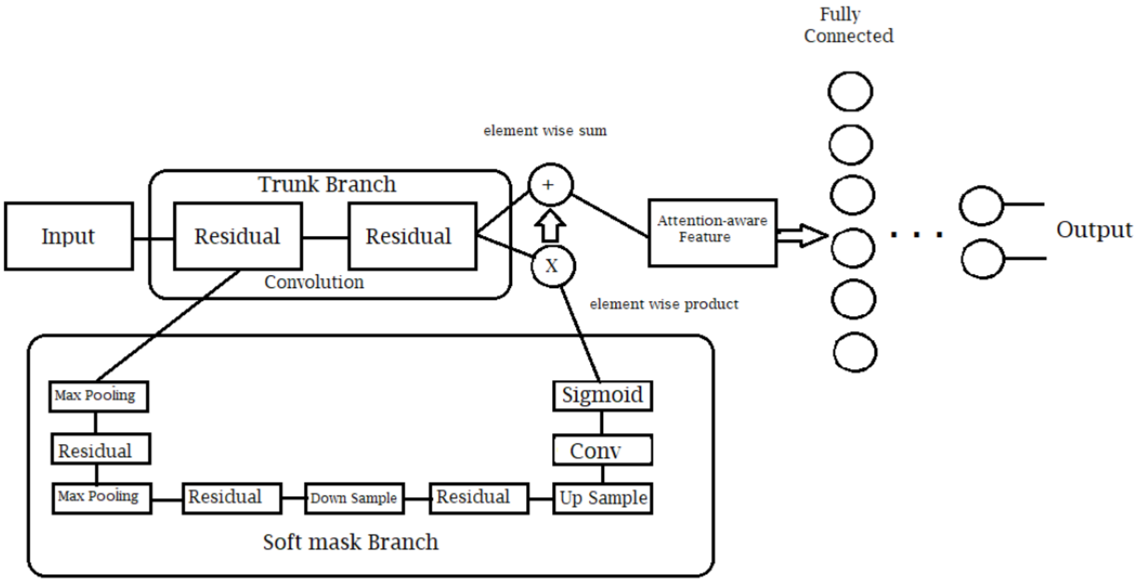

Figure 3.4: Residual Attention Neural Network

independent subspace. The number of subspaces is determined with multi-head attention transformers, and scaled dot-product attention is performed on each subspace at the same time as instance correlations are simultaneously recorded in each of the subspaces. MIL pooling is a vital stage in the process of linking instances and bags, and different applications have varied preferences when it comes to MIL pooling approaches [22]. The max-pooling, mean pooling and log-sum-exp pooling algorithms are being implemented on each individual instance as time permits, as the MIL pooling deadline for instance-level MIL pooling approaches. When combined with bag scores, attention-based MIL pooling can help to boost bag scores even more. The sigmoid function will then be used to turn it into the probability of a bag being filled[13]. By reducing the need for human data collecting and filtering and dealing with them manually rather than automatically. It can extract the most essential information from a large number of low-quality data, saving costs. Its greatest strength is large-scale data collecting, but it is also its greatest strength in small-scale and large-scale data collection[13].

3.4 ConvNeXt:

We're looking into the prospect of using ConvNeXt as a means of training our neural network to improve our model even further. In terms of performance, ConvNeXT is expected to surpass Vision Transformers' ConvNet. Swin Transformers make use of a multi-stage layout with varied feature map resolutions when it comes to the macro design[10]. The following are the two most important considerations to make about the design: The ratio is determined based on ”stem cell” structure as well as stage structure. The ”stem” cell is the one that is in charge of the first image processing that the network does [33]. The FLOPs versus accuracy compromise that ResNeXt requires is best addressed by ResNet. Through the use of depth-wise

convolution, network FLOPs may be lowered while accuracy can be enhanced. An additional essential component of the Transformer block is the inverted bottleneck that is produced by the MLP block. This bottleneck has a concealed dimension that is four times larger than the dimension that is being input. There is a higher number of FLOPs generated by depthwise convolution, and it is possible to reduce the total number of FLOPs generated by a ConvNeXt network by employing an inverted bottleneck design. Vision Transformers are distinguished from other Transformers by their non-local self-awareness[8]. Convolutional networks are only able to see local features that are contained within the size of the kernel.
For the purposes of this research, ResNet-50 will serve as our starting point. When it is finished being trained, it performs far better than ResNet-50 did when judged against the first technique of training. This will serve as our point of departure going forward. There are a number of design considerations that we examine in further depth, including an inverted bottleneck, a large kernel size and many layer-wise micro designs. We refer to them as ”macro design,” ”ResNeXt,” and ”macro design[8].”

Due of BatchNorm’s ability to accelerate convergence and reduce overfitting, ConvNets would be unable to work correctly without it. BN, on the other hand, comprises several complexity, all of which might have a detrimental influence on the model’s performance. BN has remained the technique of choice for the bulk of vision tasks despite several attempts to develop other normalizing methods. In Transformers, on the other hand, the more easy Layer Normalization (LN) method has been applied, resulting in an outstanding performance in a range of application conditions[8]. The performance of the original ResNet will be degraded if BN is replaced with LN directly. As a result of the adjustments made to training methodologies and the network architecture, we will now examine the benefits of using LN rather than BN. As we’ve shown, training our ConvNet model using LN isn’t problematic at all.
Another aspect of Transformers’ microscale architecture that ConvNeXt inherits is that of layer-by-layer design. This feature may be found in ConvNeXt’s microscale architecture. Moreover, a Gaussian Error Linear Unit, known as GELU, is used in place of the ReLU activation function because it is smoother. By deleting all of the GELU layers first from the residual block except for one, it creates an effect that is similar to that of the Transformer block. GELU can be expressed as follows for an input z,[7]

$$\mathrm{GELU(z) = zP(Z \leq z) = z\phi(z) = z} \cdot \tfrac{1}{2}\ [1 + \mathrm{erf}(\tfrac{z}{\sqrt{2}})]$$

LN is used in ConvNeXt to avoid the drawbacks of batch normalization, which is widely used in deep CNN architectures (e.g., computational cost and discrepancy between training and inference). As a result of setting the mean and variance of summed inputs for each layer, LayerNorm may remove the covariate shift issue. Changes to the output from one layer often have a direct effect on the inputs utilized by the next layer, so this is done with this in mind [6].
Furthermore, ConvNeXt also employs depthwise convolution, which is a kind of group convolution in which the number of channels and groups is proportionate to one another. The depthwise convolution process is, in fact, quite similar to the per-channel weighted sum operation that is utilized in the self-awareness mechanism

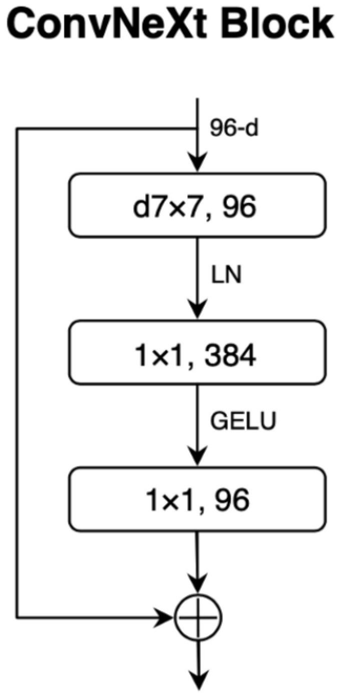

Figure 3.5: Schematic diagram of ConvNeXt block.

(mixing information in the spatial dimension). ConvNeXt adds a separate downsampling layer in between the various conversion stages. It uses a stride of two and a 2 x 2 Conv layer structure for downsampling [15].

Every transformer block in a swine transformer has an inverted bottleneck. Concatenating the output of four blocks increases the hidden dimensions by four times. To achieve a fourfold growth ratio, ConvNeXts used this strategy to create an inverted bottleneck. Because of this, the model performs better [12].

The kernels must be increased in size if vision transformer models with a global receptive field are to have the same power. The vision transformer's seeing style is characterized by a level of self-attention that embraces the whole image. According to research, those that use Swin transformers have a reduced window of self-attention. If the ResNets window is expanded bigger, the same tradeoff is achieved [32].

Fewer activation functions are applied in our model. Transformers have fewer activation functions than ResNet blocks, but that's about the only difference between the two. Additionally, take into account the two linear layers and the key/query/value linear embedding layers of an MLP block with a transformer block. Only one activation function may be contained in the MLP block. On the other hand, attaching an activation function to each convolutional layer, even the 1 x 1, is a typical procedure. This is done to enhance the model's accuracy [32].

Chapter 4

Result

4.1 Result and Analysis

Google colab was used to run each and every one of the models. The first step of the research was the gathering of data for the purpose of training models. This dataset included those of both males and females. A total of around 348 images were obtained, of which 300 were used for training and the remaining 48 were used for testing in accordance with the instructions. After the images have been trained, they are passed on to the classification stage, where each image is evaluated individually and categorized as either normal or malignant. The output of each image, along with its accuracy in determining whether it is benign or cancerous, is shown. The following is the formula that is used to determine accuracy:

$$\textbf{Accuracy: } \frac{TP+TN}{TP+TN+FP+FN}$$

Parameters	Numbers
Trained Images	300
Tested Images	48
True Positive	10
True Negative	33
False Positive	1
False Negative	4
Accuracy	**89.58%**

Table 4.1: Measuring accuracy using ConvNeXt.

As a result, ConvNeXt is used to produce the output, which comprises image detection and classification. The results were determined to be accurate 89.58% of the time, the results were shown in table 4.1 and figure 4.1.

The algorithms were also evaluated based on their precision, recall, and F-measure scores, all of which are metrics that are often used in the fields of text mining and machine learning. The four different kinds of classified items are true positive (TP – objects that have been correctly labelled as belonging to the class), false positive (FP – items that have been falsely labelled as belonging to a certain class), false negative (FN – items that have been incorrectly labelled as not belonging to a certain class),

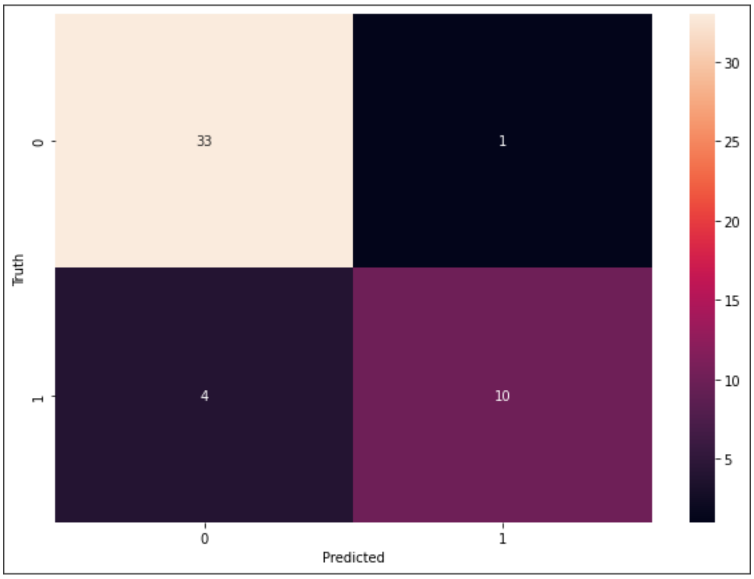

Figure 4.1: Confusion Matrix

and true negative (TN – items that have been correctly labelled as not belonging to a certain class) (TN -items correctly labelled as not belonging to a certain class). It is possible to calculate recall by using the following formula, which takes into account the number of true and false positives [24].

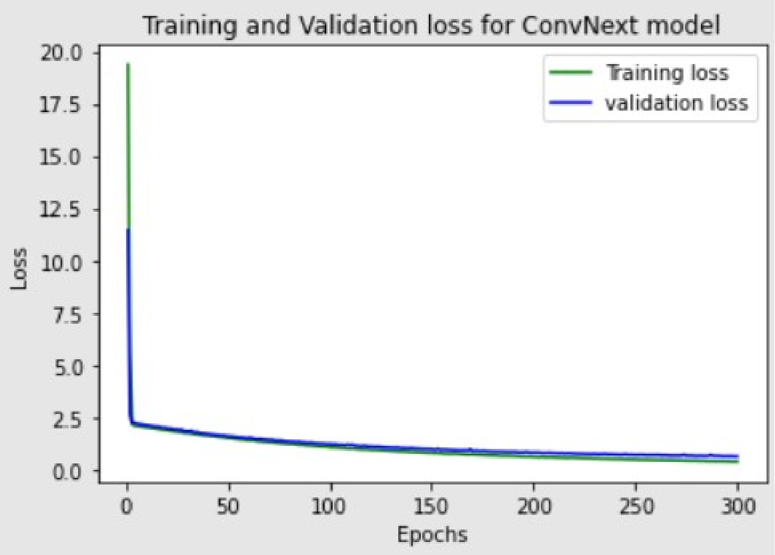

Figure 4.2: Training and validation loss for ConvNeXt

$$\textbf{Recall}= \frac{TP}{TP+FN}$$

In certain circles, the recall is also referred to as the ”sensitivity” or the ”absolute positive rate” [24]. Precision, also known as ”positive predictive rate,” is determined by comparing the number of graded items that are true positives to those that are false positives in the following way:

$$\textbf{Precision}=\frac{TP}{TP+FP}$$

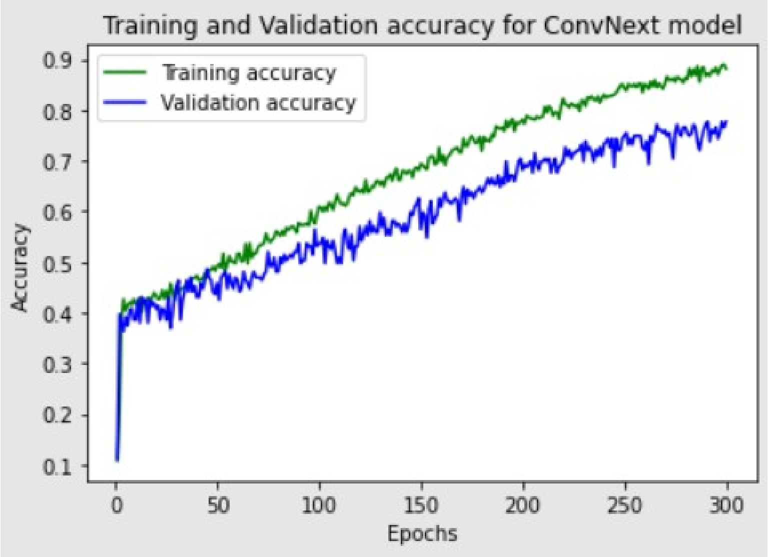

Figure 4.3: Training and validation accuracy for ConvNeXt

The F-measure is a measurement that combines precision and recall, and its definition is as follows:

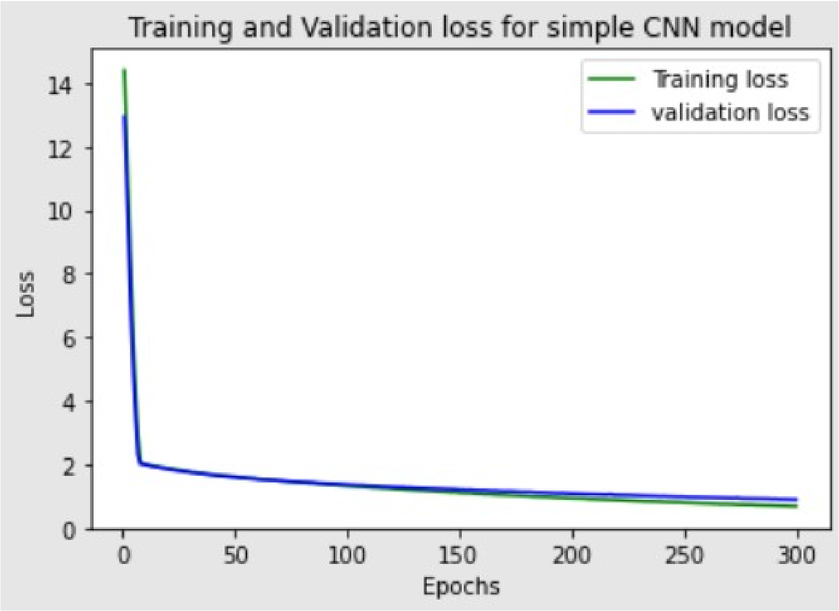

Figure 4.4: Training and validation loss for CNN

$$\mathbf{F}= \frac{((1+\beta)\cdot Recall\cdot Precision)}{\beta x(Precision+Recall)}$$

where represents the precision's relative value and indicates the relative value. Recall and precision are considered to be of equal significance when the value is set to 1, which is a common configuration. When the number is lower, it means that accuracy is more significant, and when the value is greater, it shows that recall is more important. [24].

Method	Precision	Recall	F-Measure
CNN	84.65%	83.53%	84.25%
ConvNext	90.90%	71.43%	80%

Table 4.2: Measuring overall precision, recall and F- Measure

The outcomes of the models in terms of CNN and ConvNeXt are as follows: In the results area, both the accuracy and the precision that occurred during training and validation are presented.

The accuracy and precision of CNN and ConvNeXt can both be seen in this comparison, which is shown in figure 4.5. Comparatively, ConvNeXt has an accuracy of 89.58%, while CNN's is just 84.72%. Also, we find that CNN has a precision of 84.65%, whereas ConvNeXt has 90.90%.

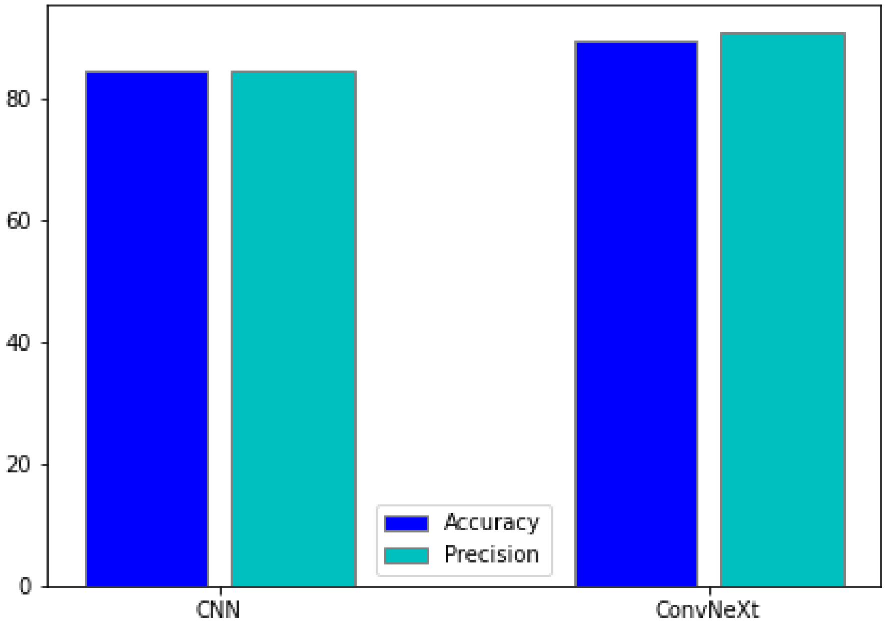

Figure 4.5: Accuracy and Precision Analysis.

Experiments show that the ConvNeXt backbone improves model performance even though the channel size of every stage is greatly decreased, showing that redundancy in the channel dimension still exists. It will be easier for people to find architectural channels in the future. Use a neural architecture approach to jointly increase performance as well as a throughput to make it more feasible to utilise it on edge devices.

The accuracy and loss data are recalculated after every 30 epochs. The precision of the training has been steadily improving, while the amount of loss experienced during training has been steadily decreasing. To begin with, the model is presented with a validation set, which will be put to use once its training has been completed in order to construct the appropriate change network. If the validation accuracy of a model is the same as the training accuracy, then the model is considered to be superior or a near approximation of the training accuracy. Training is another

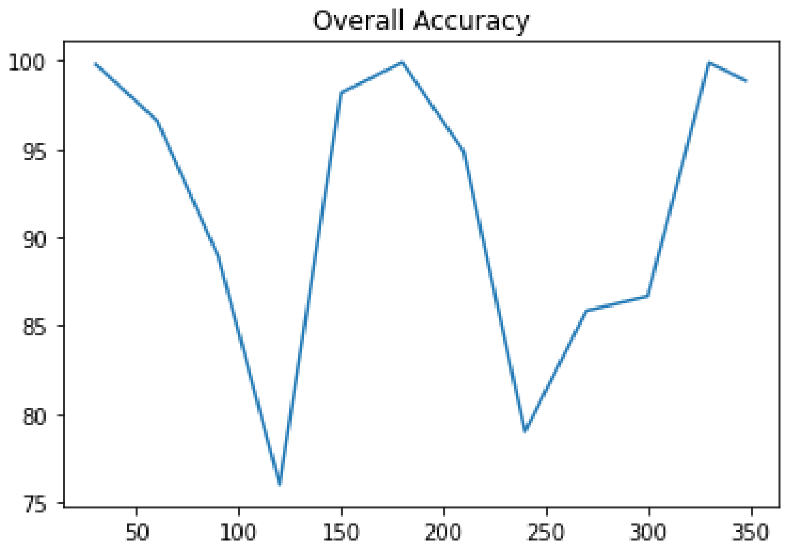

Figure 4.6: Overall accuracy by averaging each training over time.

essential component of the overall process. Validation's accuracy deteriorates as the loss grows. The ConvNeXt model's first training accuracy was not very great at the start of the training process. The accuracy of the validation is lower than that, which shows that the model was not calibrated in the correct manner. To put it more simply, the model already has the training data but needs validation. The datasets are completely brand new. The model has never been exposed to new sets of data previously. In order to make the model more effective, the system now has an increased number of neurons as well as layers. The more time and practice you put in, the more accurate your training will become.

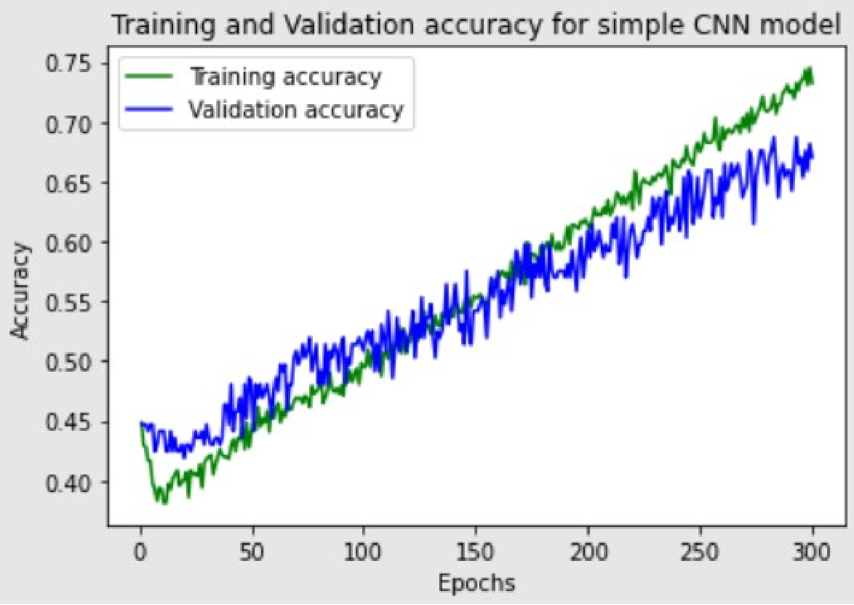

Figure 4.7: Training and validation accuracy for CNN

Figure 4.6 and 4.3 presents an illustration of the fluctuating accuracy of training over time. In addition, the validity of the validation is not very reliable. There have been moments when people have questioned whether or not the training is accurate and precise. Through validation, similar degrees of precision can be attained. As the eons progressed, there was a rise in the quantity of instruction.

After training the model, it was found that the accuracy increased to 89.58 percent, demonstrating that there was an overall gain in model accuracy after each training period. According to this figure 4.3, we are able to conclude that if we train our model using a large number of datasets, our model will have an accuracy that is superior to this.

4.2 Future Work

The framework we have provided for categorising lung cancer into various basic types may be modified in several ways. It's possible, for example, to add more advanced patient data that wasn't available at that time into our study, try out newer CNN designs, and allow anybody to review our data and results. Instead of using a 2D picture, the categorisation might benefit from including the patient's whole 3D CT image. Additional information on the patient's medical history, their DNA sequence, and more might be incorporated into this study in order to get a complete picture of how this disease affects them. Additional information on the models that have been used to integrate lung cancer detection is the major goal of this research. By comparing all of the models that have been implemented, we are able to establish the precise outcome of the system, which also gives us a broad sense of how accurate each model is. A more advanced preprocessing method may be used in order to enhance the clarity and reduce artefacts in CT images. It's feasible that a more modern and effective preprocessing strategy might help enhance things like accuracy. During the past several years, the area of deep learning for computer vision has grown tremendously as the speed of GPUs has improved significantly. Experimental results reveal that the ConvNeXt backbone enhances model performance even when the channel size of each stage is considerably reduced, suggesting that redundancy in the channel dimension still exists. This means that the architectural channels will be searchable in the future.

4.3 Limitations

ConvNeXt, a pure ConvNet model, outperforms a hierarchical vision transformer on image classification, object recognition, for instance, as well as semantic segmentation. Computer vision applications are far more diverse than our evaluation assignments. Transformers could be better than ConvNeXt for some jobs. Multi-model learning may benefit from a cross-attention module that mimics feature interactions across senses. Transformers are versatile for discontinuous, sparse, or organized outputs. The chosen architecture should be simple but satisfy the project's demands.

4.4 Conclusion

A neural network architecture known as convNeXt was introduced in this study with the purpose of identifying lung cancer in its earliest stages. The performance of the ConvNeXt network is superior to that of older CNN models. The accuracy of the models employed to predict the existence of malignant or benign cells is affected by the quality of the dataset. In this research, an increase in model performance is achieved by changing the structure of the feature extraction backbone in ConvNeXt. According to the findings of this research, a radiomics method that makes use of deep learning accurately identifies the malignant and benign potential of lung cancer cells. Deep learning for computer vision can achieve expert-level accuracy with a pre-trained neural network. This technique might automate biomedical and medical imaging operations with human-like precision. Our technology may be utilized to achieve significant leaps forward in terms of accuracy in a variety of difficult areas of medical imaging. However, the human experience and the model that we tested are the only things that restrict the accuracy of our predictions and the results. Unless significant advances have been achieved in human detection; therefore, there will be no further progress on this subject. Because of the circuitry of the neural network, we do not have a clear understanding of the cause-and-effect links that exist in the data or in the classification. As a direct consequence of this, we are able to rapidly detect and classify the onset of lung cancer, in addition to performing classification. According to our research, pre-training ConvNeXt models on huge datasets is beneficial in terms of data. While our method relies on the freely accessible Luna-16 dataset, some users may prefer to pre-train and use their own data. A more thoughtful and appropriate approach to data collection is required in order to remove any data biases.

Bibliography

[1] K. Awai, K. Murao, A. Ozawa, M. Komi, H. Hayakawa, S. Hori, and Y. Nishimura, “Pulmonary nodules at chest ct: Effect of computer-aided diagnosis on radiologists’ detection performance,” *Radiology*, vol. 230, no. 2, pp. 347–352, 2004.

[2] V. Ambrosini, S. Nicolini, P. Caroli, C. Nanni, A. Massaro, M. C. Marzola, D. Rubello, and S. Fanti, “Pet/ct imaging in different types of lung cancer: An overview,” *European journal of radiology*, vol. 81, no. 5, pp. 988–1001, 2012.

[3] S. Sivakumar and C. Chandrasekar, “Lung nodule detection using fuzzy clustering and support vector machines,” *International Journal of Engineering and Technology*, vol. 5, no. 1, pp. 179–185, 2013.

[4] J. Cabrera, A. Dionisio, and G. Solano, “Lung cancer classification tool using microarray data and support vector machines,” in *2015 6th International Conference on Information, Intelligence, Systems and Applications (IISA)*, IEEE, 2015, pp. 1–6.

[5] K. O’Shea and R. Nash, “An introduction to convolutional neural networks,” *arXiv preprint arXiv:1511.08458*, 2015.

[6] J. L. Ba, J. R. Kiros, and G. E. Hinton, “Layer normalization,” *arXiv preprint arXiv:1607.06450*, 2016.

[7] D. Hendrycks and K. Gimpel, “Gaussian error linear units (gelus),” *arXiv preprint arXiv:1606.08415*, 2016.

[8] B. Abdillah, A. Bustamam, and D. Sarwinda, “Image processing based detection of lung cancer on ct scan images,” in *Journal of Physics: Conference Series*, IOP Publishing, vol. 893, 2017, p. 012 063.

[9] S. Albawi, T. A. Mohammed, and S. Al-Zawi, “Understanding of a convolutional neural network,” in *2017 international conference on engineering and technology (ICET)*, Ieee, 2017, pp. 1–6.

[10] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” *Advances in neural information processing systems*, vol. 30, 2017.

[11] E. R. Velazquez, C. Parmar, Y. Liu, T. P. Coroller, G. Cruz, O. Stringfield, Z. Ye, M. Makrigiorgos, F. Fennessy, R. H. Mak, *et al.*, “Somatic mutations drive distinct imaging phenotypes in lung cancer,” *Cancer research*, vol. 77, no. 14, pp. 3922–3930, 2017.

[12] A. Howard, A. Zhmoginov, L.-C. Chen, M. Sandler, and M. Zhu, “Inverted residuals and linear bottlenecks: Mobile networks for classification, detection and segmentation,” 2018.

[13] M. Ilse, J. Tomczak, and M. Welling, "Attention-based deep multiple instance learning," in *International conference on machine learning*, PMLR, 2018, pp. 2127–2136.

[14] S. Makaju, P. Prasad, A. Alsadoon, A. Singh, and A. Elchouemi, "Lung cancer detection using ct scan images," *Procedia Computer Science*, vol. 125, pp. 107–114, 2018.

[15] Z. Zhang and M. Sabuncu, "Generalized cross entropy loss for training deep neural networks with noisy labels," *Advances in neural information processing systems*, vol. 31, 2018.

[16] T. Sajja, R. Devarapalli, and H. Kalluri, "Lung cancer detection based on ct scan images by using deep transfer learning.," *Traitement du Signal*, vol. 36, no. 4, pp. 339–344, 2019.

[17] J. Sang, M. S. Alam, H. Xiang, *et al.*, "Automated detection and classification for early stage lung cancer on ct images using deep learning," in *Pattern Recognition and Tracking XXX*, International Society for Optics and Photonics, vol. 10995, 2019, 109950S.

[18] J. Wang, R. Gao, Y. Huo, S. Bao, Y. Xiong, S. L. Antic, T. J. Osterman, P. P. Massion, and B. A. Landman, "Lung cancer detection using co-learning from chest ct images and clinical demographics," in *Medical Imaging 2019: Image Processing*, SPIE, vol. 10949, 2019, pp. 365–371.

[19] A. Elnakib, H. M. Amer, and F. E. Abou-Chadi. (2020). "Early lung cancer detection using deep learning optimization," [Online]. Available: https://www.learntechlib.org/p/217904/.

[20] A. Kulkarni, K. Jadhav, A. Mishra, and A. Kumar. (2020). "Lung cancer detection using deep convolutional neural network," [Online]. Available: https://sersc.org/journals/index.php/IJFGCN/article/view/28246.

[21] C. Li, Y. Tan, W. Chen, X. Luo, Y. Gao, X. Jia, and Z. Wang, "Attention unet++: A nested attention-aware u-net for liver ct image segmentation," in *2020 IEEE International Conference on Image Processing (ICIP)*, IEEE, 2020, pp. 345–349.

[22] A. Sadafi, A. Makhro, A. Bogdanova, N. Navab, T. Peng, S. Albarqouni, and C. Marr, "Attention based multiple instance learning for classification of blood cell disorders," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2020, pp. 246–256.

[23] H. M. Tayade. (2020). "Early detection of laryngeal cancer using multiple instance learning based neural network," [Online]. Available: http://norma.ncirl.ie/4475/1/harshalmilindtayade.pdf.

[24] D. M. Abdullah, A. M. Abdulazeez, and A. B. Sallow, "Lung cancer prediction and classification based on correlation selection method using machine learning techniques," *Qubahan Academic Journal*, vol. 1, no. 2, pp. 141–149, 2021.

[25] T. L. Chaunzwa, A. Hosny, Y. Xu, A. Shafer, N. Diao, M. Lanuti, D. C. Christiani, R. H. Mak, and H. J. Aerts, "Deep learning classification of lung cancer histology using ct images," *Scientific reports*, vol. 11, no. 1, pp. 1–12, 2021.

[26] T. A. M. Devi and V. M. Jose, “Three stream network model for lung cancer classification in the ct images,” *Open Computer Science*, vol. 11, no. 1, pp. 251–261, 2021. [Online]. Available: https://www.degruyter.com/document/doi/10.1515/comp-2020-0145/html.

[27] M. Kim and B.-D. Lee, “Automatic lung segmentation on chest x-rays using self-attention deep neural network,” *Sensors*, vol. 21, no. 2, p. 369, 2021.

[28] X. Lu, Y. Nanehkaran, and M. Karimi Fard, “A method for optimal detection of lung cancer based on deep learning optimized by marine predators algorithm,” *Computational Intelligence and Neuroscience*, vol. 2021, 2021.

[29] V. K. Vipparla, P. K. Chilukuri, G. B. Kande, *et al.* (2021). “Attention based multi-patched 3d-cnns with hybrid fusion architecture for reducing false positives during lung nodule detection,” [Online]. Available: https://www.scirp.org/html/1-1731512_108272.htm.

[30] L. Eldridge. (2022). “Stage 4 lung cancer life expectancy,” [Online]. Available: https://www.verywellhealth.com/what-is-stage-4-lung-cancer-life-expectancy-2249420.

[31] M. A. Hassanien, V. K. Singh, D. Puig, and M. Abdel-Nasser, “Predicting breast tumor malignancy using deep convnext radiomics and quality-based score pooling in ultrasound sequences,” *Diagnostics*, vol. 12, no. 5, p. 1053, 2022.

[32] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, “A convnet for the 2020s,” *arXiv preprint arXiv:2201.03545*, 2022.

[33] A. Singh. (Mar. 2022). “Convnext: The Return Of Convolution Networks | by Aditya Singh | Augmented Startups | Medium.” [Online; accessed 2022-05-23].