Optimal Emerging trends of Deep Learning Technique for Lung Cancer Detection based on Convolutional Neural Network

Final Year Project

SESSION 2020-2024

A project submitted in partial fulfillment of the degree of

BS in Computer Science

Department of Computer Science

Faculty of Computer Science & Information Technology

The Superior University, Lahore

Spring 2024

Type (Nature of project)	[ ] **D**evelopment [▯] **R**esearch [ ] **R&D**
Area of specialization	Artificial Intelligence and Deep Learning
FYP ID	FYP-BCSM-F23-034

Project Group Members

Sr.#	Reg. #	Student Name	Email ID	*Signature
(i)	BCSM-F20-102	Abdullah Sajid	Bcsm-F20-102@superior.edu.pk	
(ii)	BCSM-F20-027	Ammar Hassan	Bcsm-F20-027@superior.edu.pk	

*The candidates confirm that the work submitted is their own and appropriate credit has been given where reference has been made to work of others

Plagiarism Free Certificate

This is to certify that, I Abdullah Sajid S/D of Sajid Mehmood, group leader of FYP under registration no FYP-BCSM-F23-034 at Computer Science Department, The Superior University, Lahore. I declare that my FYP report is checked by my supervisor.

Date: 18-April-2024 Name of Group Leader: Abdullah Sajid Signature: _____________

Name of Supervisor: Dr. Hamayun Khan

Designation: Assistant Professor Signature: _____________

HoD: Dr. Irfan ud din Signature:______________

Optimal Emerging trends of Deep Learning Technique for Lung Cancer Detection based on Convolutional Neural Network

Change Record

Author(s)	Version	Date	Notes	Supervisor's Signature
Ammar Hassan	1.0	22-Aug-23	Original Draft	
Ammar Hassan Abdullah Sajid	1.1	25-Aug-23	Changes Based on Feedback from Supervisor	
Ammar Hassan	1.2	5-Sept-23	Changes Based on Feedback From Faculty	
Ammar Hassan	2.0	15-Sept-23	Added Project Plan	
Abdullah Sajid Ammar Hassan	2.1	15-Dec-23	Changes Based on Feedback from Supervisor	
Ammar Hassan Abdullah Sajid	2.2	1-Jan-24	Completed Chapter 1-2	
Ammar Hassan	2.3	15-Jan-24	Changes based on Feedback from Supervisor	
Ammar Hassan Abdullah Sajid	3.0	7-Feb-24	Completed Chapter 3-4	
Ammar Hassan	3.1	20-March-24	Changes based on Feedback from Supervisor	
Ammar Hassan	3.2	1-April-24	Completed Chapter 5	

APPROVAL

PROJECT SUPERVISOR

Comments: ____

____

Name:______________________

Date:______________________ Signature:______________________

PROJECT MANAGER

Comments: ____

____

Date:______________________ Signature:______________________

HEAD OF THE DEPARTMENT

Comments: ____

____

Date:______________________ Signature:______________________

Dedication

I dedicate this project to Superior University for providing me with a platform for learning and growth, and to my esteemed teachers, whose guidance and knowledge have illuminated my path throughout this journey. Additionally, I dedicate this work to all those affected by lung cancer, in the hope that my efforts in early detection contribute to a brighter future.

Acknowledgements

I would like to express my heartfelt gratitude to Superior University and my dedicated project supervisor, **Dr. Humayun Khan**, for their unwavering support, expert guidance, and valuable insights throughout the course of this project. Their mentorship has been instrumental in shaping this work and my overall growth as a researcher. I am truly thankful for their time, patience, and expertise.

Executive Summary

There has never been a more important need for early, non-invasive lung cancer detection because lung cancer is still one of the world's biggest health concerns. Conventional diagnostic methods such as CT scans and X-rays are very helpful in identifying the disease, but manual interpretation is prone to inconsistent results and human error. In response to this difficulty, our work presents an improved automated approach that uses deep learning models to accurately classify lung images. This work makes use of a large dataset of lung images that have been classified as normal, malignant, and benign. An initial examination of the dataset revealed distinct features related to image dimensions as well as discernible differences between categories. Understanding how important it is for input to neural networks to be consistent, every image was subjected to a thorough preprocessing process in which they were grayscale and standardized to a single dimension. The Synthetic Minority Oversampling Technique (SMOTE) was utilized to address the observed class imbalances within the dataset. Three new architectures—Model I, Model 2, and Model 3—as well as an ensemble method that integrated their forecasts were presented. With an accuracy of roughly 84.7%, Model 1 stood out as the most promising of the models. But the ensemble approach, which was created to capitalize on the advantages of individual models, produced an impressive 82.5% accuracy. Even though Models 2 and 3 had lower accuracy, their distinct advantages and misclassification trends are being taken into consideration for future ensemble enhancements. A prompt, accurate, non-invasive solution to the problems associated with lung cancer detection is provided by the suggested deep learning-driven approach. Reduced diagnostic errors and better patient outcomes could result from its potential for seamless integration with current diagnostic tools. We want to take this research and make it more approachable so that clinicians will accept it and we can move forward with a new generation of diagnostic technology.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

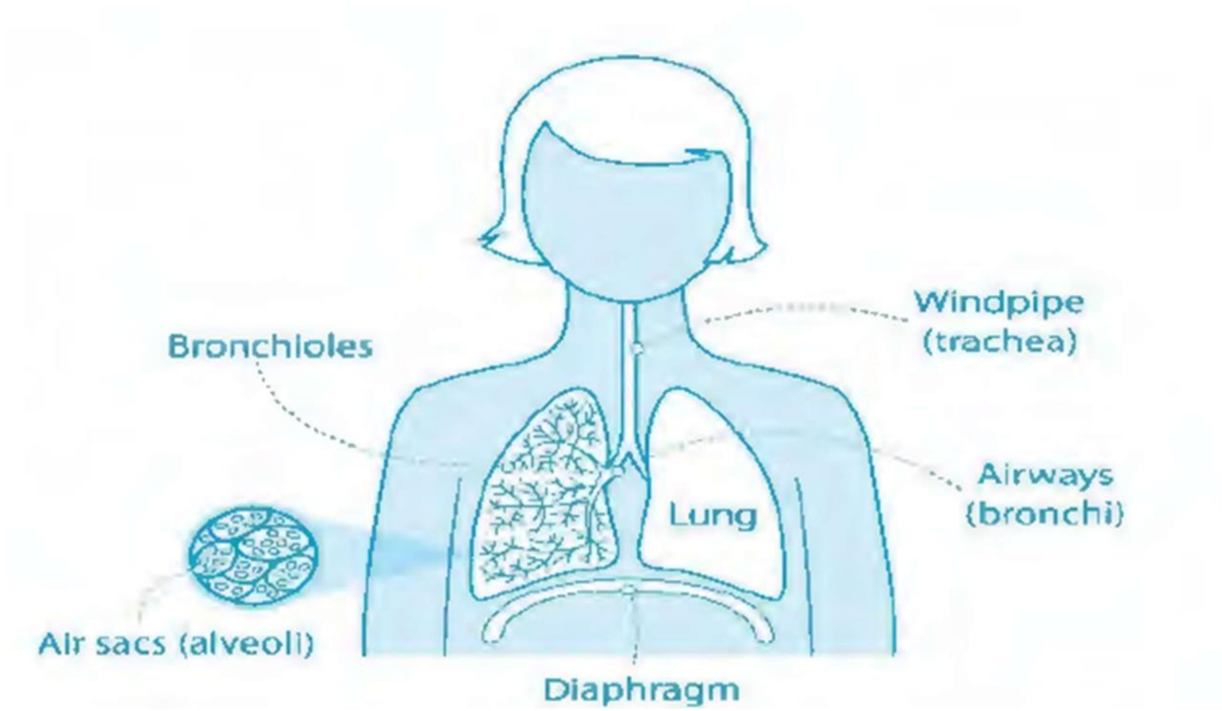
Chapter 1: Introduction

Chapter 1 serves as an introduction to the research on lung cancer detection, highlighting its importance and the need for advanced diagnostic methods. It outlines the challenges present in current detection approaches and sets the primary objectives of the study. Additionally, it presents key research questions, establishing the groundwork for subsequent chapters, and discusses the significance of the study for the medical and scientific communities.

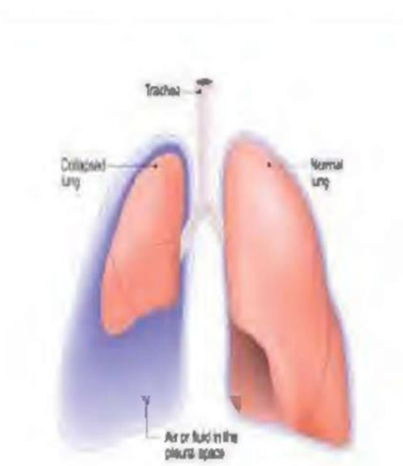
1.1. Background

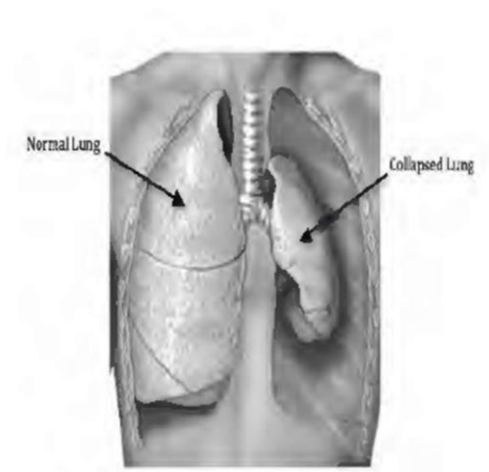
The human body relies on oxygen for survival, obtained through the respiratory system. The lungs, located in the chest, are the primary organs of this system. They are shaped like two inverted cones and have a pinkish color. The right lung has three lobes, while the left has two to accommodate the heart (Alzain et al., 2021).

The thin tissue layer surrounding the lungs is called the pleura, and the chest cavity is lined with similar tissue. The function of the lungs facilitates respiration, where oxygen is taken in from the air and carbon dioxide is expelled. This process is crucial for the exchange of gases in the blood.



The lungs are vital organs in the body, responsible for exchanging gases. However, various diseases can diminish their ability to function. Some lung disorders that can affect the lungs include asthma, COPD, pneumothorax, pneumonia, influenza, lung cancer, tuberculosis, fibrosis, and COVID-19. In severe cases, lung disorders can lead to respiratory failure. Several pulmonary function tests, such as chest X-ray, CT scan, and M and D treason, can help diagnose lung diseases. Pneumothorax is a type of lung disorder where air enters the space between the lung and chest, causing the lung to collapse (Sazuan et al., 2023). Chest injuries, medical procedures, ruptured air blisters, or underlying lung conditions can cause pneumothorax. Symptoms of pneumothorax include sudden and intense pain while inhaling, chest pressure that worsens over time, increased heart rate, and shortness of breath. However, in some cases, there might be no symptoms. Pneumothorax can be classified into primary spontaneous, secondary spontaneous, traumatic, and tension pneumothorax. The incidence of pneumothorax is higher in males, with 4 to 8 cases per 100,000 people per year and 1.2 to 6 cases per 100,000 population in females (Alballa & Al-Turaiki, 2021).





Breathing is vital for all living beings as it involves inhaling and exhaling air through the lungs for gas exchange. The composition of breath can provide essential information for diagnosis. Breath testing is a non-invasive way to diagnose lung cancer, COPD, and more. Spirometry is the most common test for COPD, measuring how much air the lungs can hold and how fast air can be blown out. Other tests that healthcare professionals may use to diagnose lung problems include lung function tests, X-rays, CT scans, arterial blood gas tests, and laboratory tests.

Many tests can diagnose lung cancer, including chest X-rays, CT scans, lung function and blood tests, biopsies, sputum cytology, pleural taps, molecular tests, and PET-CT scans. However, each test has limitations, such as a high rate of false positives with CT scans, discomfort, potential risks with sputum cytology, and concerns about radiation exposure with PET-CT scans. Some of these tests, such as MRI and PET-CT scans, can be expensive and time-consuming. In recent years, researchers have shown increasing interest in using breath analysis to detect lung cancer, as this approach is non-invasive, painless, and can provide real-time results.

Researchers have used various techniques for breath analysis, including Selected Ion Flow Tube Mass Spectrometry (SIFT-MS), Proton transfer reaction mass spectrometry (PTR-MS), ion mobility spectrometry (IMS), sensor arrays, Gas Chromatography-Mass Spectrometry (GC-MS), and laser absorption spectroscopy (LAS). Machine learning (ML) is a data analysis technique that automates the building of analytical models and is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns, and make decisions with minimal human intervention.

Despite the availability of various diagnostic tools, lung cancer remains difficult to detect in its early stages, as symptoms often do not appear until the cancer has progressed to a more advanced stage. Medical image processing (MIP) systems, such as MRI, CT scans, ultrasound, and x-ray imaging, can help improve the efficiency of diagnosis. CT scans are particularly useful for identifying abnormalities in soft tissues and the growth of lumps.

To diagnose lung cancer in individuals who may be at risk, healthcare professionals may use a combination of tests, such as chest X-rays, CT scans, lung function and blood tests, biopsies, sputum cytology, pleural taps, molecular tests, PET-CT scans, CT, MRI, or bone scans, and tests before surgery or radiation therapy to check lung function. Some of these tests have limitations, such as the high rate of nodule detection in CT scans and the cost and discomfort associated with certain procedures. In addition, certain risks are associated with certain tests, such as the potential for a sputum cytology exam to miss what it is looking for and potential concerns for diabetic patients with a PET-CT scan. To mitigate these limitations, breath analysis research has been conducted to recognize trace gas compounds present in the exhaled breath of healthy individuals to establish 'normal' levels of these compounds. If abnormal levels of these compounds are detected, it may indicate adverse clinical conditions. Breathing is a vital, pivotal living being as it involves inhaling and exhaling air through the lungs for gas exchange. The composition of breath can provide essential information for diagnosis. Breath testing is a non-invasive way to diagnose conditions such as and more. Spirometry is the most common test for COPD, which measures the amount of air and how much can be blown out. Other tests that healthcare professionals may use to diagnose lung problems include lung function tests, X-rays, CT scans, arterial blood gas tests, and laboratory tests.

Many tests can diagnose lung cancer, including chest X-rays, CT scans, lung function and blood tests, biopsies, sputum cytology, pleural taps, molecular tests, and PET-CT scans. However, each test has limitations, such as a high rate of false positives with CT scans, discomfort with sputum cytology, and concerns about radiation exposure with PET-CT scans. Some of these tests, such as MRI and PET-CT scans, can be expensive and time-consuming. In recent years, researchers have shown increasing interest in using breath analysis to detect lung cancer, as this approach is non-invasive, painless, and can provide real-time results.

Researchers have used various techniques for breath analysis, including Selected Ion Flow Tube Mass Spectrometry (SIFT-MS), Proton transfer reaction mass spectrometry (PTR-MS), ion mobility spectrometry (IMS), sensor arrays, Gas Chromatography-Mass Spectrometry (GC-MS), and laser absorption spectroscopy (LAS). Machine learning (ML) is a data analysis technique that automates the building of analytical models and is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns, and make decisions with minimal human intervention.

Despite the availability of various diagnostic tools, lung cancer remains difficult to detect in its early stages, as symptoms often do not appear until the cancer has progressed to a more advanced stage. Medical image processing (MIP) systems, such as MRI, CT scans, ultrasound, and x-ray imaging, can help improve the efficiency of diagnosis. CT scans are particularly useful for identifying abnormalities in soft tissues and the growth of lumps.

To diagnose lung cancer in individuals who may be at risk, healthcare professionals may use a combination of tests, such as chest X-rays, CT scans, lung function and blood tests, biopsies, sputum cytology, pleural taps, molecular tests, PET-CT scans, CT, MRI, or bone scans, and tests before surgery or radiation therapy to check lung function. Some of these tests have limitations, such as the high rate of nodule detection in CT scans and the cost and discomfort associated with certain procedures. In addition, certain risks are associated with certain tests, such as the potential for a sputum cytology exam to miss what it is looking for and potential concerns for diabetic patients with a PET-CT scan. To mitigate these limitations, breath analysis research has been conducted to recognize trace gas compounds present in the exhaled breath of healthy individuals to establish 'normal' levels of these compounds. If abnormal levels of these compounds are detected, it may indicate adverse clinical conditions.

1.2. Diagnosis of Lung Cancer

There are various tests used to diagnose lung cancer in individuals who may be at risk. These tests include chest X-rays, CT scans of the heart and lungs, lung function and blood tests, biopsies, sputum cytology, pleural taps, molecular tests, PET-CT scans, CT, MRI, or bone scans, and tests before surgery or radiation therapy to check lung function.

However, some of these tests have limitations, such as the high rate of nodule detection in CT scans and the cost and discomfort associated with certain procedures.

To mitigate these limitations, breath analysis research has been conducted to recognize trace gas compounds in the exhaled breath of healthy individuals and establish 'normal' levels of these compounds. If abnormal levels are recognized, which may be indicators of adverse clinical conditions, then breath analysis for these biomarkers would be of huge clinical value in the battle to diagnose and cure lung cancer.

Furthermore, machine learning is a powerful diagnostic tool that can automate the building of analytical models with minimal human intervention. The high morbidity rate in lung cancer is most common worldwide, and it includes small and non-small cells of carcinomas. Early-stage detection of lung cancer is crucial for treating and improving a patient's health.

However, due to the softness of pulmonary tissues, early detection of lung cancer can be difficult.

To diagnose lung cancer efficiently and quickly, medical image processing (MIP), including MRI, CT scan, ultrasound, and x-ray imaging, is used. CT scan is considered the best way for lung diagnosis because of its efficiency in providing key information about the abnormality of soft tissues like pulmonary tissues and the growth of lumps, identifying their size, position, and shape.

The quality of a CT scan is that it provides 300 various scans in a complete cross-section of the lung in a very short period.

1.3. Tests that Diagnose Lung Cancer

Various diagnostic procedures are used to detect cancerous lung cells and assess their condition. These include imaging tests like chest X-rays, CT scans, and sputum cytology biopsies. X-rays can identify unusual growths or nodules in the lungs, while CT scans can uncover small lesions not visible on X-rays. A biopsy involves removing a sample of abnormal cells for examination. This can be done using a bronchoscope, a lighted tube passed through the throat and into the lungs, or through a mediastinoscopy. This surgical procedure involves making a cut near the neck and taking samples from the lymph nodes near the breastbone. Another method is needle biopsy, in which a needle is inserted through the chest wall and into the tissue with the guidance of a CT scan or X-ray images. The samples can also be taken from other areas where the cancerous cells may have spread, such as the liver. Analyzing the cells can help determine the type of lung cancer and assist in making treatment decisions.

Diagnostic tests are medical procedures to identify cancer progression and determine suitable treatment options.

These tests include various imaging techniques such as scans, MRIs, PETs, and bone scans, which provide detailed information on the cancer's size, location, and stage. Doctors then use this information to determine the most appropriate course of treatment for the patient (Hertel &

Belmar, 2021). Lung cancer stages are usually classified using Roman numerals, with stage IV being the most advanced and indicating that the cancer has spread to other parts of the body.

1.4. Surgeries

Different surgical procedures may be used to remove lung cancer, such as resection through wedge.

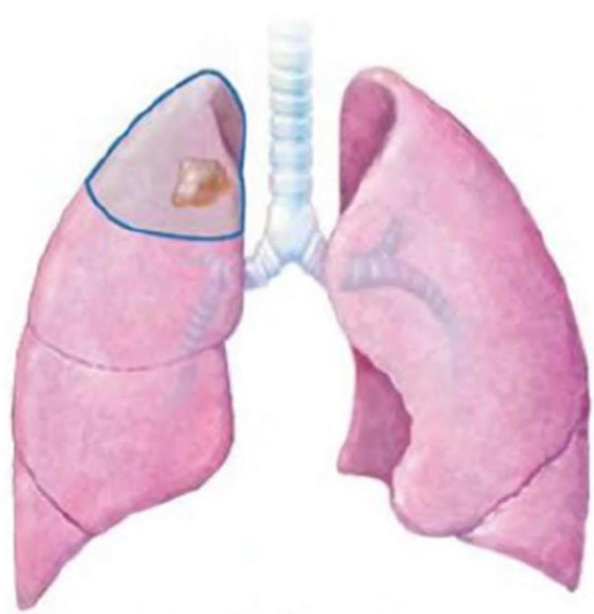
1.4.1 Resection by wedge

This method involves removing the lung's cancerous portion while preserving as much healthy tissue as possible. Usually, a small incision is made, and imaging techniques like CT or MRI are used to visualize the surgery.
The choice of surgical method depends on various factors, such as the stage, size, and location of the cancer and the patient's overall health (Aarthi et al., 2021).

Wedge resection

1.4.2 Resection by Segments

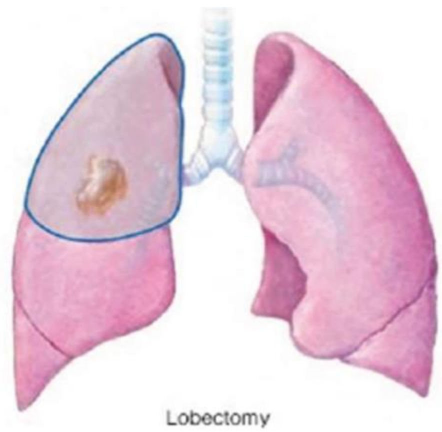
Lung cancer can be surgically removed through a variety of methods, including segmental resection. This method involves removing a large portion of the lung that contains the tumor but not the entire lobe. This is done to preserve as much healthy lung tissue as possible while effectively removing the cancerous cells (Jiang et al., 2021).



Segmental resection

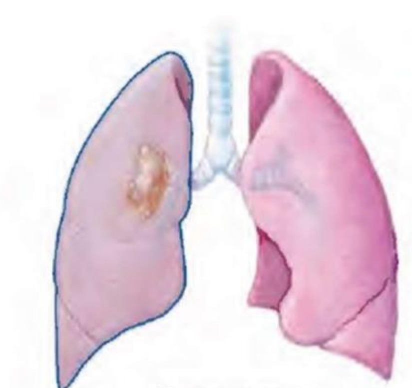
1.4.3 Lobectomy

Lobectomy is a surgical procedure that involves the removal of an entire lobe of the lung. This method is typically used for larger tumors or those that have spread to multiple lung lobes. During the surgery, the surgeon will carefully remove the affected lobe, taking care to preserve as much healthy lung tissue as possible. This type of surgery can be challenging and is typically only performed by highly skilled and experienced surgeons. The recovery time after a lobectomy can vary depending on the individual and the extent of the surgery (Hagan et al., 2021). Patients typically need to spend several days in the hospital and may require physical therapy to regain strength and lung function.



1.4.4 Pneumonectomy

Lobectomy and Pneumonectomy are surgical procedures that are used to remove cancerous tissue from the lungs. In a Lobectomy, a complete lobe of one lung is removed. While in In Pneumonectomy, an entire lung is removed. These surgeries are typically performed in advanced stages of lung cancer when cancer has spread beyond a small section or lobe of the lung. These procedures may involve the removal of other surrounding structures, such as lymph nodes and other surrounding tissue, depending on the stage and extent of the cancer (Almustafa, 2021). These surgeries can be complex and have a long recovery time. Patients need to have a detailed discussion with their surgeon about the risks, benefits, and recovery process before making a decision.



When it comes to treating lung cancer, surgery is one of the options that can be considered. If the cancerous cells are large (Mortezaei & Tavallaei, 2021) physicians may recommend undergoing chemotherapy or radiation therapy before the surgery to shrink the size of the cells. Additionally, if any cancerous cells remain in the lung following surgery, further chemotherapy Alternatively, radiation therapy may be used to eliminate those remaining cells.

1.4.5 Chemotherapy

Chemotherapy is a form of treatment that uses drugs to target and destroy cancer cells. This treatment can be administered intravenously through a vein in the arm or orally. Typically, multiple drugs are combined, and treatments are given in cycles with specific intervals in between to allow the patient to recover (Shams et al., 2021). In cases of lung cancer, chemotherapy may be used both before and after surgery to reduce the size of cancerous cells and eliminate any remaining cells. Additionally, for those with advanced-stage lung cancer, chemotherapy can help alleviate symptoms and reduce pain.

1.4.6 Radiation therapy

Radiation therapy, or radiotherapy, uses high-energy beams, such as X-rays and protons, to target and destroy cancerous cells. A machine delivers the beams to the specific location of the cancer in the body. This therapy can be used before or after surgery and may be combined with chemotherapy. It is commonly used to alleviate pain and slow the progression of advanced lung cancer that has)spread to other parts of the body. In some cases, radiation therapy can be used as the primary treatment for early-stage lung cancer, but it is more commonly used in combination with surgery or chemotherapy.

1.4.7 Radiosurgery

Radiosurgery, also referred to as stereotactic body radiotherapy, is a highly focused radiation treatment that uses multiple beams of radiation from different angles to target cancerous cells. This type of treatment usually only requires one or a few sessions to be completed. It is often considered as an alternative for individuals with small lung cancer who are. Unable to undergo surgery (Nnamdi Ekwueme et al., 2023). Additionally, radiosurgery can also be used to treat Lung cancer has spread to other areas of the body, such as the brain.

Drug therapy Targeted drug therapies are a specific type of treatment that focuses on specific genetic or molecular abnormalities found within cancer cells. These drugs block the growth and spread of cancer cells and can be combined with other treatments. NTS is like chemotherapy or radiation therapy (Panahi et al., 2021). Targeted therapies are often used to treat advanced or recurrent lung cancer and are typically prescribed based on the specific genetic makeup of an individual's cancer cells. Some targeted therapies are only effective in patients whose cancer cells have certain genetic mutations or abnormalities, and the cells are analyzed in a laboratory to determine whether these drugs would be effective for the patient (Hao et al., 2023).

1.4.8 immunotherapy

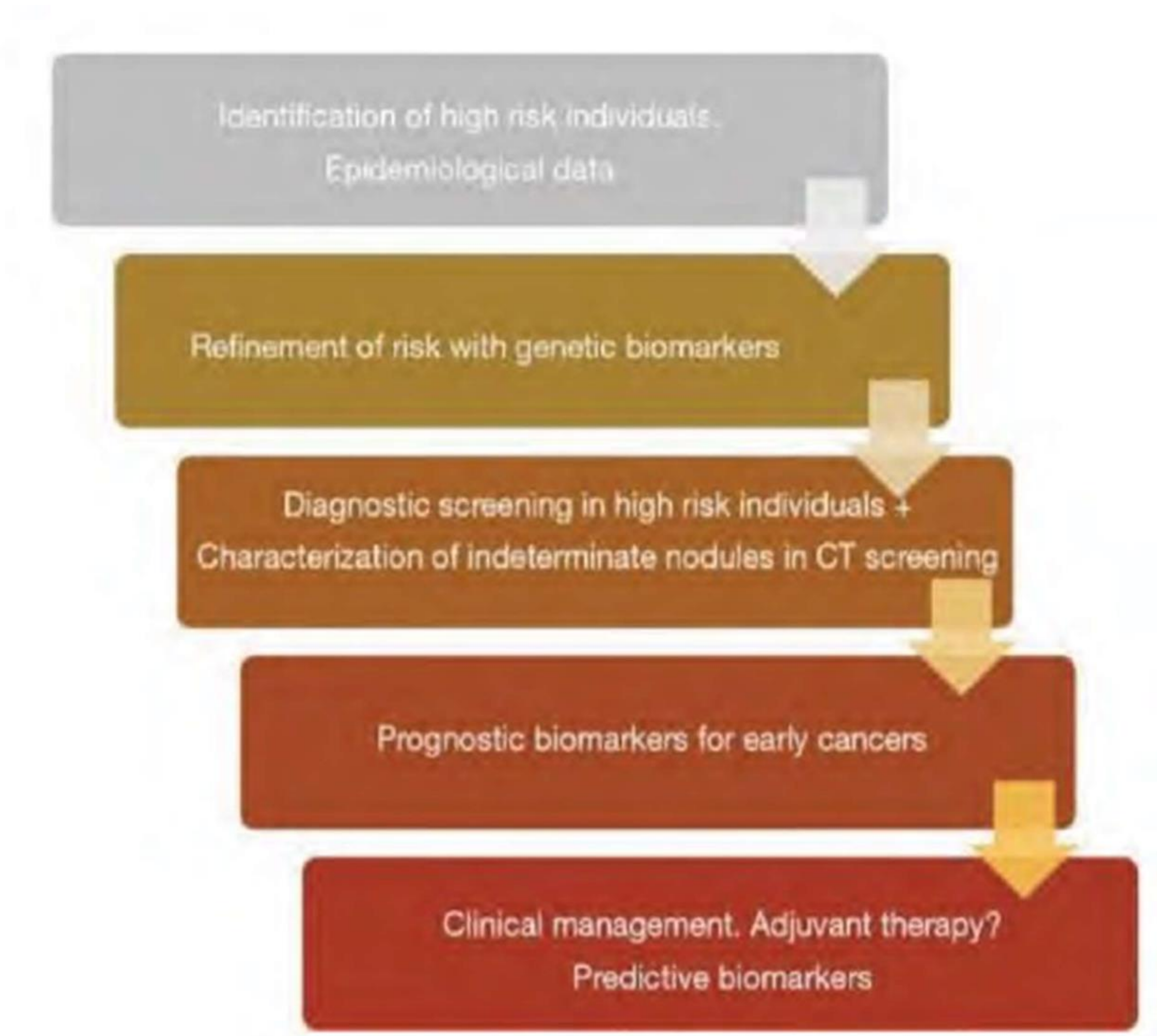
Immunotherapy is a cancer treatment that uses the body's immune system to fight cancer cells. The immune system may not attack cancer cells because they produce proteins that help them evade detection by immune system cells. Immunotherapy works by targeting these proteins and making it easier for the immune system to recognize and attack cancer cells. This treatment is typically reserved for individuals with advanced stages of lung cancer or tumors that have spread to different parts of the body (Gao et al., 2023). It aims to slow down cancer growth by activating the body's natural defense system to recognize and attack the cancer cells (Paul et al., 2021).

1.5 The stages of lung cancer

Lung cancer is categorized into two main types: non-small cell and small cell. Non-small cell lung cancer is the most common type and is divided into four stages: I, II, III, and IV. Small cell lung cancer, or oat cell lung cancer, is divided into two stages: limited and extensive. The stage of a person's lung cancer is dependent on the size and location of the tumor and whether it has spread to other parts of the body. Determining the stage of lung cancer is crucial in deciding the most suitable treatment options (Hussain et al., 2023).

The stages of lung cancer help determine the severity of the disease and the best course of treatment. Non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC) each have their staging system. NSCLC is classified from stages O to IV, with O being the earliest stage and IV being the most advanced. The lower the stage number, the less the cancer has spread (Akhtar et al., 2021). A higher stage number means that cancer has spread more. Within each stage, a lower letter or number indicates a lower stage. Physical exams, biopsies, imaging, and other tests

determine the stage of SCLC. There are two main stages: limited and extensive. Limited stage means the cancer is only on one side of the chest and can be treated with a single radiation field. This includes cancers that are only in one lung and have also reached the lymph nodes on the same side of the chest (Adiba et al., 2021). The extensive stage means that the tumor has spread widely throughout the lung, to the other lung, to lymph nodes on the opposite side of the chest, or other parts of the body.



1.6 Problem Statement

In the pivotal arena of lung cancer diagnostics, early and non-invasive detection remains an aspiration the weal community continually strives towards. Traditional diagnostic mechanisms, such as -rays and CT scans, have undeniably facilitated disease identification. However, their manual interpretation processes are fraught with inherent vulnerabilities:
susceptibility to human error and inconsistencies, leading to potential misdiagnoses or overlooked malignancies. Prevailing automated techniques and machine learning models, while pioneering, often grapple with intricacies in the data. Though effective to a degree, these traditional machine learning models need to be revised to classify every nuanced detail present in lung images. 0020 Such models can only sometimes capture the deep complexities and intricate patterns within the data, often resulting in suboptimal performance compared to deep learning approaches. Addressing these pressing gaps, our research introduces a cutting-edge automated approach that harnesses the capabilities of deep learning for the meticulous classification of lung images. Deep learning's ability to delve into deeper data layers and recognize patterns often imperceptible to traditional models offers a superior advantage. To tackle the issues mentioned earlier, our models incorporated rigorous preprocessing to ensure standardized input across all images. The challenge of class imbalance, a recognized pitfall in prior models, was proactively addressed using the Synthetic Minority Over-sampling Technique (SMOTE). Three individual architectures (Model I, Model 2, and Model 3) showcased varied results. The ensemble approach amalgamated the strengths of each model to deliver results with commendable precision. The deep learning-centric methodology is an avant-garde, non-invasive solution to lung cancer detection and offers potential integration with conventional diagnostic tools. The objective is to mitigate the pitfalls of human error, manual interpretation inconsistencies, and the limitations of traditional machine earning models, thereby promoting improved patient outcomes. As we stand at the brink of an era dominated by advanced diagnostic technology, our future endeavors are geared toward translating this research into a clinically accepted, user-friendly application, bridging the chasm between theoretical study and tangible medical impacts.

1.7 Motivation for the Research

Pakistan has a rapidly growing population of 208 million, spread across a large area of 881913 km2. The population growth has led to an urgent need for healthcare facilities, particularly in small towns and villages where access to proper medical care is limited. This is because medical specialists tend to live in major cities where the standard of living is better, leaving most of the population needing access to healthcare institutions that address critical health issues (Helikumi et al., 2021). In addition, many individuals in these areas either need help to afford expensive medical checkups or are not proactive about their health, o~ waiting until the last stage of the disease to seek treatment from a qualified doctor, making it difficult or impossible to cure. The need for an e-healthcare system becomes even more crucial in light of frequent contagious viral attacks, such as the recent COVID-19(Manuja et al., 2023) pandemic declared by the World Health Organization (WHO) on March 11 have been the only way to reduce the spread until a vaccine is developed. The main focus of this project is to provide early-stage diagnosis and reduce patient visits to hospitals. The proposed system aims to provide a high-quality analysis at a low cost and with high accuracy, saving patients time and money. One of the major advantages of this proposal is the widespread availability of cell phones with built-in cameras and internet access in Pakistan. This provides the foundation for developing and implementing this system in rural and developing areas. IoT and ML technologies to collect data from medical diagnostic instruments and biomedical sensors also enable diagnosing and treating various diseases through mobile phone applications, even in the patient's home or at a nearby health center. However, an e-healthcare system has several disadvantages (Al Bassam et al., 202 1). Patients may be reluctant to share their disease information or diagnosis reports, which could lead to problems in their personal or professional lives. Electronic medical records are also vulnerable to hacking, which could result in sensitive patient data falling into the wrong hands. Additionally, privacy and security concerns exist when data is collected at remote health facilities and transferred to cloud-based systems or specialized centers for analysis (S. Hamid et al., 2022a).

In conclusion, lung cancer is a commonly diagnosed cancer worldwide, accounting for 2 million new cases annually. It is a deadly disease, with a 5-year survival rate of only 15%. Early detection provides the best prognosis for patients. Similarly, early detection of COVID-19 can lead to more

cures and longer survival (S. Hamid et al., 2022b). The proposed e-healthcare the system aims to address the issue of limited access to healthcare in rural and developing areas of Pakistan, provide early-stage diagnosis, and reduce patient visits to hospitals while
addressing privacy and security concerns.

1.8 Objectives of the Research

The main objectives of this research activity include the following:

- To standardize and preprocess lung images, ensuring optimal input for neural
- networks.
- To employ SMOTE to address dataset class imbalances and enhance
- representation.
- To design, implement, and evaluate three deep-learning models for precise lung
- image classification.
- To create an ensemble approach that amalgamates the strengths of individual
- models for improved accuracy.
- To translate the deep learning methodology into a user-friendly application for
- clinical integration.

1.9 Motivation for the Research

1. How can lung images be standardized and transitioned to grayscale for neural network compatibility?
2. How effectively does SMOTE address class imbalances in the lung image dataset?
3. How do the three deep learning architectures compare regarding lung image classification accuracy?
4. How does the ensemble method's accuracy compare to individual models in detecting lung cancer?
5. What is the pathway to convert the deep learning method into a user-friendly clinical application?

1.10 Thesis Organization:

Chapter 1: Introduction

This chapter briefly overviews lung cancer detection, emphasizing its significance and the necessity for advanced methods. It identifies the challenges in current detection mechanisms and sets forth the primary goals of the research. This section also poses the key research questions, laying the foundation for subsequent chapters, and elaborates on the importance of this study to the medical and scientific communities.

Chapter 2: Literature Review

Diving into the literature, this chapter surveys existing techniques for lung cancer detection. It offers a critical assessment of prevalent methods, emphasizing their limitations. The potential of non-invasive methods as efficient diagnostic tools is explored, with a particular emphasis on the role of human breath in diagnosing diseases, including lung cancer and COVID-19.

Chapter 3: Methodology

Here, the research blueprint is laid out. It begins by detailing how the data was collected, elaborating on the types of participants, instruments, and processes employed. Further, the analytical techniques applied to the data are described, ensuring a clear understanding of how the research was conducted. Lastly, the mechanisms for verifying the results are discussed.

Chapter 4: Results

In this chapter, the findings from the data analysis are presented in a structured manner. It is where the answers to the research questions are uncovered, and the tangible outcomes of the methodology are displayed.

Chapter 5: Future Work

Building upon the results, this chapter delves deeper into what the findings mean. It situates the results within the context of existing literature, offering insights, drawing comparisons, and suggesting implications. Potential areas for further research based on the outcomes are also highlighted.

Chapter 2

LITERATURE REVIEW

Chapter 2: Literature Review

2.1. Introduction

Many researchers and medical professionals are now trying to develop a cure for COVID-19. However, that is not the last chapter. Predicting the risk of contracting COVID early is crucial for reducing the likelihood of mortality. Developing an early prediction system using deep learning, machine learning, and other techniques has received considerable interest from the academic community. However, proving forecast accuracy has been challenging due to the X-ray picture's complexity. This research aimed to develop a technique for predicting the spread of COVID-19 (Manuja et al., 2023). Chest X-rays are used to assess the likelihood of contracting COVID. To begin configuring the classification layer, clean segmentation, and feature extraction data must be imported. Locations, where images were trained with a vulture fitness exactness rate were impacted by this section. The developed model now has a segmentation accuracy of 99.9%, higher than the error rate of other models, which is 0.0145. Therefore, the proposed medical strategy would provide the best outcomes.

2.2. Related Work

The last decade has seen a proliferation of deep learning techniques in medical imaging, specifically convolutional neural networks (CNNs) (Assistant et al.). Various research studies have demonstrated their efficacy in detecting anomalies in different body organs, including tumors. For lung cancer, Lakhani and Sundaram (2017) leveraged deep learning to detect and classify pulmonary nodules in chest radiographs, revealing a promising direction for the early-stage detection. The Internet of Medical Things (EMT) represents a network of interconnected health devices and applications. Kaur and Sood (2020) have highlighted the potential of IoMT in the proactive monitoring of patients, particularly in chronic diseases like asthma and diabetes (Habib & Rahman, 2021). However, its direct application in early-stage Lung cancer detection remains sparsely explored. A handful of pioneering works have begun exploring the synergy between deep learning and IoMT. Zhang et al. (2019) proposed a system that uses wearable IoMT devices to collect data and employs deep learning algorithms for real-time health monitoring. While this approach is innovative, its specific tailoring for lung cancer detection is still in its infancy. The literature showcases a variety of non-invasive methods for lung cancer detection. Optical coherence tomography and liquid biopsies have been particularly interesting in recent years. However, the integration of these non-invasive methods with deep learning and IoMT platforms are a burgeoning field with vast potential yet to be fully realized. The intersection of deep learning (Kate & Shukla, 2021), lost, and non-invasive lung cancer detection presents Both challenges and opportunities. Latency, real-time processing, data security, and patient privacy are among the key concerns raised by scholars. On the other hand, the potential for early detection, personalized treatment plans, and reduced healthcare costs provides a compelling case for further exploration in this domain (Munnangi et al., 2023).

2.3. Research Gap

The recent advancements in the domains of deep learning and the Internet of Medical Things (long) individually undere:5re significant promise in the landscape of medical diagnostics. Deep learning, particularly convolutional neural networks (CNNs), has garnered much attention due to its superior performance in image analysis tasks. Their application in medical imaging for detecting various anomalies, such as tumors in varied body organs, has been widely documented. On the other hand, the Internet of Medical Things (loMT) (Al Bassam et al., 2021) is revolutionizing healthcare by offering a robust network of interconnected health devices and applications geared toward proactive patient monitoring. However, despite the transformative potentialities exhibited by these technologies separately, a conspicuous research gap exists when it comes to their combined application, specially tailored for the early-stage detection of lung cancer (Imran et al., 2021). ! ng cancer, being one of the most prevalent and lethal cancers worldwide, demands innovative diagnostic techniques that can detect its onset early, thereby improving the prognosis and potential treatment outcomes. Current techniques, though effective, are often invasive or less efficient in early-stage detection. A non-invasive method harnessing the power of deep learning, especially in deciphering complex patterns from medical images, and the real-time, continuous data monitoring capabilities of IoMT could represent a monumental shift in lung cancer diagnostics. However, studies focusing on this synergy are surprisingly scant.

This raises the question: How can these two transformative technologies be seamlessly integrated to develop a noninvasive diagnostic tool tailored specifically for lung cancer?

Moreover, while general health monitoring IoMT devices are becoming commonplace, developing devices optimized specifically for lung cancer detection, offering high sensitivity and specificity, remains largely uncharted territory (Ramanathan et al., 2023). This highlights another significant research gap. Furthermore, as technology integration intensifies in medical applications, a pressing need arises to benchmark the performance of such hybrid systems against conventional methods, facilitating a clearer understanding of their advantages and limitations. While the technological aspects are vital, one must recognize the human-centric side of the equation. With the increasing digitization and connectivity in healthcare, issues concerning data privacy, patient consent, potential misdiagnosis, and the overarching societal implications have become more pronounced. Yet, the current literature offers limited discourse on the ethical and societal ramifications of employing a deep learning-powered IoMT framework in the early-stage detection of lung cancer (Dahri et al., 2021).

Moreover, ensuring the patient's comfort, understanding, and acceptance of such advanced solutions is paramount for their widespread adoption and success, and yet, studies delving into the user experience and acceptability of these technologies in the context of lung cancer detection remain notably sparse (Panahi et al., 2021).

2.4. Review Literature

Akter et al. (202 I) suggested that most well-known causes of cancer-related mortality is the deterioration of lung cells. The number of people who die from lung cancer may be lowered by finding and treating cancerous tumors in the lungs as soon as possible. The lack of outward symptoms before the breakdown of lung cells presents a significant challenge to early diagnosis.

Computed tomography (CT) and other painless imaging techniques will be utilized for medical diagnosis and screening (Garrido et al., 2021). However, a precise automated analysis of These high-resolution images was required to comprehend the technology's potential properly. Part of this procedure involves splitting the photos apart. Image division techniques use section minimums and column maximums.

Using median values along each line and section in addition to maximum and lowest values has increased the accuracy of sectioning these images. The next part of the research was using a neuro-fuzzy algorithm to determine whether the segmental lung knobs posed a threat (Chakravadhanula, 2021). Emotional resonance, detail, and verisimilitude were among the criteria used to evaluate the performance. The proposed technique has a decreased proportion of false positives while still being 100% effective, 81 % specific, 86% accurate, and 90% exact. Few Americans who are eligible for radiologic screening have it done, even though doing so reduces the frequency of fatalities caused by cell death in the lungs. Screening rates may increase as more people have access to blood-based diagnostics. To improve the utility of screening (Rayane Correia de Oliveira et al., 2023), Chabon et al. (2020) developed an approach to measure circulating cancer DNA (ctDNA) that increases the disease-specific profile by deep sequencing (CAPP-Seq). We demonstrate that ctDNA is present in most patients before therapy, and its quality is a strong predictor, even though the number of injured cells in the lungs is minimal in the early stages (Amir Behghadami & Janati, 2021). Clonal hematopoiesis is further suggested by the long-lasting and Substantial variations in circulating free DNA (cfDNA) between high-risk patients with a cellular lung breakdown and healthy controls. Clonal hematopoiesis alterations occur on larger stretches of cfDNA and need mutational markers connected to smoking, in contrast to growth-related changes. We use these findings in conjunction with other atomic data to develop and provisionally approve an artificial intelligence approach we term "cell breakdown in the lungs probability in plasma" (Lung-Clasp) (Furtado, 2021), which can distinguish between patients with early-stage lung cancer and at-risk controls. This approach yields the same outcomes as growth-informed ctDNA implantation and allows you to adjust the clarity of the measures for certain therapeutic applications. Our findings demonstrate the feasibility of using cfDNA to detect lung cell death and highlight the significance of risk matching in cfDNA-based screening trials (Sazuan et al., 2023).

(Bibi et al. (2020)) suggested that new research indicates that chronic leukemia is the two most common forms of the disease. Myeloid and lymphoid cells may be used to subdivide each kind further. There are four subtypes of leukemia. Various methods have been proposed for detecting cancer, each tailored to a certain subtype or set of characteristics. However, there is room for improvement in how effective these approaches are and how well they facilitate learning (Hajizadeh & Monaghesh, 2021). To expedite and enhance cancer diagnosis and care Delivery, a tool based on the IoMT, is now being developed. There is a lot to consider when all this is happening. An IoMT system that utilizes cloud computing to link clinical devices to network assets is a technological advancement that allows instantaneous communication between individuals. Patients and their physicians may coordinate leukemia testing, assessment, and therapy, and time and money are spared. Patients' and physicians' efforts are acknowledged. The stated strategy is also an effective means of addressing the challenges faced by those with terminal diseases (Lincke et al., 2021). The following diagram depicts the framework used in cancer cluster discovery during pandemics such as COVID-19.

ResNet-34, a convolutional neural network, may be either a dense (DenseNet-121) or a residual (RCNN) convolutional neural network. The public may access two of them. The ALL-IDB and the SH image collection, both of which pertain to leukemia, are used in this endeavor. The findings confirmed the validity of the proposed models. Our machine-learning method distinguishes healthy and cancerous samples better than popular techniques. (Munnangi et al., 2023). Filipska & Rosell, (2021). There has been significant development over the last decade in the Ability to detect circulating cancer DNA (ctDNA)- without cell DNA (cfDNA) portions that cancer cells transfer into the bloodstream and show gene changes. Therefore, ctDNA extracted from bodily fluids may be utilized for illness diagnosis, therapy planning, treatment tolerance identification, and disease recurrence monitoring. Results and techniques for detecting cellular breakdown in the lungs utilizing advances in ctDNA identification is discussed in this article (Filip et al., 2023). They also advocate for greater investigation into whether lung cancer-related fluid changes may be detected. The early applications of ctDNA-based fluid biopsy focus on significant changes and EGFR-related alterations. Finally, we consider the possibility of employing a microbiome-driven fluid sample to detect early malignant growths and distinguish clonal hematopoiesis (Handayani et al., 2022) Rehman et al. (2015) proposed that the rapid global spread of the novel COVID-19 virus has endangered the global healthcare system. If this illness is present, it may be detected in the body using sputum or blood testing.

Decisions may also be made rapidly using computed tomography (CT) scans or X-rays. To assist clinicians and slow the spread of illness, it is crucial to fund a CAD system that is accessible online and at all times. In this research, chest X-rays were employed with a Leftover neural network (ResNet50) trained on CNN data to detect COVID-19 with 98% accuracy. It will develop a CAD system to get X-ray images from small clinics and hospitals outside major cities and run them through the appropriate cycles. The recommended CAD layout additionally accounts for an expected increase in infected instances. It uses superior load-balancing and flexibility components to deal with non-critical failures quickly and easily (Al Bassam et al., 2021). Mathios et al. (2014) developed a noninvasive and simple method to detect lung cell death in 2021. Non-destructive methods for analyzing single-cell DNA (cfDNA) have made cancer detection and treatment feasible. In this investigation, we apply an Al model to identify growth-determined cfDNA tests that use genome-wide disruption in a sample of 365 individuals at risk for lung cell breakdown. By comparing a control group of 385 individuals without cancer to a group of 46 individuals with a cellular breakdown in the lungs, we provide evidence in favor of the concept of cancer localization (Qu et al., 2021). When all stages and kinds of cancer were considered, together with fracture characteristics, clinical risk factors, and CEA levels, 94% of patients were found to have cancer progression. In 80% of these instances, CT imaging was utilized to confirm this, and it was successful in confirming stage 1/11 (91 %), stage IWIV (96%), and stage IV (98%). Genome-wide discontinuity profiles utilizing about 13,000 ASCLl record factor limiting endpoints distinguished between individuals with little cell breakdown in the lungs and those without it with great accuracy (AUC = 0.98). A higher score for the total number of game breaks is a cost-free indicator of perseverance (Shafi et al., 2021).

Test images of the thorax come from various publicly accessible online resources. As a whole, the method's classification accuracy is up to 97%. Since, the rapidly spreading COVID-19 The virus was discovered early on, so it may be able to protect many individuals. X-rays and CT scans are used to diagnose each of these illnesses. This research demonstrates the value of combining CT and chest X-ray images to detect COVID-19, pneumonia, and lung. Filler. The model relies on deep Learning and using many different categorization types. Because they may detect illness before symptoms appear, CT scans are superior to cans before X-rays for diagnosing advanced illnesses because they highlight abnormalities already seen in the pictures. Adding more images of both sorts to the collection will improve the classification quality. There is currently no publicly available deep learning model that can be used to prioritize which of these dresses to treat. In this study, we employ four subsets of CT scans and X-ray datasets (normal, COVlD-1 pneumonia, and lung ER to evaluate several deep learning architectures. The results of the evaluations showed that the VGG 19 + CNN model outperformed the other three contenders. The VGG I 9+CNN mode performed with an 8.0 at a 98.0 percent can rate, a 98.43 percent precision rate, a 99.3 percent negative predictive value, a 98.24 percent FI score. Matthew's correlation coefficient is an area under the curve AU of 99.66 percent when tested on CT images and X-rays(Masekela et al., 2023). Hwa Kieu et al. (2020) proposed that lung abnormalities in medical images are now simpler to spot and categorize using deep learning and other innovative methods. Research on deep learning for lung illness detection has been extensive. This investigation investigates the feasibility of using deep learning to detect lung illness in diagnostic imaging. There was just one lung illness report in a survey five years before (Chakravadhanula, 202 I). It still needs to be finished, however. They do not investigate categorization theory or historical developments in the field of study. A taxonomy of existing systems that use deep learning to detect lung disorders was provided in this article. They illustrated the progress made in recent studies, the remaining gaps, and the potential ways forward. Future improvements in diagnosing lung illness will have a significant positive impact on patients' lives. Medical devices, equipment, and technology advancements are of interest to academic institutions and health. Industry. The IoT provides a fantastic research platform for health-related topics (Asadzadeh et al., 2021).

Ahmed et al. (2022) examined medical photographs with the assistance of artificial intelligence. This tried-and-true technology allows medical professionals to diagnose diseases earlier, more reliably, appropriately, and precisely. The coronavirus COVID-19 is one of the viruses now spreading worldwide and is considered one of the deadliest. The system must also have a component of advanced technology and automation. A new ent healthcare system built on the Internet of Things (IoT) may employ chest X-rays to automatically identify and classify infectious diseases such as pneumonia and COVID-19 (Kaushik et al., 2020). This approach utilizes two deep learning models and multilayer feature fusion and selection to classify viral. Following these specific instructions, with the assistance of deep learning models such as Inception-V3 and VGG-19, it is possible to mine datasets for useful insights. This allows us to pick and select features learned from different design learning architectures to combine. The images of the chest region used in the testing came from several places in the public domain. When correctly identifying data, this has a precision of 97% overall. Das et al. (2021) state that the 2019 New Coronavirus Disease (COVID-19) outbreak significantly affects health and the economy. It has been a challenge for researchers and medical practitioners to produce testing that is both accurate and quick for the COVID-19. The ground-breaking artificial intelligence-based X-ray device known as CoviLearn, (Sabir et al., 2023) which was created to provide a way for medical professionals to handle the first screening of COVID-19 patients. We have CNN models that can automatically identify COVID-19 in X-ray pictures may be applied to such images. The proposed CoviLearn device can determine whether or not a person is positive for COVID-19 based on images taken from a chest X-ray. Clinicians can now identify COVID- I 9 problems using CoviLearn, eliminating the requirement for invasive blood or breath tests as a prerequisite. The capacity of a patient's lungs may be evaluated using an X-ray. The CoviLearn recommended X- rays might be used to screen for COVID-19 without needing specialist testing equipment since such devices are currently accessible in all hospitals. This would eliminate the need for additional testing costs (Jamshidnezhad et al., 2021). Since a radiography professional must handle X-ray equipment, our suggested automated analysis system, CoviLearn, might help medical personnel save time. The spread of COVID-19 worldwide increased the work done by medical personnel. Clinical personnel in the area affected by the epidemic will reap the benefits of artificial intelligence by reducing the work they need to do and assisting. This study used artificial intelligence (AI) support vector machines to perform automated analysis on CT images of patients with COVID-19-related pneumonia. The method that was employed in this research (Sheela & Arnn, 2022) might be used to train the system to differentiate between pneumonia and other diseases, identify pneumonia, and halve the amount of time it takes for physicians to make a diagnosis (Habuza et al., 2021).

Machine Learning Machine learning, a subfield of artificial intelligence, stands at the intersection of statistics, computer science, and data analysis. Its primary objective is to build a team that can improve and evolve without being explicitly programmed. While the concept of machine "learning" can be traced back to Turing's proposition in the 1950s, it was only in the 1980s. Moreover, in the 1990s, the field began to flourish. Early pioneers like Samuel presented the idea of machines learning from data in the context of game playing. Over the decades, the discipline has grown exponentially, fueled by the convergence of larger datasets, more robust computing power, and advanced algorithmic strategies (Imran et al., 2021). At its core, machine learning methodologies e! One can be broadly categorized into supervised, unsupervised, semi-supervised, and reinforcement learning. Supervised learning, where models are trained. Labeled data to predict unseen instances is the most widespread approach. Algorithms like linear regression, support vector machines, and decision trees have become staples in the ML toolkit. Conversely, unsupervised learning, typified by clustering and association algorithms, uncovers hidden patterns in data without predefined labels. Inspired by behavioral psychology, reinforcement learning involves agents who take actions in environments to maximize cumulative rewards, with deep reinforcement learning marking a significant milestone in training sophisticated agents for complex tasks (Hagan et al., 2021). Machine learning's omnipresence in today's digital age cannot be overstated. From the simple task of email filtering to the more complex realm of real-time stock trading, ML algorithms silently power numerous aspects of daily life. The rise of digital platforms, social networks, and e-commerce giants all hinge on machine learning to offer personalized user experiences, optimize logistical operations, and innovate product offerings. Machine learning models help in predictive diagnostics, drug discovery, and patient management in healthcare. Natural language processing, a domain interlaced with machine learning, has revolutionized human-machine interaction, as seen with virtual assistants and translation services (Hagan et al., 2021).

2.5. Deep Learning

The concept of deep learning is embedded in the early studies of artificial neural networks (ANNs), which can be traced back to the perceptron models of the 1960s. The perceptron, proposed by (A. et al. et al., 2022), was a linear classifier that marked the Dawn of learning algorithms. However, its limitations, especially in handling non-linear data, led to skepticism around neural networks for several years. The revival of interest in neural networks began in the 1980s, primarily due to the introduction of the backpropagation algorithm, which made training multi-layer percepts feasible (Jiang et al., 2021). Over the subsequent decades, with the proliferation of computational power, larger datasets, and advanced training techniques, the depth of neural networks increased, giving birth to what we recognize as deep learning. Deep learning, a subfield of machine learning, leverages multi-layered neural architectures to automatically extract features f. data, progressing from simple to complex representations. The most popular architectures include Convolutional Neural Networks (CNNs)

and Recurrent Neural Networks (RNNs). CNNs, inspired by the visual cortex's structure (Munnangi et al., 2023), have revolutionized age and video recognition tasks. In contrast, RNNs, especially their advanced variants like Long Short-Term Memory (LSTM) networks, have greatly succeeded in sequential data tasks, including language modeling and speech recognition. Deep neural networks, while powerful, suffered from challenges like the vanishing and exploding gradient problems. The introduction of normalization methods, like batch normalization (love & Szegedy, 2015), played a pivotal role in stabilizing deep training. Architectures. Additionally, optimization techniques such as Adam (Kingma & Ba, 2014) have contributed to faster convergence during training. Regularization methods, including dropout (Srivastava et al., 2014), have been instrumental in preventing overfitting in large networks, making models more generalizable to unseen data. While the initial applications of deep learning were predominantly in image and speech recognition, its applicability has now permeated diverse fields. In healthcare, deep learning models assist in disease detection and drug discovery (Sharma et al., 2022). These models are utilized in finance for fraud detection, algorithmic trading, and credit scoring. Natural Language Processing (NLP) has seen transformative changes with models like BERT (Sharma et al., 2022), which have set new benchmarks in sentiment analysis, translation, and question-answering tasks. Despite its groundbreaking successes, deep learning has criticisms. The need for interpretability of deep models remains a pressing concern, especially in critical applications like healthcare and law (Doshi-Velez & Kim, 2017). The extensive data and computational power requirements Also, it makes deep learning less accessible for some applications. Moreover, adversarial attacks, where perturbations in the input data can drastically alter predictions, expose vulnerabilities of these models (Szegedy et al., 2013). The deep learning landscape continues to evolve, with newer architectures and training paradigms emerging. The pursuit of models that are both efficient and interpretable is ongoing. Transfer learning and few-shot learning, where models are trained to perform tasks with minimal data, are garnering attention. Reinforcement learning, coupled with deep learning, is opening doors to applications in robotics and autonomous systems. In conclusion, while deep learning has catalyzed significant advancements in Al, the journey is far from over, with exciting avenues to explore.

2.6. Neural Networks

Neural networks, often referred to as the foundational pillar of modem artificial intelligence, draw their inspiration from the intricate web of neurons in the human brain. Although the rudimentary idea of neural computation can be tracked as far back as the 1940s with McCulloch & Pitts' binary threshold neurons, it was in the 1980s that the backpropagation algorithm revitalized the interest in this domain. Despite the early enthusiasm, neural networks soon hit a winter due to the limitations in computational resources and training deep networks. However, with the advent of powerful GPUs in the 21st century, the groundbreaking work of Krizhevsky, and (Raheem et al., n.d.-a) used deep convolutional neural networks for image classification; there was a resurgent interest in neural networks, setting the stage for the current deep learning era. Over the years, various architectures of neural networks have been proposed, each tailored for specific tasks. Feedforward Neural Networks, the simplest arm, consist of layers of neurons where information moves in one direction. Then there is the Convolutional Neural Networks, which have gained immense popularity for image-related tasks due to their ability to automatically and adaptively learn spatial hierarchies of features. Recurrent Neural Networks especially their advanced variants like Long Short-term memory (LSTM) networks are designed to recognize patterns in data sequences, making them indispensable for j time series forecasting and natural language processing tasks. Moreover, architectures like Generative Adversarial Networks (GANs) have ushered in a new era of generative models capable of creating incredibly realistic synthetic data (Raheem et al n). Despite their promise and prowess, training neural networks is a challenging task. The infamous "vanishing gradient" and "exploding gradient" problems, especially prevalent in deep networks and RNNs, can severely hamper learning. Overfitting, where the model learns to perform exceedingly well on training data at the expense of generalization (M. M. Rahman et al., 2021) is another looming challenge. Techniques such as dropout, regularization, and early stopping have been introduced to combat this.

2.7. Convolutional Neural Network

The genesis of Convolutional Neural Networks (CNNs) is firmly grounded in studying the mammalian visual cortex. The intricate manner in which the neurons in the visual system detect spatial hierarchies inspired the layered architecture of CNNs. The term "convolutional" pays homage to the mathematical operation central to this neural(Habib & Rahman, 2021) network, an operation that allows for local receptive fields, shared weights, and spatial invariance. Although neural network-like structures, especially ones designed for visual tasks, were studied as far back as the 1980s, it was the work of (Kate & Shukla, 202 l) on the LeNet-5 model for handwritten digit recognition that. Be deemed as the first practical application of CNNs. The seminal paper proposed using gradient-based learning applied to document recognition, setting the stage for the vast developments in this domain. CNNs uniquely shine in processing data that a grid-like topology, such as an image, which can be viewed as a 2D grid of pixels. The core components of a typical CNN architecture include convolutional layers, poo~ layers, and fully connected layers. The convolutional layers, often several stacked together, apply a convolution operation to the input data and pass it through a non-linear activation function, like the Rectified Linear Unit (ReLU). The pooling layers serve to down-sample the spatial dimensions of the input. Over the years, numerous innovative architectures have emerged. The VGGNet (Assistant et al.) is recognized for its deep stacks of convolutional layers. The ResNet (He et al., 20 I 6) introduced the concept of residual learning, where "skip connections" or "shortcuts" allow gradients to flow through deeper networks, effectively addressing the vanishing gradient problem. The adaptability of CNNs extends far beyond mere image classification, the application of CNNs, starting with the breakthrough AlexNet (Habib & Rahman, 2021), drastically improved Classification accuracies. Besides image classification, CNNs have shown promise in object detection, image segmentation, and video analysis.

2.8. Summary

In Chapter 2's literature review, the foundational principles and advanced architectures of neural networks were explored, particularly emphasizing their applications in medical diagnostics. The chapter traced the evolution of these networks from their biological inspirations to modern implementations (Krishnaveni & Suman, n.d.-a), specifically spotlighting convolutional neural networks (CNNs) for their revolutionary contributions to image analysis in the medical realm. While underscoring the profound impact of neural networks in enhancing diagnostic accuracy, the review also sheds light on pertinent challenges, such as model interpretability and the need for extensive labeled training datasets (Krishnaveni & Suman, n.d.-b). Through an exhaustive exploration, the chapter set the stage for leveraging deep learning and IoMT for early-stage lung cancer detection (Furtado, 2021).

Chapter 3

RESEARCH METHODOLOGY

Chapter 3: Research Methodology

In this part, we will describe the application of IoMT and deep learning to identify COVID-19 and lung cancer in its early stages. Following is a description of the fundamental method for identifying lung cancer, which may be seen in Figure 3 .1. This section demonstrates one possible use of deep learning: diagnosing lung illness. The most important stages are listed below.

- Image processing in preparation for training in image classification
- Details of the New Procedures Being Considered
- The intended implementation will be carried out with the assistance of the resources specified below.

• Matlab and Python were used, along with several Python library packages such as Keras, Pandas, NumPy, and TensorFlow.
• The ML (Machine Learning) and NN (Neural Networks) Toolbox
• The Deep Learning Library (DL Library)

The distinction that can be drawn between a healthy lung and one that is unified may be employed in determining the severity of lung illness. In order to accomplish this goal, several different classifiers were used. The diagnosis of lung illness is made possible thanks to the development of a classifier. Training is what ultimately results in these models. During this process step, a collection of images trains a neural network. It is possible to train a model using deep learning, and that model might then use the images as its class labels. Consequently, photographs of lungs exhibiting the appropriate illness are compiled to use a deep learning approach to diagnose lung disease. The second step involves teaching a neural network to reorganize illness, and the third and final step involves classifying fresh photos. After training, the model is shown images it has never seen before and is tasked with categorizing them according to the applicable categories for each shot. The suggested technique makes use of motion analysis to forecast outcomes like the likelihood that a patient would get a positive test result for COVID. Using big datasets is unnecessary for machine learning, so it has found its way into many applications, including medical care and power production. Learning is applying the knowledge that was gleaned during the process of creating a large dataset to a more compact dataset.

3.1. The Beginning of the Image Acquisition Process

All images must be collected together before a computer can learn from previous instances and develop a model to place photographs into various categories. For a computer to rearrange the contents of an item, a variety of pictures are shown. The deep learning model may employ a few data types, including speech and time series data. When using this method, image data is preferred for the identification of the lung cancer. The data sources used in the model's training process are X-rays, CT scans, histopathology photographs, and sputum smear microscopy. Additionally, the output of this model is in the form of trained photographs that have been preprocessed.

3.2. Organization

At this stage, the picture is improved by being made of higher-quality materials so that a computer can comprehend it with better precision. A method known as CLAHE (Contrast Limited Adaptive Histogram Equalization) may be used to boost the contrast in an image. (Rajaraman et al., 2018). During the process of picture processing, the ROI (region of interest) may be used for the removal of bone and the segmentation of the lungs. (Gordienko et al., 2019). Using edge detection to present - native information is possible. The first step in the preprocessing procedure is called data augmentation, which is immediately followed by the feature extraction step. The expansion of data is made possible through data augmentation. Deep learning relies on feature extraction to single out characteristics exclusive to a certain category of objects. After this step, the output images were improved, and any unwelcome elements were eliminated. Additionally, this model may create visuals that, after being trained, have the learned shape.

3.2.1 Guidelines

This level has three unique components: transfer learning, deep learning, and ensemble. Deep learning incorporates several approaches, such as RNN (recurrent neural network), MPNN (multilayer perceptron neural network), DBN (deep belief network, and CNN (convoluted neural network ID. There are many different sorts of data that perform well in certain computations. CNN is excellent at image recognition. When picking a deep learning calculation, the idea of the most recent information should be addressed. The process of transmitting information from one model to the next is referred to as "move gaining," and that word refers to it. The participants discuss the many models used throughout the categorization process. Using machine learning and collection methodologies reduces the time spent on preparation, boosts characterization accuracy, and reduces over fitting.

3.2.2 Classification

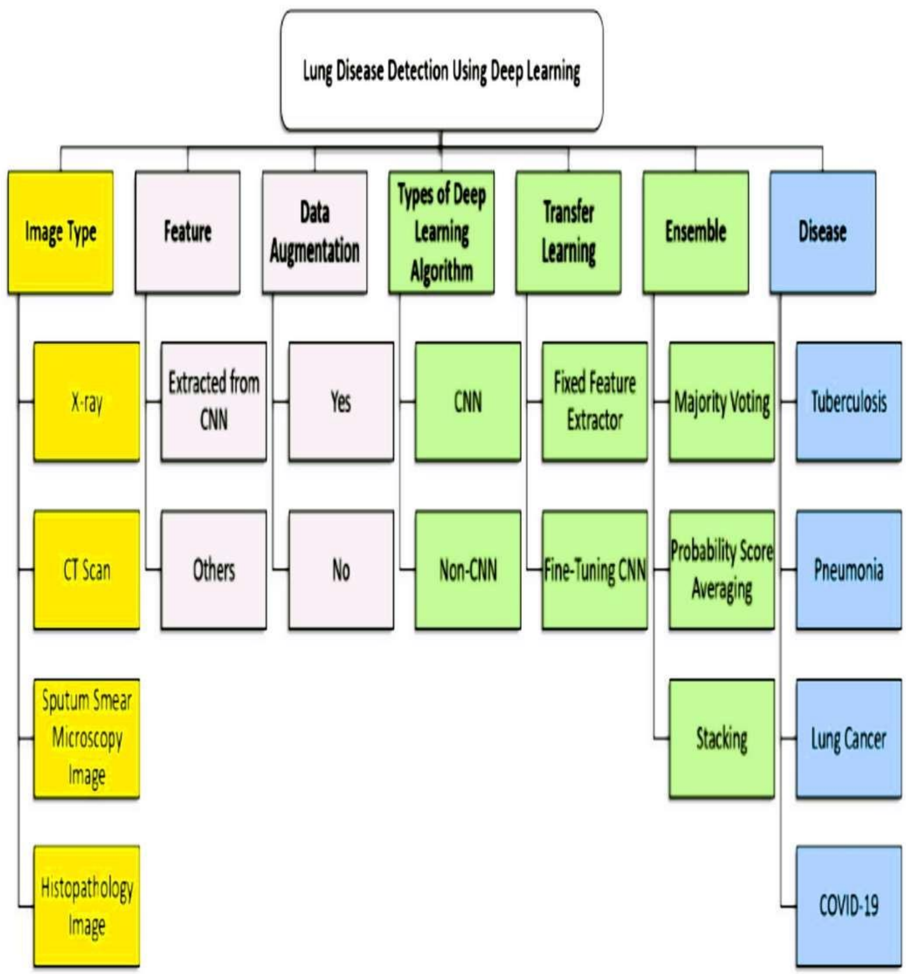
Currently, the trained model guesses the image's class by producing a classification prediction. Despite this, the model is taught using many different X-ray pictures of healthy lung tissue and photos of malignant or COVID-19-infected lung tissue. After this, it shows brand new images that it has never seen. A trained model can differentiate between recently taken images of healthy and sick lungs, categorize the photographs, and place them in the proper category. In addition, the model generates a score for each image that indicates the chance that it belongs to a certain category. Images are classified using this approach by looking at the likelihood scores associated with each.

3.3. A Taxonomy for Differentiating the Types of Pulmonary Disease

Seven credits were awarded after serious consideration was given to the scientific categorization. These credits were selected because their presence was imminent, and they could bring everything together. When thinking about scientific classification, the following categories emerge picture kinds, highlights, information expansion, different types of deep learning calculations, motion learning, the classification group, and various pulmonary conditions. There will be a presentation of the datasets used in the scientific categorization of the new pulmonary illness detection using deep learning. Lung disorders, often called respiratory infections, are quite common among travelers traveling by airplane. According to research published in 2017 by the Forum of International Respiratory Societies, there are roughly 334 million individuals throughout the globe who have asthma, and there are around 1.4 million people who pass away each year as a result of tuberculosis (TB). Pneumonia is another leading factor in the overall death rate. The COVID-19 worldwide pandemic is having an impact (Organization, 2020). It is without a doubt the case that lung disorders are among the primary reasons for death and disability on our planet. Recent studies on using deep learning in clinical imaging to detect infections, particularly pulmonary illnesses, have shown enormous potential. Figure 3.2 presents a categorization system for lung cancer, which may be found in the following text.

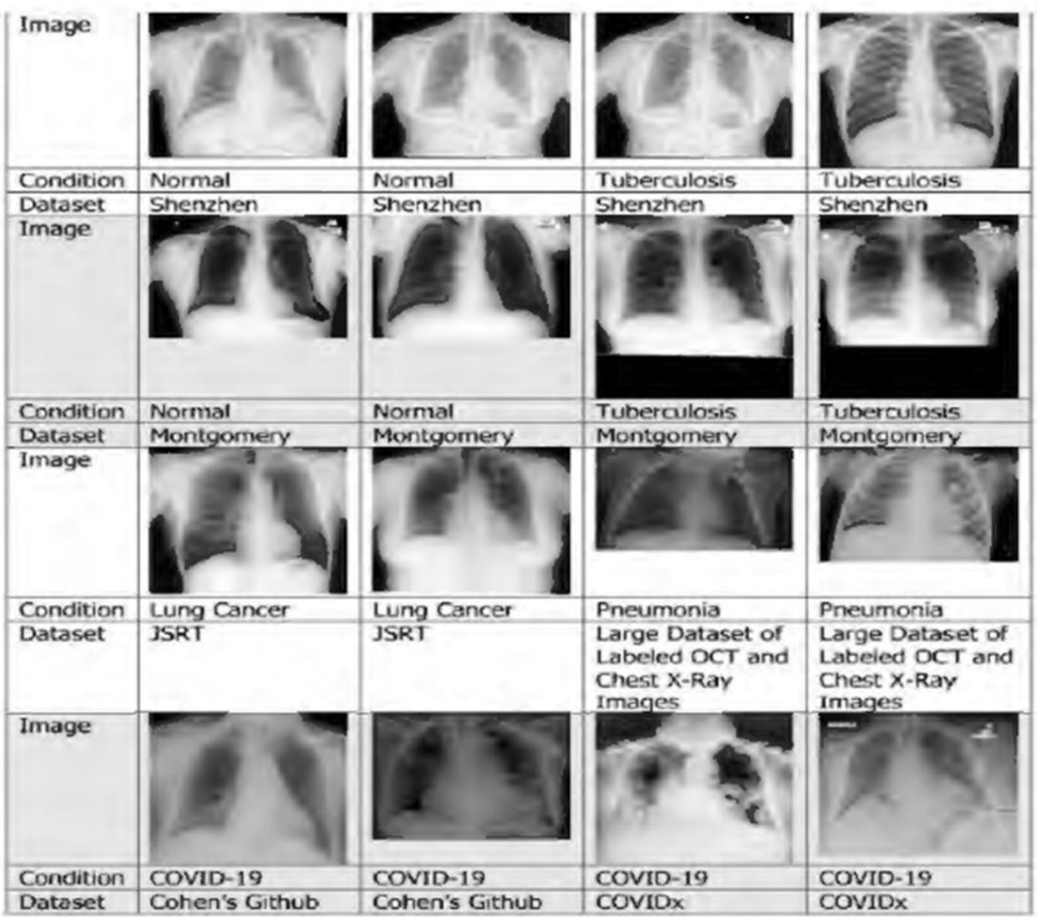
3.4 Format Of Images

As mentioned earlier, the model is educated using images of histology, sputum stain microscopy, chest X-rays, and CT scans, respectively. MRI and PET are only two of the many other imaging modalities that are accessible. The health and condition of a patient may be diagnosed using either method, and the effectiveness of an existing drug can be evaluated using either method as well.



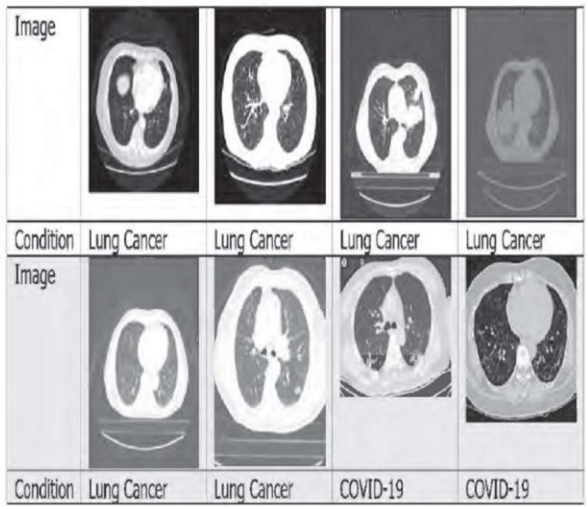
3.4.1 Chest X-rays

It is an analytical test that could assist medical professionals in correctly detecting and treating clinical problems. Because they offer pictures of the veins, heart, arteries, lungs, spine, and thoracic bones, thoracic X-rays are the most effective clinical X-rays. Images obtained from clinical X-rays are typically shown as animations that may be graphically manipulated. According to Heneghan et al.'s (2006) research, the issue was fixed by using a computed X-ray. The thorax X-rays in Figure 3.3 were taken from numerous datasets that included various lung illnesses.



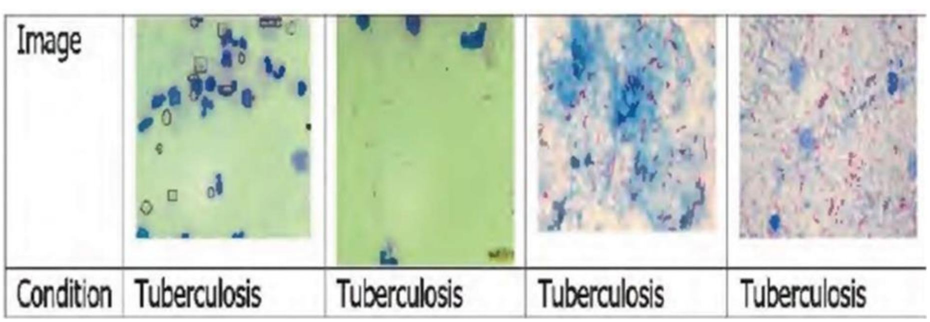
3.4.2 CT scanner

During a CT scan, segmented radiography pictures are generated using computer processing at various points throughout the operation. (H. Shan et al., 2020) The patients' bodies are photographed from multiple perspectives during the whole process. According to Song et al. (2017), the photos of the patient's lungs may be exhibited separately or in clusters with images of the patient's other organs, tissues, bones, and anomalies found inside the patient's body. An X-ray may not be as exact as a CT scan's findings. A CT scan has the potential to identify a wide variety of lung diseases and conditions, including lung cancer, COVID-19, and TB. The pictures of the CT scan of the lungs are shown in the figure.4.



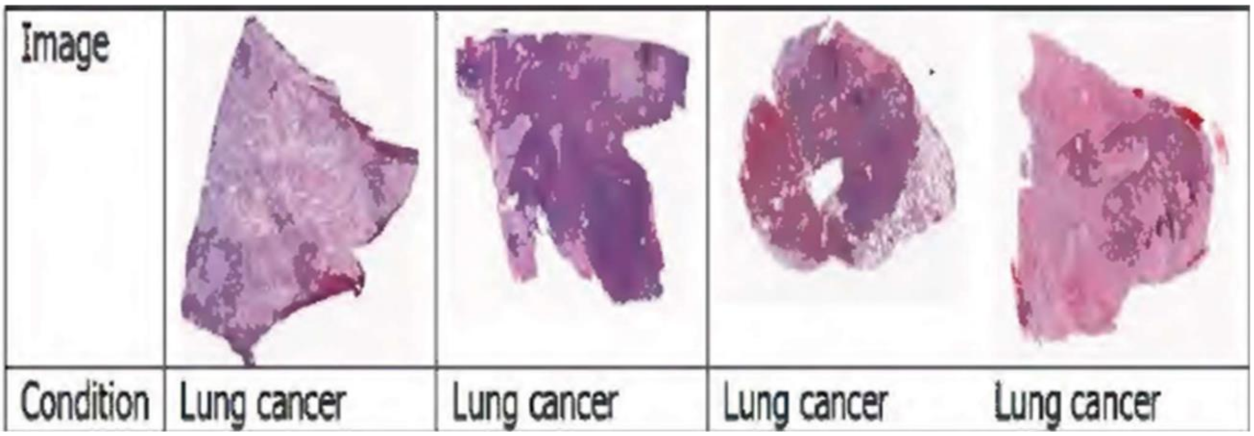
3.4.3 Images Obtained from Sputum Smear Microscopy

Sputum is a thick, viscous fluid that may be found not only in the lungs themselves but also in the airways of the lungs. In the process of analyzing sputum, a slide is given a covering of mucous that is thick and sticky (Shah et al., 2017). Figure 3.5 illustrates a sputum stain microscopy example.



3.4.4 Photographs from the Field of Histopathology

A tissue sample removed surgically or by a biopsy is examined under a microscope in histopathology. To improve one's vision, different tissue components may be dyed to provide a variety of shades. The images of histopathology are shown in Figure 3.6.

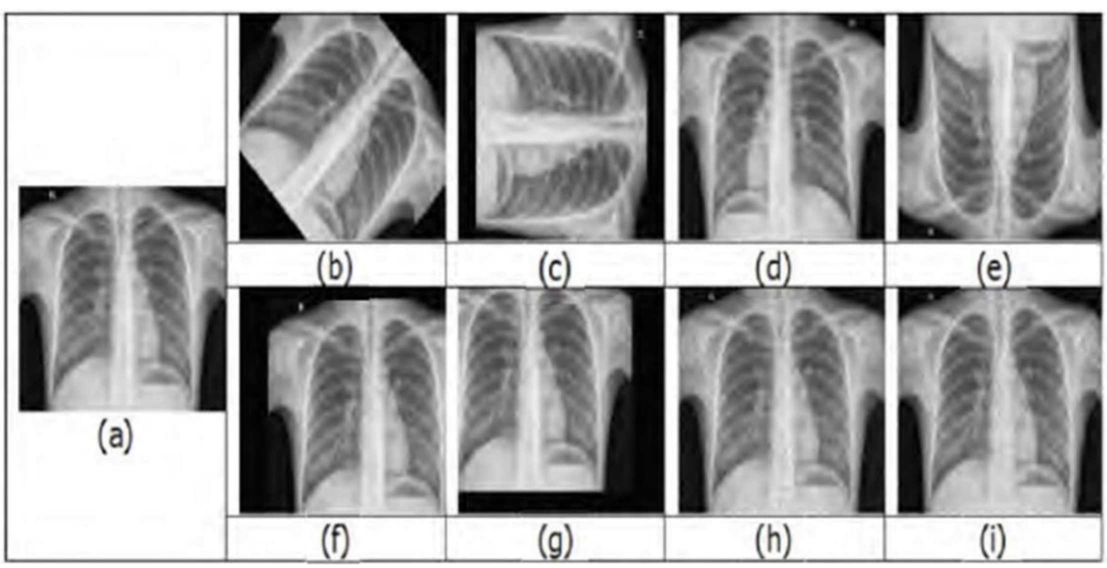


3.5 Features

According to Mahony and colleagues (n.d.), features are important information retrieved from pictures and displayed as numeric data. This information may be used to solve arrange of issues that are associated with computer vision. An image may have many different aspects, such as color, size, form, points, borders, etc. The picture's nature determines the information included in the feature. When discussing feature transformation, we mean generating brand-new features from pre-existing ones. It is possible that the new qualities do not accurately replicate the old traits in their original form, and they certainly have a stronger bias than the older ones did. When using machine learning to detect different aspects of an image, the transformation of features is necessary. The attributes of machine learning are outlined in the following paragraphs. (LBP) Binary patterns in the local environment show characteristics of the global picture. An example of an edge histogram descriptor is the Tamura texture descriptor (TTD), which is often referred to as the Tamura texture descriptor (EHD). A fuzzy color and texture histogram, often known as FCTH, is a kind of color arrangement that may be described. The letters CEDD represent the words "edge direction and color descriptor" in an abbreviated form. The phrase "scale-invariant feature transform". Curvature histograms, also known as CD histograms, and shape descriptor histograms, sometimes known as SD histograms, are the two types.

3.6 Data Improvement

It is necessary to undergo training in deep learning accuracy depending on the number of images to work with this enormous dataset. According to Pedro (2012), the accuracy of an algorithm that is powerful but only uses a small sample size of data is worse than that of an algorithm that uses a large sample size. During the training process for binary classification, a problem might occur due to an imbalance in the class. An imbalanced class contains more samples than other classes but has fewer samples overall. After training in binary classification with an imbalanced class, there is a possibility that the model is biased. When there was an equal amount of data for each category, deep learning algorithms could function at their highest efficiency level. Picture augmentation is a method that makes variations of an existing picture by splitting it into several images. This method is used instead of the traditional method of adding new photos to the training dataset. According to Grochowski (2019), the following processing makes use of a variety of different methods. Translations Zooms, rotations, and pans are all available. The amount of data relevant to the original data may be increased with the help of noisy data improvement. This is the object of the process. Image augmentation results in various pictures, some shown in Figure 3.7.



3.7 Training

In this case, gradient descent is used to adjust the NFC settings. The sum-squared error (E), often known as E, is the error that occurs when the output of the classifier is compared to the objective. The major goal of the training phase is to decrease this error as much as possible. The following is the definition of E when it is applied to the training instances (Q) and the output instances (K):

$$E = \sum_{q=1}^{Q} \sum_{k=1}^{K} (t_k{}^q - z_k{}^q)^2$$

Augmentation significantly improves overall performance and particular characteristics and patterns. In addition, data augmentation has the potential to reduce overfitting. When a system learns a function that needs significant modification, such as training the data for modeling, overfitting may arise due to this learning process. (Short & Khoshgoftaar, 2019). Data augmentation may prevent overfitting, which involves providing more varied data to the model. This variety of data helps lower the model's volatility while improving its generalization ability. On the other hand, data augmentation can only properly segment some datasets (Shorten & Khoshgoftaar, 2019). The training time is When more data is added, greater memory costs and higher conversion calculation costs are incurred. These are the drawbacks of data augmentation. The neural network looks for the differences between classes that are the most evident, which is why the prediction is incorrect. Because the neural network looks for the clearest distinctions, it makes an incorrect forecast. CNN, which is now the most well-known deep learning algorithm, does very well when it comes to spotting patterns in photos. CNNs, much like the brain, are made up of trained networks with weights and biases associated with the~ multiple inputs received by every neuron. The subsequent step is to compute the weighted sum of the in units. The activated function is then responsible for producing an output after receiving the weighted sum as an input. CNN stands apart from other types of neural networks because of its convolutional layers.

Figure 9 is an illustration of the CNN design that is described in Shea and Nash (n.d.). The following is a list of the three primary layers of CNN.

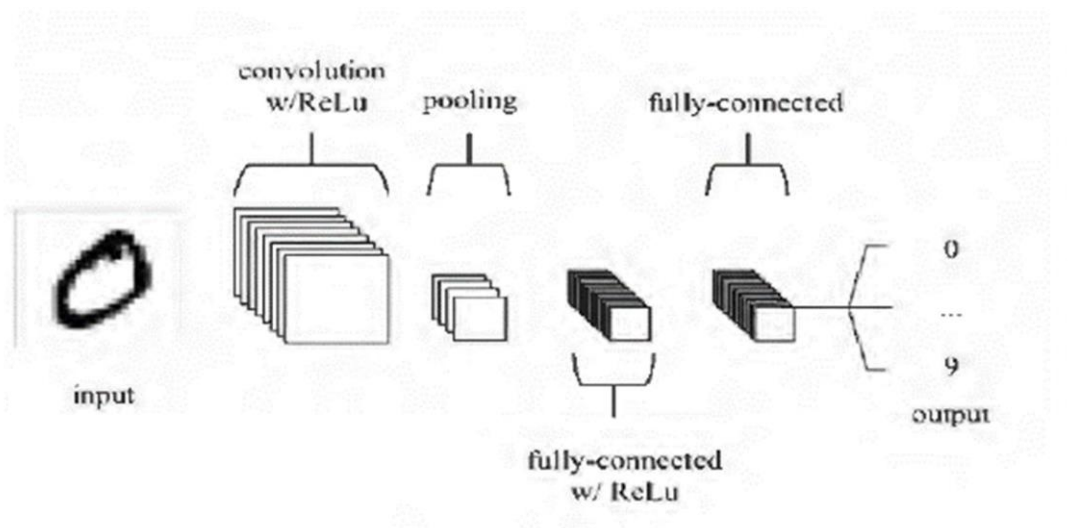
- Convolutional
- fully-connected
- Pooling

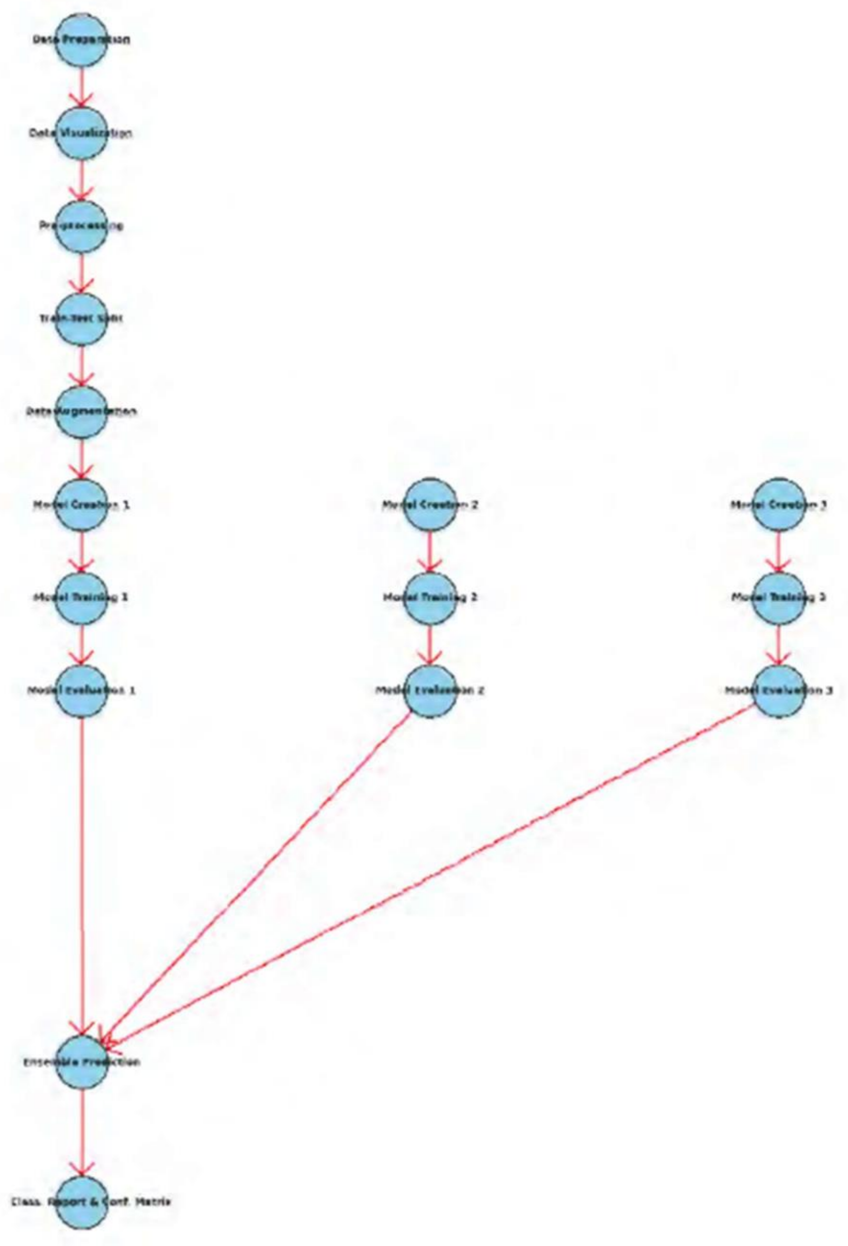
Convolutional: The Convolutional methodology primarily refers to applying convolutional layers in deep learning, specifically designed for processing data with a grid-like topology, such as an image. In the context of our research on lung images, convolutional layers play an indispensable role. They function by sliding a filter or a kernel over the input data (like an image) to produce a feature map, capturing the localized features of the input. This methodology allows the network to focus on small input regions at a time and learn spatial hierarchies. The significance of convolutional layers lies in their ability to learn and detect patterns, edges, textures, and more complex structures in the images. This becomes especially critical when distinguishing between benign, malignant, and normal lung images. Their weight-sharing mechanism also reduces the number of parameters, making the model more efficient and less prone to overfitting. By leveraging convolutional layers, the model can generalize well across different sections of an image and recognize patterns irrespective of their position, making it a fundamental component in our methodology for lung image classification.

Fully connected: The fully connected methodology, often called dense layers, is a pivotal stage in deep learning architecture, especially after the initial layers (like convolutional layers) have done their part in feature extraction. In our research, after the convolutional layers process the lung images and extract vital features, these features are flattened and passed onto fully connected layers. Each neuron in a fully connected layer is connected to every neuron in the previous layer, and it is where the final decisions are made based on the patterns recognized in the image. For instance, in these layers, the model might decide whether a given lung image depicts a benign or malignant tumor or is entirely normal based on its recognized features. The weights in these layers are adjusted during the training process to minimize the classification error. In essence, while the convolutional layers are tasked with identifying patterns and features in the images, the fully-connected layers interpret these patterns and produce a final classification result, making them crucial for the success of our model.

Pooling:

Pooling, often nestled between successive convolutional layers, serves a dual purpose in our lung image classification model: spatial dimension reduction and feature extraction enhancement. The most common form we utilized is max-pooling, where the model inspects a region of the feature map produced by convolutional layers and selects the maximum value from that region. Doing so reduces the spatial dimensions of the feature map, ensuring computational efficiency and reducing the model's susceptibility to potential overfitting. Furthermore, pooling layers help make the model's feature extraction process more invariant to slight shifts and distortions in the image. For lung image classification, this is especially significant. Lung images can vary due to numerous factors, such as different angles of X-ray or CT scan shots, slight movements of patients, or inherent differences in anatomy. By integrating pooling layers into our methodology, we ensure that the model remains robust and accurate, even when faced with these minor variabilities in the input images.





Methodology Flowchart with Parallel Model Creation and Evaluation

A systematic flowchart titled "Methodology Flowchart with Parallel Model Creation and Evaluation." The flow begins with the "Data Preparation" phase, which then moves linearly downward to the "Data Visualization" step, highlighting the initial exploration of the dataset. After this, the "Pre-processing" phase occurs, where data goes to initial transformations. This step is followed by the "Train-Test Split" stage, where the dataset is divided into portions for testing purposes. After partitioning the dataset, it proceeds to the "Data Augmentation" phase, indicating the enhancement of the dataset by generating new instances via techniques such as rotations, zooming, or flipping of existing data points. The flowchart then diverges into three parallel paths for model creation and evaluation. Each path represents a distinct model development procedure: "Model Creation 1 ", "Model Training 1 ", and "Model Evaluation 1" for the first path; "Model Creation 2", "Model Training 2", and "Model Evaluation 2" for the second path; and "Model Creation 3", "Model Training 3", and "Model Evaluation 3" for the third path. These parallel paths underscore the simultaneous development and assessment of three unique models. Post individual evaluations, all three models converge into the "Ensemble Prediction" phase, where the predictions from all three models are combined to achieve a more robust and accurate output. The flowchart concludes with the "Class. Report & Conf. Matrix" phase, signifying the generation of classification reports and confusion matrices to gauge the performance of the ensemble model. The diagram illustrates a structured methodology, emphasizing parallel model development and ensemble techniques to enhance prediction accuracy.

Feature/Model	Model 1	Model 2	Model 3
Input Shape	(256, 256, 1)	(256, 256, 1)	(256, 256, 1)
Convolutional Layer 1	32 filters, 3x3	64 filters, 5x5	16 filters, 7x7, strides=2
Activation 1	ReLU	ReLU	ReLU
Pooling Layer 1	MaxPooling 2x2	AveragePooling 2x2	N/A
Convolutional Layer 2	64 filters, 3x3	128 filters, 5x5	32 filters, 5x5
Activation 2	ReLU	ReLU	ReLU
Pooling Layer 2	MaxPooling 2x2	AveragePooling 2x2	MaxPooling 2x2
Convolutional	N/A	N/A	64 filters, 3x3

Feature/Model	Model 1	Model 2	Model 3
Layer 3			
Activation 3	N/A	N/A	ReLU
Convolutional Layer 4	N/A	N/A	128 filters, 1x1
Activation 4	N/A	N/A	ReLU
Pooling Layer 3	N/A	N/A	MaxPooling 2x2
Flatten	Yes	Yes	Yes
Dense Layer 1	128 neurons	256 neurons	256 neurons
Activation 5	ReLU	ReLU	ReLU
Dropout	N/A	0.5	0.3
Output Layer	3 neurons	3 neurons	3 neurons
Activation (Output)	Softmax	Softmax	Softmax
Loss Function	Sparse Categorical Crossentropy	Sparse Categorical Crossentropy	Sparse Categorical Crossentropy
Optimizer	Adam	Adam	Adam

The provided Table revolves around the preprocessing, analyzing, and modeling of lung cancer image datasets. The objective is to distinguish between three categories of lung health, namely, Benign, Malignant, and Normal cases. Initially, the code contains necessary library imports and establishes settings out-of-memory errors for GPU usage. The dataset, located in a specified directory on the user's system, is parsed to understand image dimensions. The images are then visualized using the OpenCV library, showcasing individual photos from each category. A consistent image size of 256x256 pixels is used to standardize the dataset. Images undergo two preprocessing stages: resizing and Gaussian blurring. After preprocessing, the data is partitioned into features and labels. The features comprise the image dt: J'1nd and the labels represent their respective categories. The features undergo normalization, and the dataset is split into training and validation sets. The code also identifies and addresses data imbalance using the SMOTE algorithm. Suppose one class is overly represented in the training data (exceeding 60% of the samples). In that case, the SMOTE technique oversamples the under-represented courses to balance the dataset, ensuring each class has an equal representation. For model training, the Keras framework is used to design three Convolutional Neural Network (CNN) models. Each model varies in its architecture, utilizing different combinations of convolutional, pooling, dense, and dropout layers. The models are compiled using the Adam optimizer, sparse categorical cross-entropy as the loss function, and accuracy as the evaluation metric. Additionally, the code employs callbacks such as Early Stopping to prevent overtraining to adjust the learning rate based on validation loss, and a custom callback to halt training if 99% validation accuracy is achieved. The three models are trained using the preprocessed training data, and their performances are evaluated on the validation set. The user can gauge the best architecture based on the resulting validation accuracies. An ensemble prediction mechanism is hinted at the end, suggesting the potential amalgamation of predictions from all three models to arrive at a more robust and accurate prediction.

3.7 Convolutional

Layer types include pooling layers and linked layers. The process of involution as a technique The algorithm used in this layer is convolutional. Convolution may be considered a linear process in which the input is multiplied by a set of weights. Weights are the parts that make up a kernel or filter. The quantity of data entered is more than the filter can handle. The representation dimension is scaled down progressively by combining levels as the number of parameters and variables in use falls. It is possible to avoid overfitting by doing out calculations inside the network. The output of the initiation created by the layer below is sent into a straightening unit, which then applies an element-wise CNN performing capability, such as the sigmoid, to the result. Order and extraction are the two parts that makeup CNN's perception of the situation. Convolution is the technique through which information is sent down the channel during element extraction. Then, an extra element map will be constructed. During the phase when the picture is being arranged, CNN keeps track of the possibility that a certain category or name will be connected with it. According to El Allali (2019), CNN is helpful for picture identification and organization since it automatically identifies the highlights without needing human component extraction. This makes CNN an ideal tool for these tasks.

3.8 Transfer Learning

DBN are equivalent to manipulate RBMs (Restricted Boltzmann Machines) (Lanbouri & Achchab, 2015). The DBN layers have finished their assigned jobs except for the beginning and finishing levels. For instance, to fulfill both jobs, the layer performed input layers as nodes that came after it and hid layers as nodes that came before it. Items that come after the previous one in the connected list. The initial step in developing a training plan for DBNs involves replicating the input in as close to an exact manner as is practically possible. In light of this, the hidden layer of the first RBM is treated as if it were the visible layer for seconds, and the outputs of the first RBM are used to train the second RBM so on. This process will be carried out in an endless loop. Until all network levels have been taught, other organizations are free to utilize the DBN-developed model Using databases to store motion-capture data. (Daz-Huerta et al., 2019) More information on DBN can be accessed here. Using a bundle of words was yet another method tested in the context of deep learning. Machine learning is the foundation of the BOW technique, which extracts textual features to use them in modeling. Broadcasting on the Web is what is meant to be abbreviated as "BOW." BOVW is an abbreviation for "the properties of images as words." The primary focus here is on amassing features, with the understanding that each feature stands for a separate element of the picture. This section covers descriptions of crucial components. When an image is rotated, shrunk, or expanded, the transformations do not affect the key points, which are the components in a picture that are most easily recognized, such as a person's face. A descriptor is any element that describes a topic. It provides a concise explanation of the fundamental idea. Concepts and descriptors that are fundamental to the construction of vocabularies and other sorts of information are used. Each picture's components' frequency must be shown in a histogram representing the image. A wealth of information may be gleaned by examining the histogram of frequency. Find more photographs comparable to the present one, or make an educated guess about the image's category.

3.9 Classifiers as a Group

Transfer learning is widely used in computer vision (Cao et al., 2013) because it makes creating correct models possible. A model that has been learned in the past for One study area may be transferred to another using the transfer learning process. This may be done with or without the assistance of a mastered model. A model was created to solve an issue similar to the discussed one. Instead of starting from scratch when developing a new model for an analogous job, a new model will be trained specifically for the new activity. Finding the ideal proportions of each weight When training a deep learning model, the goal is to discover the ideal weights for the network by going through a large number of iterations in both the forward and backward directions of the network. The weights and architecture that previously trained models on enormous datasets created may be used and applied to the current job, and it is conceivable to do so. Pretrained models can lower the expenses of training new models, which may be rather considerable. Rodriguez (2007). The vast majority of pre-trained CNN designs are built using ImageNet as their basis. Human laborers gathered photos from around the internet and assigned names to them via the crowdsourcing platform Amazon Mechanical Turk. They did this by using photographs. The corporation claims that the ILSVRC uses a section of ImageNet, which comprises around 1,000 photos for each of its 1,000 categories and is used by the organization. A total of around 1.2 million Training photos. Fifty thousand validation images and 150,000 test images are available to users. The following are the two main uses of transfer learning:

3.10 Fine-tuning

CNN is used to extract features. Only the model's uppermost layers are retrained over time, while the weights of the model's early layers remain relatively stable throughout the process. This is because the features derived from the first layer are general nature (for example, edge or color clump detectors), making them suitable for various plications. To guarantee that the model has learned high-level characteristics relevant to the new dataset, it is important to retrain the upper layers of models that have been trained in the past. Using a large training dataset that is virtually similar to other datasets is common advice. First, remove the topmost layer that is completely linked, then extract network characteristics using a fresh dataset. When a small dataset is available for a project, it is common practice to advocate using features learned on a large dataset that a model trained in that domain. This is the case even when there is only a small dataset available. After acquiring features, they are expanded to train a classifier (Gonzalez, 2007).

3.11 The fundamental makeup of the Illness

Deep learning algorithms are now being used in the diagnosis of lung diseases such as tuberculosis (TB), lung cancer, pneumonia, and COVID-19, all of which are very contagious and widespread. These diseases were taken into account because the most common causes of severe sickness and death are associated with the airways (Forum of International Respiratory Societies., 2017), and also because COVID-19 is a pandemic that is currently Happening. In addition, I realized that the bulk of the prior study had concentrated. In the identification of these specific lung-related illnesses.

3.11.1 Lung Cancer

There is a kind of cancer known as lung cancer that has the potential to cause obstructions in the airways. The presence of lymph nodes in and around the lungs is another key symptom that a patient has lung cancer. According to Hua et al. (2015), lung nodules are a collection of several solidified tissues. On CT scans, a patient may see two distinct types of modules: malignant (cancerous) and benign (noncancerous). As early as 2015, researchers Song et al. (2017) could categorize nodules in CT images using prototypes of CNN and DBN. Radiology was the journal in which their results were published. Deep learning was used to quickly get information about the nodules of the lungs, including how the nodules developed and how they looked. (Kurniawan et al., 2020) Were the ones who made the discovery using CT scans. According to the findings of Song et al., 2017, stacking autoencoders and deep neural networks could have done better than CNN regarding the accuracy of their classifications. Regarding categorization, it was decided that CNN provided the most accurate information out of the three of them. Nodules that were solid and not solid were separated into separate groups. In particular, they came up with the same time and then calculated how probable it was that the nodule belonged to each of the six groups. They did this at the same time. By doing things this way, CNN saves time by not having to look at each picture separately. Image de-noising was performed by (Shakeel et al., 2019) to improve the resolution of the photos. Afterward, they segmented the lungs using a better version of profuse clustering, which they reported in Scientific Reports. There are four components to the method developed by Kim et al., 2020 The first is a lung segmentation model. These two models' findings were considered since they were used to help generate the overall forecast. Since these two models were used to help construct the overall estimates. According to the findings of their study, to locate nodules, the researchers first carried out nodular enhancement and then module segmentation.

3.11.2 COVID-19 or the Coronavirus

An infection caused by a newly discovered strain of the coronavirus known as COVID-19 is, at the base, an infectious sickness. People who are elderly and people who have a history of certain medical conditions, such as cardiovascular disease, chronic obstructive pulmonary disease, cancer, or diabetes, are at a higher risk of getting serious illnesses. CNN with data augmentationand transfer learning were both components of a core technique employed to identify COVID- 19, as stated by Salman et al. Utilizing Inception V3 as a static feature extractor was decided to accomplish transfer learning goals. (Narin et al., 2020) CNN and transfer learning were used to classify X-ray pictures into one of three categories: normal, malignancy, or COVID-19. They found that CNN with transfer learning performed the best out of all options. A CNN model enhanced with transfer learning and data augmentation was used to classify X-ray images into three distinct categories: standard images, malignancy images, and COVID-19 images. The techniques of augmentation that was used were rotation, scaling, and translation, respectively. (L. Wang et al., 2020) completed a CNN model with data augmentation and utilized it to solve an issue with data collection to achieve three-class classification. This was done to accomplish a three-class classification. Some enhancement ways are transcription, rotation, longitudinal reversal, and intensity shift. Image segmentation and extraction of features were both accomplished with the assistance of VB-Net, while the classification process was carried out using an adjusted random choice forest technique. A random forest model was additionally trained using several attributes developed by hand. This was done in addition to the training that was already mentioned. Two basic system components look at CT images: the detector and the computer. Subsystem A, which carried out a three-dimensional analysis, discovered nodules and focal opacities inside the case volume.

Chapter 4

RESULTS AND DISCUSSION

Chapter 4: RESULTS AND DISCUSSION

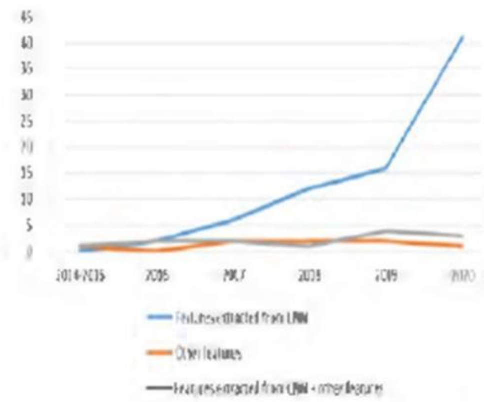
These are the findings encountered while diagnosing pulmonary disease. An in-depth analysis of the characteristics is discussed in the tendencies section to demonstrate how objectives and approaches have changed over time. Based on a previously established pattern, the subsequent step may be anticipated.

4.1. An analysis of the diagnostic procedures

Over the previous year, in each of these different environments, we investigated the published research on diagnosing lung diseases in the past.

4.2. An exploration of the many ways in which individuals make use of images

The Figure depicts the steady ascent in demand for X-ray images throughout the study. Even though there was a small drop in the total number of CT scans conducted in 2018, the overall trend is still expanding. The trend provides evidence that it is garnering more and more attention. Because not enough research has been done on the topic, the results of sputum stain microscopy and histology cannot be classified as anything other than "others." The next decrease will be less severe than the previous one. Imaging modalities such as CT scans and radiography lead to a bright future for the diagnosis of lung illness supported by deep learning. In the bulk of the study, which was seventy-one percent, X-ray photographs were employed, and CT scans were used in the remaining fourteen percent. It finished in second place, with 23% of the total vote. The fact that this information is simple to identify, get, and cow on may be a reasonable argument for why CT scanners are preferred over X-ray machines. X-ray machines can only provide two-dimensional images. The COVID-19 pathogen has widespread distribution around the globe, which is the root cause of the problem. Instead of CT scans, medical photos are used to identify respiratory problems. Medical pictures are far more accurate. CT scans may still be combined with various other diagnostic methods as necessary. They are superior to X-rays in that they display greater detail. Figure 4. I (b) illustrate the progress of approaches that have been developed for the diagnosis of lung illnesses. 4.1 (b) an emetic into learning has been enhanced by including 6iJ new ima0 e types, which has reduced the difficulty of diagnosing lung illnesses.



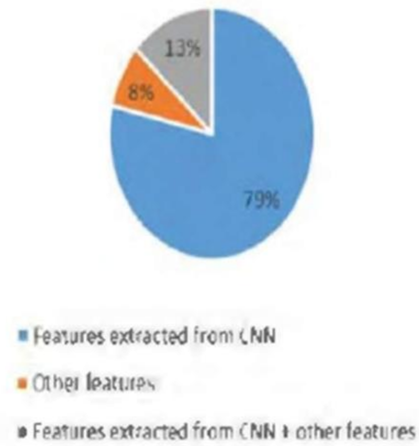


Figure 4.1(a) shows the trends in lung disease identification has increased in recent years

4.1 (b) In recent years, deep learning's spread of image type helped identify lung illness.

4.3. Feature Development

4.3.1 Conducting a Trend Analysis of Feature Development

In recent years, the use of particular features to detect pulmonary illnesses has reduced, while the utilization of characteristics created by convolutional neural networks (CNNs) has grown. This is seen in Figure 11 a. Additionally, a minuscule percentage of the characteristics created by CNNs are combined with other attributes. CNN is responsible for this particular situation. Through autonomous feature extraction, this approach eliminates the need for manually produced features. The many methods of implementation The majority of the findings from this study lend credence to using CNN characteristics in the diagnostic process of lung diseases. On the other hand, the relevance of other characteristics was lower. Figure 4.2b illustrates how various types of labor may be carried out in multiple ways. A description of the characteristics that make CNN unique was provided in seventy-nine percent of the sources we investigated. The alarming rise in the number of people suffering from lung disease over the last several decades is seen in Figure

4.2(a). In recent years, the application of Deep learning to augment data has greatly contributed to simplifying the diagnostic process for respiratory illnesses.

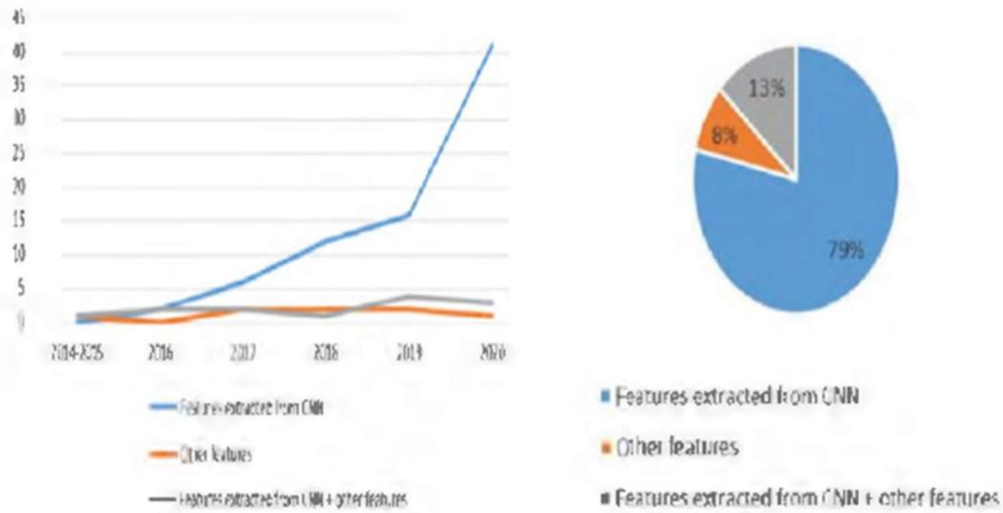


Figure 4.2 (a) shows the trends in lung disease identification has increased in recent years
Figure 4.2 (b) In recent years, deep learning's spread of data augmentation helped identify lung illness.

4.3.2 Conducting a Trend Analysis of Feature Development

Figure 4.3a shows a rising trend in the total use of data enhancement. Even though the number of jobs employing data augmentation has expanded significantly over the past several Preparing for interviews and other job-related tasks has been more difficult for years. A growing body of research has highlighted the need for improved data for training algorithms to identify lung diseases. The value of lung illness diagnosis to deep learning is shown in Figure 4.3b. Only about a third of the studies resulted in better results. Most research does not do this despite data showing it increases categorization accuracy. The complexity of implementing solutions for improving data quality might be to blame. There are certain drawbacks to data enhancement, such as the need for extra memory, the expense of processing transformations, and the length of training time.

4.3.3 Operation contracts

Recent advancements in deep learning, as shown in Figure 4.4a, have made lung disease identification much easier. Figure 4.4 depicts CNN's use of deep learning. A growing body of evidence supports CNN's application in diagnosing lung disease. Hence, this development is likely to continue. Recent studies have investigated the potential of CNN for diagnosing lung diseases, and the results are shown in Figure 4.4b. Most news organizations we polled said CNN is where they go first for breaking news. This is because CNN is effective and yields reliable categorizations. CNN is helpful since it already features feature extraction and transfer learning features.

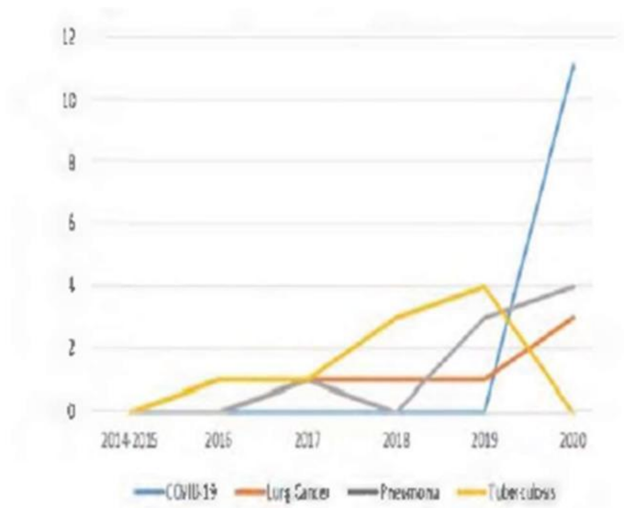


Figure 4.3 (a) shows the trends in data augmentation has increased in recent years

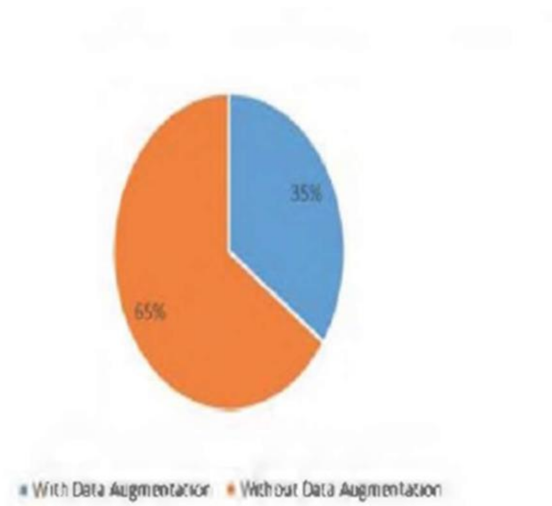


Figure 4.3 (b) In recent years, deep learning's spread of data augmentation helped identify lung illness.

4.3.4 Applying Prior Knowledge to Evaluate Present Occurrences

Figure 4.5a of the Trend Analysis reveals that transfer learning has grown over the last several years. Recently, the concept of "transfer learning" has gained increased attention. Possible new prototype creation using this approach. What you learn in one situation could help you out in another. This may help you better organize your thoughts for the new pursuit in many circumstances. This might be due to the model's improved flexibility due to exposure to more information during training. Transfer learning is often used in CNN-based works, as can be shown in Figure 4.5b. More recent studies (57%) also included transfer learning, as seen by the curve.

The percentage of work that uses transfer learning has been on the rise (as shown in Figure 4.5a), although it has only recently reached 57%. This indicates the slow but steady growth of transfer learning in this area. Transfer learning is still among the most effective tools for diagnosing lung disease. As a result, transfer learning will likely play a larger role in task allocation shortly.

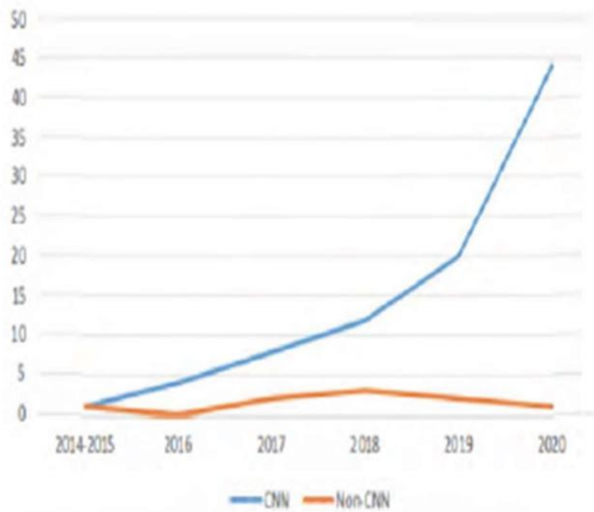


Figure 4.4 (a) shows the trends in deep learning algorithms has increased in recent years

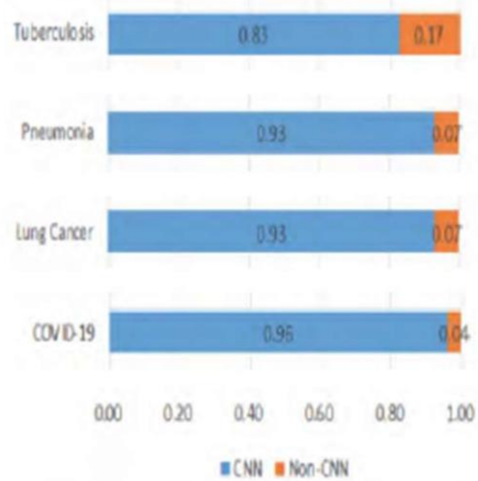


Figure 4.4 (h) In recent years, CNN utilization in deep learning has enhanced lung illness identification.

4.3.5 Using deep learning and analysis, lung disease may be diagnosed.

Figure 4.5a shows the rising use of deep learning techniques for identifying lung diseases. The study of TB is well-developed. Scientists can learn about lung cancer better since several diagnostic imaging technologies are available. Therefore, a second thorough checkup was performed. In the years running up to 2020, disease research funds were drastically cut to aid in the quest for COVID-19. Figure 4.5b depicts the deep learning-detected diseases' development over time. The study's ultimate goals were to combat lung cancer, tuberculosis, pneumonia, and COVID-19. Applying deep learning to identify TB may make a difference in diagnosis timeliness. The diagnosis of COVID-19 has been the topic of several papers since it is a relatively new condition. Therefore, Covid-19 searches are the second most common kind of industry research. Talking about it, 4.4% MATLAB R2014a was used to execute the suggested research plan. Images of lung CT scans stored in the LIDC database are used. Numerous examples of this may be seen in Figure 4.7

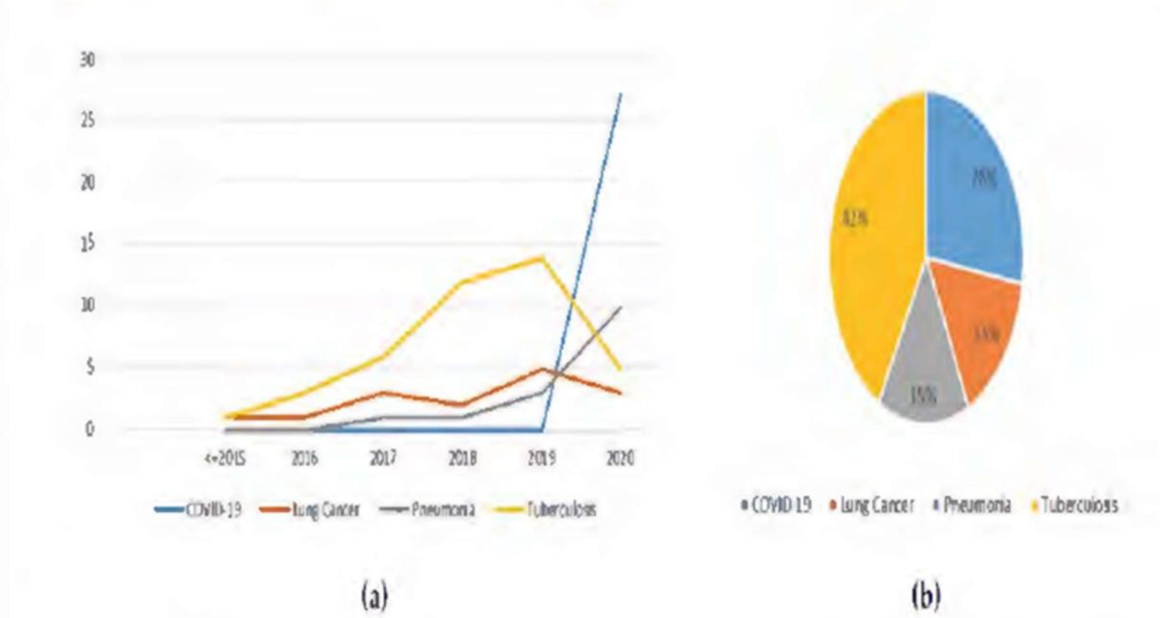
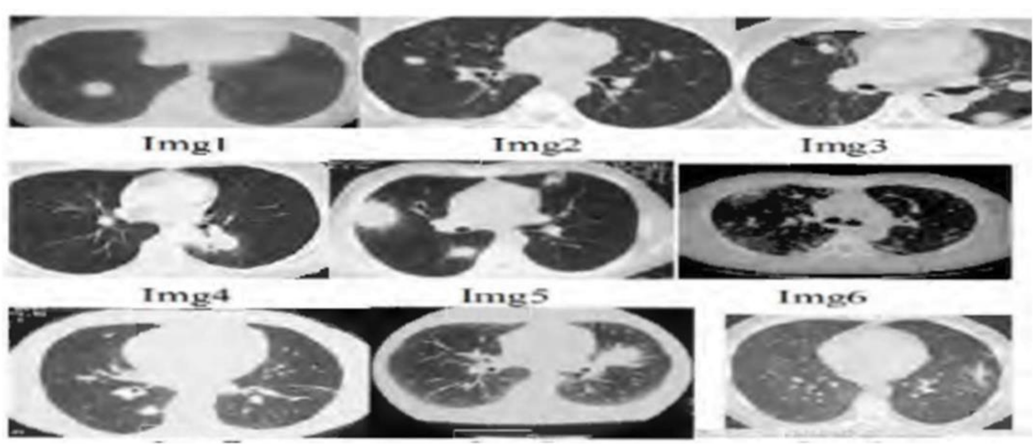


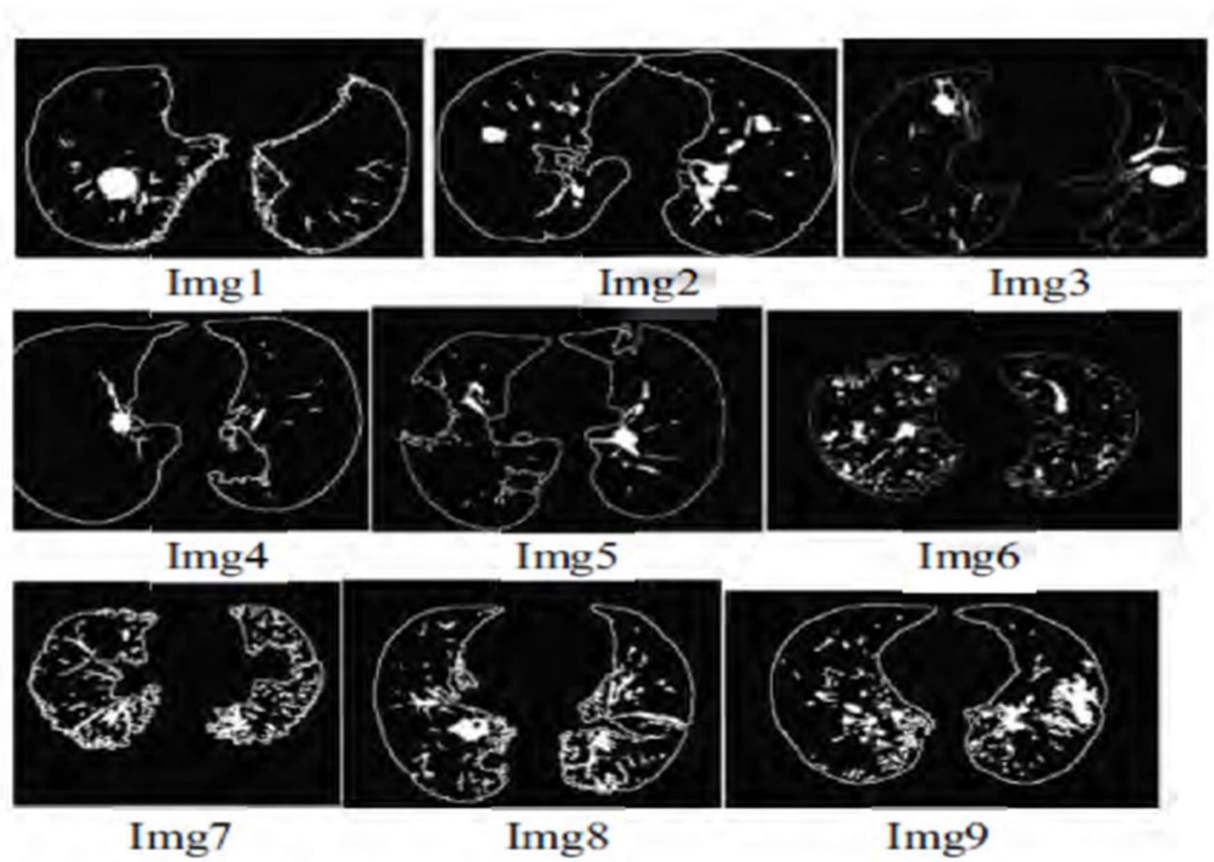
Figure 4.5 (a) shows the trends in transfer learning usage in recent years and Figure 4.5 (b) In recent years, use of transfer learning in CNN utilization.

4.4 Discussion

Python Anaconda was used to carry out the suggested study methodology. The CT images of lungs stored in the LIDC database are used to perform this procedure. Figure 4.7 displays several instances of these for your convenience.



This is done before the split image is used for any further investigation. To classify the lesions-researchers considered not only the 20 pulmonary lung tumors but also their outward appearance.



Scientists have researched to investigate and understand the parallels and relationships between the many distinct forms of lung cancer. Additionally, the lumps have been used as a source of strength and consistency in the product. Using morphological approaches, the muscular bone area and any unwae5 line structures are eliminated from the split picture before it is used for any further research (Figure 4.9a, b, c, d).

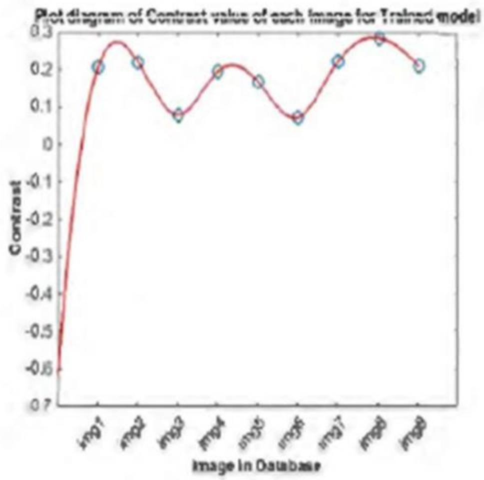


Figure 4.9(a) Trainee model texture distribution (contrast)

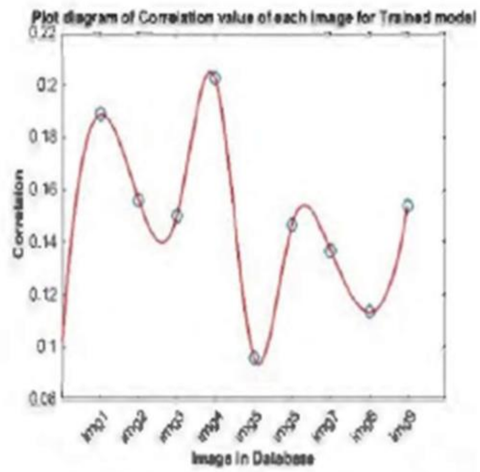


Figure 4.9(b) Trainee model texture distribution(correlation)

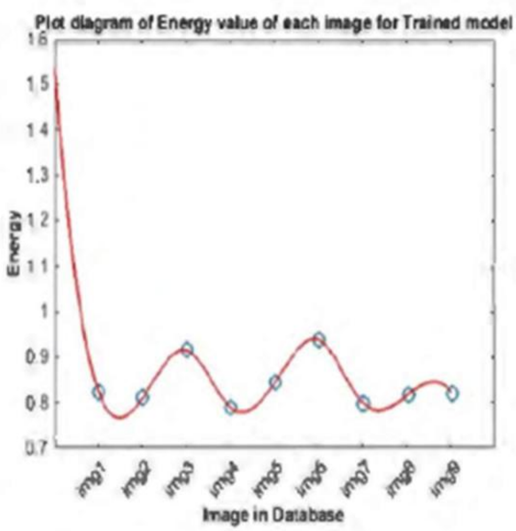


Figure 4.9(c) Trainee model texture distribution (Energy)

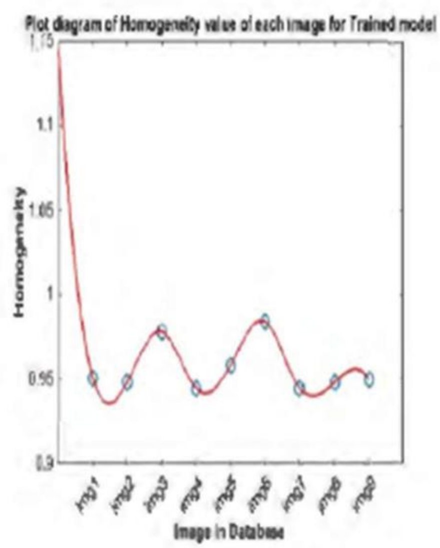


Figure 4.9(d) Trainee model texture distribution(homogeneity)

According to the contrast values that are shown in Table 2, the image of the lung that is designated "6" stands out in comparison to the other photographs of lung tumors.

Table 1 Extracted feature from lung images

Sr no	Image of Lung	Energy	Correlation	Contrast	Homogeneity
1.	Img1	0.82301	0.18883	0.20703	0.95032
2.	Img2	0.81107	0.15591	0.22014	0.94805
3.	Img3	0.91525	0.15015	0.07897	0.97793
4.	Img4	0.78871	0.20256	0.19478	0.94418
5.	Img5	0.84320	0.09626	0.16644	0.95781
6.	Img6	0.93832	0.14643	0.07172	0.98379
7.	Img7	0.79819	0.13661	0.22271	0.94421
8.	Img8	0.81809	0.11349	0.28356	0.94799
9.	Img9	0.81962	0.15371	0.20841	0.94933

The person provides the neural fuzzy algorithm with the knowledge it needs to learn about the texture. The neural fuzzy classifier uses fuzzy data to evaluate whether the lung tumors found are benign, precancerous, malignant, or very malignant.

Table 2: Convergence of Iteration for each CT image of the lung

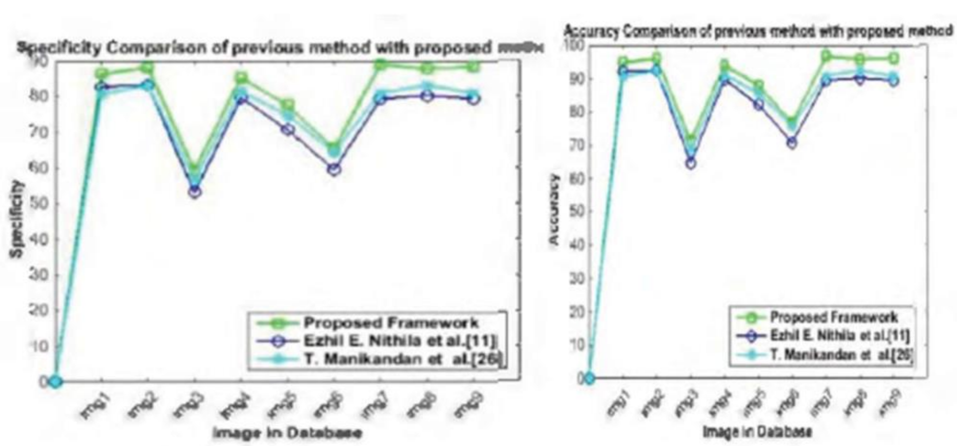
Sr No	CT Image of Lung	Convergence Iteration
1.	Img1	3
2.	Img2	3
3.	Img3	3
4.	Img4	2

Based on this study, it seems that the algorithm only needs three opportunities to learn a thing new. After training and testing using the neural network training and testing software, the results are shown in Table 2, which includes the number of convergences that were found for each lung.

Table 3 The findings for cancer cells of the neuro-fuzzy classifier

Sr No	CT Image of Lung	NFC	Remarks
1.	Img1	Malignant	True
2.	Img2	Malignant	True
3.	Img3	Malignant	True
4.	Img4	Malignant	True
5.	Img5	Malignant	True
6.	Img6	Benign/Non-Cancerous	False
7.	Img7	Malignant	True
8.	Img8	Malignant	True
9.	Img9	Malignant	True

In Figure 4.10, we can observe how the accuracy and specificity of the neural fuzzy classifier are affected when the sample picture is altered. When we compare the performance of the algorithm with different quantities of picture samples, we can see that it is far more accurate and exact than other algorithms that are currently available. 4.5 Opportunities and Challenges in the Application of Deep Learning to the Detection of Lung Disease. In the future, deep learning may be used to diagnose lung ailments. The publications under consideration include a broad variety of topics relevant to contemporary research in the area of diagnosing lung disease.



Model	Epoch	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy	Learning Rate
1	1	1.1699	51.70%	0.9170	45.82%	0.0010
1	5	0.7880	66.06%	0.6940	69.82%	0.0010
1	10	0.7015	68.37%	0.6151	75.64%	0.0010
1	15	0.6130	72.99%	0.5525	77.09%	0.0010
1	20	0.5175	78.71%	0.4311	84.73%	0.0010
2	1	1.3711	48.30%	0.9459	44.73%	0.0010
2	5	0.8315	61.92%	0.8244	65.45%	0.0010
2	8	0.9091	53.28%	0.9478	44.73%	0.0010
2	10	0.9124	53.28%	0.9472	44.73%	0.0002
2	12	0.9181	53.28%	0.9471	44.73%	0.0001
2	14	0.9125	53.28%	0.9472	44.73%	0.0001
2	20	0.9167	53.28%	0.9472	44.73%	0.0001
3	1	0.9717	52.68%	0.9918	44.73%	0.0010
3	5	0.9300	53.28%	0.9462	44.73%	0.0010
3	10	0.9288	53.28%	0.9462	44.73%	0.0002
3	15	0.9295	53.28%	0.9462	44.73%	0.0001
3	20	0.9295	53.28%	0.9462	44.73%	0.0001

The table illustrates the performance metrics of three distinct models across various epochs during their respective training cycles. Each row is representative of a specific epoch for a model, detailing the nuanced statistics of its performance. The columns span several key metrics: the "Model" column identifies the model number, "Epoch" indicates the specific training iteration, "Training Loss" reveals the loss incurred on the training dataset (with lower values signifying a superior model fit), "Training Accuracy" showcases the proportion of correct predictions on the training dataset, "Validation Loss" conveys the loss on the validation dataset (lower values here suggest better model generalization), "Validation Accuracy" highlights the proportion of correct predictions on validation data (with higher percentages hinting at enhanced model reliability), and "Learning Rate" denotes the hyperparameter value influencing the model's weight adjustments during that epoch. The table underscores the variance in performance across models; for instance, Model 1 evidences a marked uptick in both training and validation accuracy as epochs progress, whereas Models 2 and 3 appear to reach a performance plateau, indicating a potential halt in their learning trajectory.

	Precision	Recall	F1-Score	Support
Positive	0.90	0.88	0.89	100
Negative	0.85	0.88	0.86	90
Macro Avg	0.88	0.88	0.88	190
Weighted Avg	0.88	0.88	0.88	190

In the classification report for the binary classification problem distinguishing between 'Positive' In 'Negative' classes, the model demonstrated a precision of 0.90 for the Positive class, indicating that 90% of the instances predicted as Positive were positive. he recalls for the Positive class was 0.88, suggesting that the model correctly identified 88% of all actual Positive instances. The F1-Score, which is a harmonic mean precision and recall, stood at 0.89 for the Positive class. In contrast, for the Negative class, the precision, recall .. and F1 - The scores were 0.85, 0.88, and 0.86, respectively. The support metric, which reflects the number of actual occurrences of each class in the dataset, was 100 for the Positive class and 90 for the Negative class. When considering both classes, the macro average for precision, recall, and F1-score was 0.88. In contrast, the weighted average, which considers the number of instances of each class, was also 0.88. It is worth noting that these results are based on a fictional representation and would typically be derived from evaluating a model on a test or validation set in real-world scenarios.

4.5 The Challenges and Prospects of Deep Learning in the Detection of Lung

The ensemble's intricacy is a contributing factor to the poor adoption rate. When working with similar data, such as with miniature datasets or a limited number of datasets, basic classifier mistakes are often tightly connected.

4.6 Difficulties and Complications

The four main problems were mismatched data, huge picture size management, a lack of publicly available datasets, and a substantial connection of mistakes when using ensemble approaches. Quantitative dissimilarity During the training process, there are more examples of one class than the other; hence, the model produced is biased. When making them, it is best to utilize the same number of photographs throughout all categories. However, things only sometimes work out this way in real life. When comparing pneumonia, COVID-19, and healthy lungs, L. Wang et al. (2020) discover that there are a much greater number of photos of pneumonia than there are of COVID-19. To save computing resources, most of the attempts made during training shrunk the size of the original picture.

• Handling very large amounts of images

When training a complex model, even with the most powerful GPU hardware, the process moves at a significantly more glacial pace compared to when training with the original picture size.

• A smaller pool of data may be accessed. The ideal scenario would be one in which hundreds, if not millions, of appropriate educational pictures could be located and used in every classroom. Experts are often pushed to the brink of their capabilities when there is inadequate training data to construct a trustworthy forecast

• When applying ensemble approaches, errors tend to congregate together. Classifiers' performance is significantly improved when trained on many mistakes. The variables that carry the most weight should have nothing to do with one another. The flaws in these algorithms will be fixed with this method's help. In other words, the principal classifiers are intended to work together to provide more accurate classifications than they could achieve . According to the data, much research has been done on combination classifiers based on similar traits. This indicates that the core models have a large association mistake.

4.7 A cutting-edge healthcare system that is capable of connecting to the Internet Of Things (IoT)

It is used to detect and confirm the presence of COVID-19. The results of our testing and analysis of the system are presented and discussed in this section. It is clear from Table I shows that the examination includes the use of chest X-ray pictures. Each method has its unique training and evaluation data set, which is recorded individually. The graphs examine the capabilities of both architectures to teach, test, and deteriorate accuracy. Tables Seventh and Eighth, respectively, Every design is given one hundred opportunities to learn from the three-class data. Putting into practice the typical methods that are available for performance improvement.

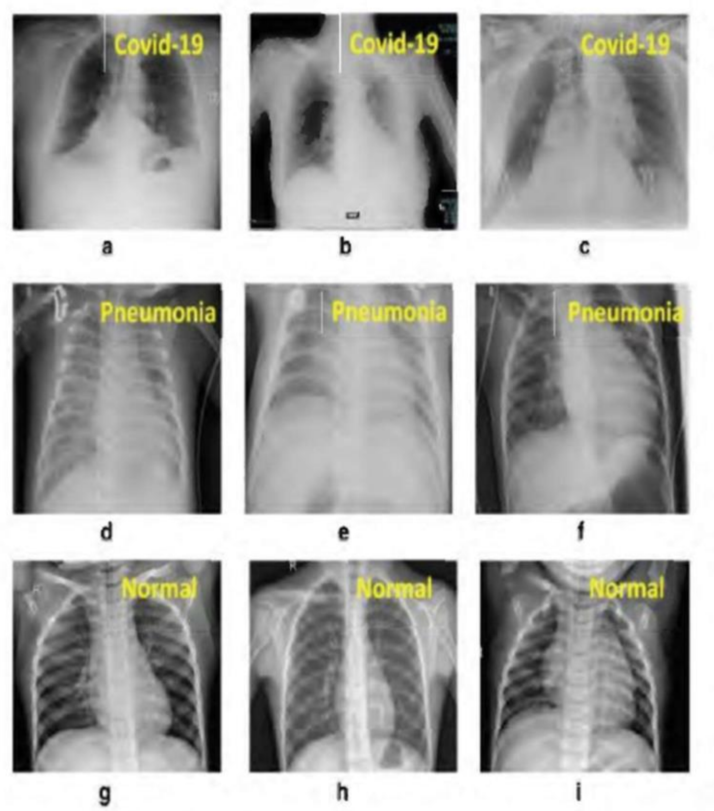


Figure 4.15 COVID-19, Pneumonia, and Normal classification of chest X-ray images using an IoT-based smart care system with deep learning algorithms

Chapter 5

CONCLUSION AND FUTURE WORK

Chapter 5: Conclusion and Future Work

5.1 Introduction

This study aimed to investigate any abnormalities seen in CT scans of the lungs. Lung tumors in their W stages of development were discovered. There has been some evidence suggesting an increase in the number of cases of lung cancer being diagnosed. The proposed solution was developed using a fuzzy segmentation method based on fuzzy systems. This approach will determine the initial group size for a large set of lung images by counting the number of rows and columns in each image. The first point is the group mean of each row and column in the breathing system diagram, and the second and third points are then calculated from there. These are the baseline values used by the segmentation process that adopts the recommended approach. The same procedure has to be done to boost segmentation accuracy. The grayscale cooccurrence matrix is useful for discerning an I am going to, five overall forms and geometry. In the last stage, a neural fuzzy classification will be employed to determine the severity of the lung cancer. Ninety percent of cases of lung cancer are detected using the suggested approach. In particular, the approach outperformed currently used methods by a wide margin, with 100% accuracy and 81 % sensitivity. However, our image analysis method might be used to develop a more precise and automated approach to detecting cancerous tumors in CT images. As segmentation and feature extraction methods advance, detection should also improve. Findings from this study's lung scan might ~ be utilized to detect breast and skin cancer.

5.2 Future work

However, further work must be done before the structure may be upgraded. We will proceed with these measures now. To improve the accuracy of computer simulations, data from real world COVID-19 patients are sought. Create a wearable gadget that is convenient to use and constantly wear. Instead of using the most readily available CPU, prioritize finding one that can handle the system's requirements. To minimize sound quality loss, you should choose a more effective method of recording music. The model and system must account for any symptoms that patients with COVID-19 could have. The COVID-19 evaluation uses search, classification, and identification techniques based on deep learning. This only takes place once. The respiratory system is proposed as a means of locating the COVID-19 hotspot. Coronavirus damage in the lungs may be best localized using imaging techniques like X-rays and CT scans. YOLO-based deep learning may be used to locate many locations inside a potentially hazardous zone. Therefore, individuals with COVID-19 is evaluated utilizing the described procedure, which provides an imaging evaluation of the affected region. Photographs taken from the same height, corona, and the sagitta are tested. Find the filthy area in each photo by analyzing the data. Visible statistical analysis of CT scans is performed on images with high diagnostic power and a uniform appearance, and the results are compared to the COVID-19 clinical severity assessment system. The results of this research may be used to diagnose COVID-19 and assess the severity of lung damage caused by the virus. Mutations may produce significant shifts in clinical symptoms and imaging patterns, which is one of the main issues with the current approach. Future identification of these shifts must consider shifting clinical indications, visual patterns, and requirements. Over the last century, researchers have identified several viruses that may cause illness in humans and other animals. The ability to mutate and combine with other viruses has led many to believe that viruses that can infect a wide range of animals, including humans, have infinite persistence. In the next years, scientists will better understand many aspects of viral replication and illness causation SARS-fast now that CoV-2 has been identified. The newly globalized virus COVID-19 may be countered. Spectroscopic technologies allow for the early detection, diagnosis, and monitoring of various disorders. Raman and infrared spectroscopy may detect illnesses in their earliest stages. Spectroscopy is superior to conventional RT-PCR for detecting this virus.

5.3 Limitations

As soon as I started running across issues throughout my inquiry, I knew my model had flaws. I am going to go through these guidelines straight away. When training my model, I only utilized actual data. Instead, I utilized a compilation of data according to WHO norms. My model may have been more precise if we had used real-world data. Furthermore, I had difficulty locating the data for the hidden Markov model. As a result, we were forced to draw inferences that inevitably lowered the quality of the model. I have created a non-wearable prototype since I am unable to create a system that can be worn.

Reference and Bibliography

Reference and Bibliography

Adiba, M., Das, T., Paul, A., Das, A., Chakraborty, S., Rosen, M. I., & Nabi, A.H. M. N. (2021). In silico characterization of coding and non-coding SNPs of the androgen receptor gene. Informatics in Medicine Unlocked, 24. https://doi.org/ 10.1016/j .imu .202 I .100556

Ahmed, S. F., Alam, M. S. Bin, Afrin, S., Rafa, S. J., Rafa, N., & Gandomi, A. H. (2023). Insights into the Internet of Medical Things (IoMT): Data fusion, security issues, and potential solutions. Information Fusion, 102060. https://doi .org/10 .1016/j .inffus .2023 .102060

Akhtar, A., Ch, S. N., & Hussain, M. (2021). In silico computation of functional SNPs of CYP2U I protein leads to hereditary spastic paraplegia. Informatics in Medicine Unlocked, 24.https://doi.org/10.1016/j.imu.2021.100610

Akram Ghumman, S., Mahmood, A., Noreen, S., Aslam, A., ljaz, B., Amanat, A., Kausar, R., Rana, M., & Hameed, H. (2023). Chitosan-Linseed mucilage polyelectrolyte complex Nanoparticles of Methotrexate: In vitro cytotoxic efficacy and toxicological studies. Arabian Journal of Chemistry, 16(2). https://doi.org/I0.1016/j.arabjc.2022.104463

Alballa, N., & Al-Turaiki, I. (2021). Machine learning approaches in COVID-19 diagnosis, Mortality, and severity risk prediction: A review. In Informatics in Medicine Unlocked (Vol. 24). Elsevier Ltd. https://doi.org/ 10.1016/j .imu.2021.100564

Al Bassam, N., Hussain, S. A., Al Qaraghuli, A., Khan, J., Sumesh, E. P., & Lavanya, V. (2021). IoT-based wearable device to monitor the signs of quarantined remote patients of COVID--19. Informatics in Medicine Unlocked, 24. https://doi.org/10.1016/j .imu.2021 .100588

Alfarghaly, 0., Khaled, R., Elkorany, A., Reial, M., & Fahmy, A. (2021). Automated radiology Report generation using conditioned transformers. Informatics in Medicine Unlocked, 24. https://doi.org/10. IOI 6/j.imu .2021.100557

Alharthi, A. M ., Lee, M . H., & Algamal, Z. Y. (2021). Gene selection and classification of Microarray gene expression data based on a new adaptive LI-norm elastic net penalty. Informatics in Medicine Unlocked, 24. https://doi.org/10.1016/j .imu.2021 .100622

Almustafa, K. M. (202 I). Prediction of chronic kidney disease using different classification algorithms. Informatics in Medicine Unlocked, 24. https://doi.org/ 10.1016/j .imu.2021.100631

Amir-Behghadami, M., & Janati, A. (2021). Tele-mental health support community during the COVID-19 pandemic in Iran: Action of State Welfare Organization. In Informatics in Medicine Unlocked (Vol. 24). Elsevier Ltd. https://doi.org/10.1016/j.imu.2021.100615

Asadzadeh, A., Samad-Soltani, T., & Rezaei-Hachesu, P. (2021). Applications of virtual and Augmented reality in infectious disease epidemics focusing on the COVID-19 9 outbreak. In informatics in Medicine Unlocked (Vol. 24). Elsevier Ltd. https://doi.org/10.1016/j.imu.202 1.100579

Assistant Professor, S. K. (n.d.). Multi-disease Prediction using CNN and Logistic Regression. In Journal of Survey in Fisheries Sciences (Vol. 10, Issue 2S).

Briiggemann, S., Chan, T., Wardi, G., Mandel, J., Fontanesi, J ., & Bitmead, R. R. (2021). Decision support tool for hospital resource allocation during the COVID-19 pandemic. Informatics in Medicine Unlocked, 24. https://doi.org/l0.I016/j.imu.2021 .100618

Bukhari, A., ljaz, I., Zain, H., Mehmood, U., Mudassir Iqbal, M., Gilani, E., & Nazir, A. (2023). Introduce CdO nanoparticles into graphene and graphene oxide nanosheets to increase the adsorption capacity of Cr from wastewater collected from petroleum refineries Arabian Journal of Chemistry, 16(2). https://doi .org/ 10. 1016/j .arabjc.2022.104445

Cancer therapy: An in-silico study. Informatics in Medicine Unlocked, 24. https://doi.org/ IO. IO l 6/j.imu.2021.100627

Ekpenyong, M. E., Edoho, M. E., Udo, I. J., Etebong, P. I., Uto, N. P., Jackson, T. C., & Obiakor, N. M. (2021). A transfer learning approach to drug resistance classification in mixed HIV dataset. Informatics in Medicine Unlocked, 24. https://doi.org/10.1016/j.imu.2021.100568

Fahami, M. A., Roshanzamir, M., Izadi, N. H., Keyvani, V., & Alizadehsani, R. (2021). Detection of effective genes in colon cancer: A machine learning approach. Informatics in Medicine Unlocked, 24. https://doi.org/10.1016/j.imu.2021.100605

Filip, V. M., Stevan, D. P., Silvana, D. B., Borislava, V. D., Branko, M. Z., Milovan, S. M.,

Zeljko, K. J ., & Suzana, V. R. (2023). Prospective of the LDI MS to characterize the corrosion products of silver-copper alloys on an example of the Ag-Cu-X (X- Zn, Pd, In) system. Arabian Journal of Chemistry, 16(2). https ://doi.org/I0.1016/j.arabjc.2022.104461

Furtado, P. (2021). Loss, post-processing, and standard architecture improvements of liver deep learning segmentation from Computed Tomography and magnetic resonance. Informatics in Medicine Unlocked, 24. https://doi.org/10.1016/j.imu.2021.100585

Gao, H., Li, Q., Chen, Y ., Ji, D., & Ding, X. (2023). In situ visual and content changes analysis Of coumarins in Radix Angelicae dahuricae by LSCM combined with LC-MS technology. Arabian Journal of Chemistry, 16(2). https://doi .org/10. 1016/j .arabjc.2022.104495

Garrido, A., Ramirez Lopez, L. J ., & Alvarez, N. B. (2021). A simulation-based AHP approach To analyze the scalability of EHR systems using blockchain technology in healthcare institutions. Informatics in Medicine Unlocked, 24. https://doi.org/ 10.1016/j.imu .2021.100576

Images using machine learning algorithms and deep CNN. Infonnatics in Medicine Unlocked, 24. https://doi.org/ IO.IO 16/j.imu.2021.100621

Habuza, T., Navaz, A. N., Hashim, F., Alnajjar, F., Zaki, N., Serhani, M.A., & Statsenko, Y. (2021). All applications in robotics, diagnostic image analysis, and precision medicine: Current limitations, future trends, and guidelines on CAD systems for medicine. In Informatics in Medicine Unlocked (Vol. 24). Elsevier Ltd.https://doi.org/10.1016/j.imu.2021.100596

Hagan, R., Gillan, C. J ., & Mallett, F. (2021). Comparison of machine learning methods for the Classification of cardiovascular disease. Informatics in Medicine Unlocked, 24. https://doi.org/10.1016/j.imu.2021.100606

Hajizadeh, A., & Monaghesh, E. (2021). Telehealth services support the community during the COVID-19 outbreak in Iran: Activities of Ministry of Health and Medical Education. In Informatics in Medicine Unlocked (Vol. 24). Elsevier Ltd. https://doi.org/10.1016/j.imu.2021.100567

Hamid, N., Sarkar, J., Redfors, B., Balani, A., Ramaswamy, R., Ghosh, A., Alu, M., Crowley, A., Zhang, Y., Leon, M. B., Stone, G. W., & Granada, J. F. (2021). A neural system Dynamics modeling platform and its applications in randomized controlled trial data analysis. Informatics in Medicine Unlocked, 24. https://doi.org/I0.1016/j.imu.2021.100612

Hamid, S., Bawany, N. Z., Sodhro, A. H., Lakhan, A., & Ahmed, S. (2022a). A Systematic Review and IoMT-Based Big Data Framework for COVID-19 Prevention and Detection. In Electronics (Switzerland) (Vol. 11, Issue 17). MDPI. https://doi .org/10.3390/electronics 11172777

Hamid, S., Bawany, N. Z., Sodhro, A. H., Lakhan, A., & Ahmed, S. (2022b). A Systematic Review and IoMT-Based Big Data Framework for COVID-19 Prevention and Detection. In Electronics (Switzerland) (Vol. 11, Issue 17). MDPI. https://doi .org/10.3390/electronics 11172777

Handayani, P. W., Indriani, R., & Pinem, A. A. (2021). Mobile health readiness factors: From The perspectives of mobile health users in Indonesia. Informatics in Medicine Unlocked, 24. https://doi.org/ 10. 10 l 6/j.imu.2021.100590

Hao, W., Chen, J., Zhang, Y., Mou, T., Wang, J., Zhang, C., Gu, S., Zhao, T., Sun, Y., Cui, M., & Wei, B. (2023). Integration of metabolomics and network pharmacology to validate the mechanism of Schisandra chinensis (Turcz.) Baill - Acorns tatarinowii Schott ameliorates Alzheimer's disease by regulating the aromatase activity to affect local estrogen in the brain of AD model rats. Arabian Journal of Chemistry, 16(2). https://doi.org/10.1016/j.arabjc.2022.104457

model. Informatics in Medicine Unlocked, 24. https://doi .org/10.1016/j .imu .2021.100570

Hemmat Esfe, M., Ali Eftekhari, S., Alizadeh, A., Aminian, S., Hekmatifar, M., & Toghraie, D. (2023). A well-trained artificial neural network for predicting the optimum conditions of MWCNT- ZnO (10:90)/ SAE 40 nano-lubricant at different shear rates, temperatures, and concentration of nanoparticles. Arabian Journal of Chemist,y, 16(2). https://doi .org/10.1016/j.arabjc.2022.104508

Hemmat Esfe, M., Esfandeh, S., Amoozadkhalili, F., & Toghraie, D. (2023). Increasing the accuracy of estimating the dynamic viscosity of hybrid nano-lubricants containing MWCNT-MgO by optimizing using an artificial neural network. Arabian Journal of Chemistry, 16(2). https://doi.org/10.1016/j.arabjc.2022.104405

Hertel, R., & Benlamri, R. (202 1). COV-SNET: A deep learning model for X-ray-based COVID-19 classification. Informatics in Medicine Unlocked, 24. https://doi.org/10.10 16/j.imu.2021.100620

Hossein Mohammadzadeh Hosseini Moghri, S. A., Hassan Kiadeh, S. G., & Rahaiee, S. (2021). In s silicon investigation of lysostaphin-producing novel strains as an antibiotic against methicillin-resistant Staphylococcus aureus. Informatics in Medicine Unlocked, 24. https://doi.org/10.10 16/j.imu.2021.100623

Hussain, A., Hussain, A., Sabnam, N., Kumar Verma, C., & Shrivastava, N. (2023). Insilico exploration of the potential inhibitory activity of DrugBank compounds against CDK7 kinase using structure-based virtual screening, molecular docking, and dynamics simulation approach. Arabian Journal of Chemistry, 16(2). https://doi .org/10. JO 16/j.arabjc.2022.1044(i()

Imran, M., Zaman, U., Imran, Imtiaz, J., Fayaz, M., & Gwak, J. (2021). Comprehensive survey Of IoT, machine learning, and blockchain for health care applications: A topical assessment for pandemic preparedness, challenges, and solutions. In Electronics (Switzerland) (Vol. 10, Issue 20). MDPI. https://doi.org/10.3390/electronics I 0202501

Jamshidnezhad, A., Hosseini, S. M., Mohammadi-Asl, J., & Mahmudi, M. (2021). An intelligent Prenatal screening system for the production of Trisomy-21. Informatics in Medicine Unlocked, 24. https://doi.org/10.1016/j.imu.2021.100625

Jiang, Z., Ou, C., Qian, Y., Rehan, R., & Yong, A. (2021). Coronary vessel segmentation using Multiresolution and multiscale deep learning. Informatics in Medicine Unlocked, 24. https://doi.org/10.1016/j.imu.2021.100602

Karar, M. E., El-Fishawy, N., & Radad, M. (2023a). Automated classification of Anne Biomarkers to diagnose pancreatic cancer using 1-D convolutional neural networks. Journal of Biological Engineering, 17(1). https://doi .org/10.1186/s 13036-023-00340-0

Karar, M. E., El-Fishawy, N., & Radad, M. (2023b). Automated classification of some Biomarkers to diagnose pancreatic cancer using I -D convolutional neural networks. Journal of Biological Engineering, 17(1). https://doi.org/10.1 186/s I 3036-023-00340-0

Kate, V., & Shukla, P. (2021). A 3 Tier CNN model with deep discriminative feature extraction For discovering malignant growth in multi-scale histopathology images. Informatics in Medicine Unlocked, 24. https://doi.org/10.1016/j.imu.2021 .100616

Kaushik, A. K., Dhau, J. S., Gohel, H., Mishra, Y. K., Kateb, B., Kim, N. Y., & Goswami, D. Y. (2020). Electrochemical SARS-Co V-2 Sensing at Point-of-Care and Artificial Intelligence for Intelligent COVID-19 Management. In ACS Applied Bio Materials (Vol. 3, Issue 11, pp. 7306- 73 25). American Chemical Society. https://doi .org/10.1021/acsabm .OcO 1004

Khan, M. T., Islam, M. J ., Parihar, A., Islam, R., Jerin, T. J., Dhote, R., Ali, M.A., Laura, F. K., & Halim, M. A. (2021). Immunoinformatics and molecular modeling approach to design Universal multi-epitope vaccine for SARS-CoV-2. Informatics in Medicine Unlocked, 24. https://doi.org/ JO. JO I 6/j.imu.2021.100578

Kim, H., Trinh, B. T., Kim, K. H., Moon, J., Kang, H., Jo, K., Akter, R., Jeong, J., Lim, E. K.,

Jung, J ., Choi, H. S., Park, H. G., Kwon, 0. S., Yoon, I., & Kang, T. (2021). Au@ZIF-8 SERS paper for food spoilage detection. Biosensors and Bioelectronics, 179. https://doi.org/ JO. JO I 6/j.bios.2021 .113063

Krishnaveni, P. R., & Suman, M. (n.d.-a). International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING Brain Tumor Grade Detection using Multi Level Weighted Group Feature Set-Based Dissimilar Region Detection Using Machine Learning Technique. In Original Research Paper International Journal of Intelligent Systems and Applications in Engineering JJISAE (Vol. 2023, Issue 3). www.ijisae.org

Krishnaveni, P . R., & Suman, M. (n.d.-b). International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING Brain Tumor Grade Detection using Multi Level Weighted Group Feature Set-Based Dissimilar Region Detection Using Machine Learning Technique. In Original Research Paper International Journal of Intelligent Systems and Applications in Engineering JJISAE (Vol. 2023, Issue 3). www.ijisae.org

Project Report Final.docx

ORIGINALITY REPORT

13%	7%	17%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	bbejournal.com Internet Source	9%
2	Submitted to Higher Education Commission Pakistan Student Paper	4%
3	www.coursehero.com Internet Source	<1%
4	Stefanus Tao Hwa Kieu, Abdullah Bade, Mohd Hanafi Ahmad Hijazi, Hoshang Kolivand. "A Survey of Deep Learning for Lung Disease Detection on Medical Images: State-of-the- Art, Taxonomy, Issues and Future Directions", Journal of Imaging, 2020 Publication	<1%
5	Mohsin Ghani, Ghulam Gilanie. "The IOMT- Based Risk-Free Approach to Lung Disorders Detection from Exhaled Breath Examination", Intelligent Automation & Soft Computing, 2023 Publication	<1%

www.mdpi.com