زانكۆی سەلاحەدین - هەولێر
Salahaddin University-Erbil

A REVIEW ON LUNG CANCER DETECTION TECHNIQUES USING MACHINE LEARNING

Research Project

This research project has been written under my supervision and has been submitted for the award of the degree of B.Sc. in **(PHYSICS**).

BY:

Hanan Askandar jawhar

SUPERVISED BY:

Assistant Prof. Dr. Sardar P. Yaba

April

2024

بِسْمِ اللهِ الرَّحْمنِ الرَّحِيمِ

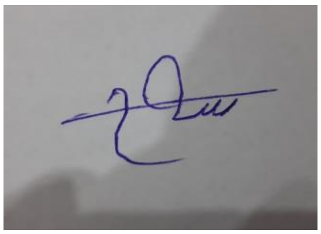
قَالُواْ سُبْحَانَكَ لاَ عِلْمَ لَنَا إِلاَّ مَا عَلَّمْتَنَا إِنَّكَ أَنتَ الْعَلِيمُ الْحَكِيمُ

صدق الله العظيم

سورة البقرة الاية32

Supervisor Certificate

This research project has been written under my supervision and has been submitted for the award of the degree of B.Sc. in (Physics)

Signature: 

Name: Assistant Prof. Dr. Sardar P. Yaba

Date: /   /2024

I confirm that all requirements have been completed.

Signature:

Name:

Head of the Department of physics

Date: /   /2024

This thesis is

dedicated:

- ✓ To My Parents
- ✓ To My Brother and Sisters
- ✓ To great teacher

Hanan

ACKNOWLEDGEMENTS

First and foremost, I must acknowledge my limitless thanks to **Allah**, the Ever-Magnificent; the Ever-Thankful, for His help and bless. I owe a deep debt of gratitude to the physics department and Salahaddin University for giving us an opportunity to complete this work. I am very appreciative to my colleagues in the physics department. I am grateful to some people, who worked hard with me from the beginning till the completion of the present research particularly my supervisor **Assistant Prof. Dr. Sardar P. Yaba** who has been always generous during all phases of the research. I would like to take this opportunity to say warm thanks to all my beloved friends, who have been so supportive along the way of doing my thesis. Last but not least, deepest thanks go to all people who took part in making this thesis real. Finally, my deepest thanks go to **Diyar Mustafa** for his support and encouragement.

SUMMARY

Lung cancer is a sort of dangerous cancer and is difficult to detect. It usually causes death for both genders men and women, therefore, so it is more necessary for care to immediately and correctly examine nodules. Accordingly, several techniques have been implemented to detect lung cancer in the early stages. This project aims to demonstrate the best methods used for categorizing and identifying lung cancer in its early phases by deep learning methods. In this project, a comparative analysis of different techniques based on machine learning for the detection of lung cancer has been presented. We have reviewed journal papers from 2018 to 2022 about this topic, with different existing methods covering machine learning (ML), deep learning, and artificial intelligence (AI), for the detection of lung cancer. The results show that Subrahamanian R. R. et al., achieved the highest performance ratio of 99.52%. Moreover, Jena S.R. et al., achieved the lowest performance ratio was %87.79.

TABLE OF CONTENTS

CHAPTER ONE

INTRODUCTION AND THEORETICAL BACKGROUND

1.1 Introduction

Lung cancer is the most important cause of death in humans. Since the symptoms of lung cancer occur in advanced stages, it leads to a high mortality rate among other types of cancer. Early diagnosis is therefore very important. Most of the lung diagnosis methods are costly and laborious. Lung cancer is a cost-effective disease that requires early diagnosis (Brunetti et al., 2022). The American Cancer Society estimates the lung cancer in the United States, in 2023, there will be 238,340 new cases of lung cancer and about 127,070 deaths from lung cancer (Siegel et al., 2018). In addition to examining the patient, blood tests, X-rays, biopsies, and computed tomography (CT) scans can be used to diagnose cancer. The sooner a patient is diagnosed, the better his or her prospects of healing and survival. In order to effectively diagnose the condition, technology is crucial. Based on their observations, several researchers have come up with solutions. In the latest years, a number of computer-aided diagnostic (CAD) procedures and systems have been proposed, implemented, and produced in order to use digital transmission to tackle this issue. Such algorithms employ a variety of machine learning and deep learning approaches, along with multiple methodologies based on image processing-based approaches for predicting cancer malignant levels. The aim of this project is to demonstrate the best methods used for categorizing and identifying lung cancer in its early phases by deep learning methods.

1.2 Theoretical background

1.2.1 The lungs

The lungs are the primary organs of the respiratory system in humans and most other animals, including some snails and a small number of fish. lungs are part of the respiratory system, a group of organs and tissues that work together to breathe. The respiratory system's main job is to transport oxygen and remove extra carbon dioxide. The lungs are a pair of spongy, air-filled organs located on either side of the chest (thorax). The trachea (windpipe) conducts inhaled air into the lungs through its tubular branches, called bronchi. The bronchi then divide into smaller and smaller branches (bronchioles), finally becoming microscopic. The bronchioles eventually end in clusters of microscopic air sacs called alveoli. In the alveoli, oxygen from the air is absorbed into the blood. Carbon dioxide, a waste product of metabolism, travels from the blood to the alveoli, where it can be exhaled. Between the alveoli is a thin layer of cells called the interstitial, which contains blood vessels and cells that help support the alveoli. The lungs are covered by a thin tissue layer called the pleura. The same kind of thin tissue lines the inside of the chest cavity -- also called pleura. A thin layer of fluid acts as a lubricant allowing the lungs to slip smoothly as they expand and contract with each breath as shown in Figure (1-1), (Hansen, 2012).

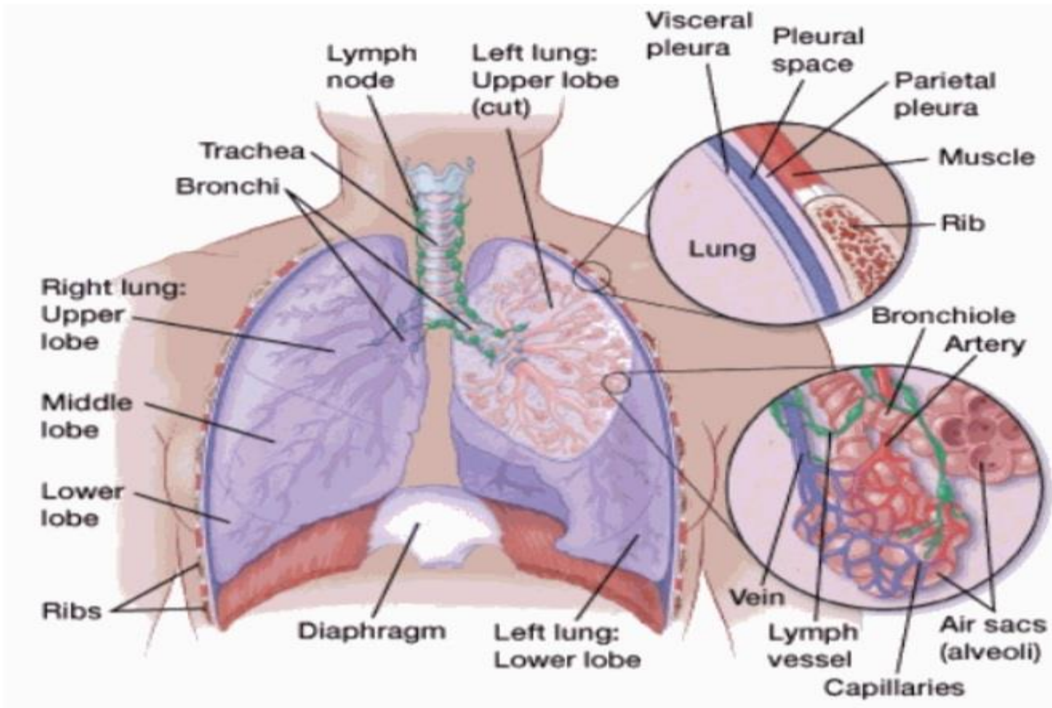


Figure (1-1): Normal Structure lung cancer (Sociaty, 2024).

1.2.2 The lungs cancer

Cancer of the lung, like all cancers, results from an abnormality in the body's basic unit of life, the cell. Normally, the body maintains a system of checks and balances on cell growth so that cells divide to produce new cells only when new cells are needed. Disruption of this system of checks and balances on cell growth results in an uncontrolled division and proliferation of cells that eventually form a mass known as a tumor. Since lung cancer tends to spread or metastasize very early after it forms, it is a very life-threatening cancer and one of the most difficult cancers to treat. While lung cancer can spread to any organ in the body, certain organs particularly the adrenal glands, liver, brain, and bone are the most common sites for lung cancer metastasis.

Lung cancer is cancer that begins in the cells of the lungs. It’s not the same as cancer that starts elsewhere and spreads to the lungs. Initially, the main symptoms involve the respiratory system. In the later stages of lung cancer, especially if it spreads to distant areas, it can affect many systems in your body. Lung cancer can affect more than just your lungs. Once you have a tumor in your lung, cancer cells can break off and form new tumors nearby or if wayward cancer cells enter the lymphatic system or bloodstream, they can travel to other parts of the body. There are two main types of lung cancer:

1- Non-small cell lung cancer (NSCLC):

About 80% to 85% of lung cancers are NSCLC. The main subtypes of NSCLC are adenocarcinoma, squamous cell carcinoma, and large cell carcinoma. These subtypes, which start from different types of lung cells, are grouped together as NSCLC because their treatment and prognoses (outlook) are often similar.

2- Small cell lung cancer (SCLC):

This type of lung cancer tends to grow and spread faster than NSCLC. In most people with SCLC, the cancer has already spread beyond the lungs at the time it is diagnosed. Since this cancer grows quickly, it tends to respond well to chemotherapy and radiation therapy. Unfortunately, for most people the cancer will return at some point (Hansen, 2012).

1.2.3 Machine learning

Artificial intelligence (AI)-based automated CT lung cancer detection is a feasible option that can assist physicians by reducing their workload, improving hospital operational throughput, and providing them with more time to develop high-quality doctor-patient relationships. Numerous studies have indicated that combining a physician/radiologist assessment with the use of an AI or machine learning model improves the detection of lung nodules on CT scans. AI models improved radiologist performance in detecting small lung nodules (less than 5 mm in diameter), which are commonly missed by eye assessment alone. Deep learning models minimize the workload on physicians, reduce fatigue- related errors of judgment, and improve the detection of nodules, which are more likely to be missed in the early stages of lung cancer. Machine learning is a field of artificial intelligence that uses algorithms to enable computers to learn from data, without being explicitly programmed. Machine learning is a subfield of artificial intelligence that focuses on developing algorithms that allow computers to learn from data, without being explicitly programmed. The field has grown rapidly in recent years and has applications in a variety of fields, including computer vision, natural language processing, speech recognition, and robotics, among others. Some popular machine learning algorithms include:

Linear Regression is used for modeling the relationship between a dependent variable and one or more independent variables.

Logistic Regression is used for modeling the probability of a binary outcome.

Decision Trees - used for classification and regression problems.

Random Forest - an ensemble of decision trees that can handle more complex datasets.

Support Vector Machines - used for classification and regression problems by finding the best hyperplane that separates the data.

Neural Networks - a complex algorithm inspired by the structure of the human brain, used for a wide range of tasks, such as image and speech recognition.

K-Means Clustering - used for clustering data points based on their similarity.

Principal Component Analysis - used for dimensionality reduction and feature extraction (Smola and Vishwanathan, 2008).

1.3 Literature Review

Lung cancer is a potentially lethal illness. Cancer detection continues to be a challenge for medical professionals. The true cause of cancer and its complete treatment have still not been discovered. Cancer that is caught early enough can be treated. Image processing methods such as noise reduction, feature extraction, identification of damaged regions, and maybe a comparison with data on the medical history of lung cancer are used to locate portions of the lung that have been impacted by cancer. This research shows an accurate classification and prediction of lung cancer using technology that is enabled by machine learning and image processing.

Tekade and Rajeswari (2018) Study lung nodules by using Computer Tomography (CT) scan images for detect and classify the lung nodules and the malignancy level of that nodules. The CT scan images are segmented using U-Net Architecture Prediction from U-Net and 3D multipath VGG-like network is combined for final results. The lung nodules are classified and malignancy level is detected using this architecture with **95.60% of accuracy**.

Shakeel et al. (2019) **a**pply deep learning instantaneously trained neural network for predicting lung cancer. Eventually, the system is examined by the efficiency of the system using MATAB based simulation results. The system ensures that **98.42% of accuracy** with minimum classification error 0.038**.**

(Lakshmanaprabu et al., 2019) Used Computed Tomography (CT) scan to find the position of tumor and identify the level of cancer in the body. The CT scan of lung images was analyzed with the assistance of Optimal Deep Neural Network (ODNN) and Linear Discriminate Analysis (LDA). The deep features extracted from a CT lung images and then dimensionality of feature is reduced using LDR to classify lung nodules as either malignant or benign. The ODNN is applied to CT images and then, optimized using Modified Gravitational Search Algorithm (MGSA) for identify the lung cancer classification The comparative results show that the proposed classifier gives the sensitivity of 96.2%, specificity of 94.2% and **accuracy of 94.56%.**

(Subramanian et al., 2020) proposed model is trained by using the Convolutional Neural Network (CNN). Pretrained ImageNet models including LeNet, AlexNet and VGG-16, are used to detect lung cancer. The proposed model uses AlexNet model and the features obtained from the last fully connected layer of the model were separately applied as input to the softmax classifier. The combination of AlexNet and softmax layer together has given an **accuracy of 99.52%.**

(Jena et al., 2021) proposed model is a network with multiple layer latent variables. Here, at every layer, the variables follow Gaussian distributions. Therefore, DGMM forms a cluster of Gaussian distributions to offer a nonlinear model and to describe the image data more flexibly. To eliminate over-parameterized solutions, Gaussian-based dimensionality reduction by designing an overfitting model is used. The simulation was carried out in a MATLAB environment, to achieve an **accuracy of about 87.79%** during the 18thepoch for training and testing the image samples. The false-positive rate could also be determined through this investigation. The anticipated DGMM-RBCNN shows a better and better trade-off than the prevailing systems.

(Shafi et al., 2022) Proposed computer-aided design (CAD) model identifies the physiological and pathological changes in the soft tissues of the cross-section in lung cancer lesions. Then, Is tested and validated using the CT scans of both patients and control patients that are not shown in the training phase. The proposed deep learning-assisted SVM-based model yields **94% accuracy** for pulmonary nodule detection representing early-stage lung cancer. It is found superior to other existing methods including complex deep learning, simple machine learning, and the hybrid techniques used on lung CT images for nodule detection.

CHAPTER TWO
MATERIALS AND METHODS

2.1 Materials

Six studies were evaluated in this review article for their accuracy in classifying lung cancer, from 2018 to 2022 (Tekade and Rajeswari, 2018, Shakeel et al., 2019, Subramanian et al., 2020, Jena et al., 2021, Shafi et al., 2022). The objective of this research work is to process and analyze the images that are framed from the CT scan and generate results on whether the images contain cancer cells or not. These images are collected from various multispecialty hospitals and diagnostic centers.

2.2 Methods

The sequence of operations for detection of tumor in CT images of the liver consists of various steps like preprocessing, segmentation, feature extraction, and classification of the tumor, which are shown in the Figure (2.1) (Neelagiri Dayanand, 2018).

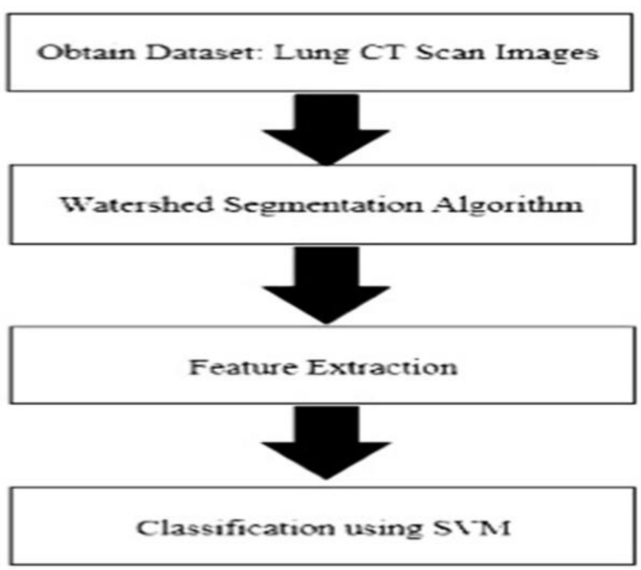


Figure (2-1). Basic Block Diagram for lung cancer detection.

CHAPTER THREE

RESULTS AND DISCUSSION

3.1 Result

According to the survey in the research for lung cancer detection techniques using machine learning. One observes various image processing techniques were used, like pre-processing, segmentation, calcification, and performance accuracy which is shown in Table (3-1).

Table (3-1) The comparison between various methods of detecting lung cancer by using the image processing technique.

Author	Data	Enhancement	Segmentation	Classification	Performance
(Tekade and Rajeswari, 2018)	**CT image**	**VGG-Like network**	**U-Net Architecture**	**LIDC LDRI**	**%95.6**
(Shakeel et al., 2019)	**CT image**	**(IPCT) Improved Profuse Clustering Technique and**	**K-mean Clustering**	**Deep learning Istantaneously trained neual network (DINN).**	**%98.42**
(Lakshmanaprabu et al., 2019)	**CT image**	**Filtering And Contrast**	**(MGSA) Modified Gravitational Search Algorithm.**	**(ODNN) Optimal Deep Neural Network.**	**%94.56**
(Subramanian et al., 2020)	**CT image**	**Adaptive bilateral filtering**	**U-Net Segmentation.**	**LeNet, AlexNet and VGG-16**	**%99.52**
(Jena et al., 2021).	**CT image**	**Collaborative Wiener Filter (CWF).**	**Multi-directional Patch-based Region growing Segmentation (MD-PRGS)**	**DGMM-RBCNN**	**%87.79**
(Shafi et al., 2022).	**CT image**	**ROI using multi-scale amplitude-modulation frequency modulation(AM-FM),and (PLSR).**	**U – Net**	**Deep learning Enable , SVM**	**%94**

3.2 Discussion

One of the most difficult challenges confronting today's researchers is cancer detection. Despite extensive study, there is still a lack of an accurate model since detection is a multidisciplinary role that depends on a variety of parameters. The advancement of accurate cancer detection techniques at an earlier stage has been the subject of extensive research. Among various methods discussed in this project, (Subramanian et al., 2020) achieved the highest performance ratio of 99.52% presented in Table (3-1) due to using of a deep learning model leveraging AlexNet pretrained model combined with a softmax layer developed and used to efficiently classify the lung CT images for cancer. Moreover, Jena S. R. et al., achieved the lowest performance ratio of %87.79 presented in Table (3-1), due to using the deep Gaussian mixture model in the region-based convolutional neural network [DGMM-RBCNN].

CHAPTER FOUR

CONCLUSION AND FUTURE WORK

4.1 Conclusion

1-When lung cancer is diagnosed at an early stage, it would be beneficial because the medication will then be initiated to prevent the disease from having a harmful result.

2-This project summarizes a survey of various machine-learning approaches to classify lung malignancies using CT scan images.

3-Many classifiers have been used by the researchers in the literature.

4-Based on the survey done in this work, it can be concluded that the highest result was about 99.52%.

4.2. Future Work

In the future, we prefer to identify the influential factors for lung cancer in the human body and use the same features of cancer prediction. An IoMT application can be developed to extract these physiological parameters. Physiological parameters for prediction include pressure, oxygen level, and body temperature.

REFERENCES

BRUNETTI, A., ALTINI, N., BUONGIORNO, D., GAROLLA, E., CORALLO, F., GRAVINA, M., BEVILACQUA, V. & PRENCIPE, B. 2022. A Machine Learning and Radiomics Approach in Lung Cancer for Predicting Histological Subtype. *Applied Sciences,* 12**,** 5829.

HANSEN, H. H. 2012. *Lung cancer: textbook for general practitioners*, Springer Science & Business Media.

JENA, S. R., GEORGE, S. T. & PONRAJ, D. N. 2021. Lung cancer detection and classification with DGMM-RBCNN technique. *Neural Computing and Applications,* 33**,** 15601-15617.

LAKSHMANAPRABU, S., MOHANTY, S. N., SHANKAR, K., ARUNKUMAR, N. & RAMIREZ, G. 2019. Optimal deep learning model for classification of lung cancer on CT images. *Future Generation Computer Systems,* 92**,** 374-382.

NEELAGIRI DAYANAND, S. 2018. *Lung cancer detection and classification using SVM.* California State University, Northridge.

SHAFI, I., DIN, S., KHAN, A., DÍEZ, I. D. L. T., CASANOVA, R. D. J. P., PIFARRE, K. T. & ASHRAF, I. 2022. An effective method for lung cancer diagnosis from ct scan using deep learning-based support vector network. *Cancers,* 14**,** 5457.

SHAKEEL, P. M., BURHANUDDIN, M. A. & DESA, M. I. 2019. Lung cancer detection from CT image using improved profuse clustering and deep learning instantaneously trained neural networks. *Measurement,* 145**,** 702-712.

SIEGEL, R. L., MILLER, K. D. & JEMAL, A. 2018. Cancer statistics, 2018. *CA: a cancer journal for clinicians,* 68**,** 7-30.

SMOLA, A. & VISHWANATHAN, S. 2008. Introduction to machine learning. *Cambridge University, UK,* 32**,** 2008.

SOCIATY, A. C. 2024. *Normal Structure lung cancer* [Online]. Available: https://www.cancer.org/cancer/types/lung-cancer/about/what-is.htm [Accessed 12/3/2024 2024].

SUBRAMANIAN, R. R., MOURYA, R. N., REDDY, V. P. T., REDDY, B. N. & AMARA, S. 2020. Lung cancer prediction using deep learning framework. *International Journal of Control and Automation,* 13**,** 154-160.

TEKADE, R. & RAJESWARI, K. Lung cancer detection and classification using deep learning. 2018 fourth international conference on computing communication control and automation (ICCUBEA), 2018. IEEE, 1-5.