Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606

Volume 06 Issue 01
2023

Review Article
https://doi.org/10.56979/601/2023

Machine Learning Approaches for Early Detection of Lung Cancer

Sayyid Kamran Hussain[1], Sadaqat Ali Ramay[1], Raheel Arshad[2], Munira Niazi[3], Zubair Mushtaq[4], Usama Ibrahim[5], Abdul Ahad Khan[5], Muhammad Bilal Afzal[5], and Ammar Aziz[5]

[1]Department of Computer Science, TIMES Institute Multan, Multan, 60000, Pakistan.
[2]Houston Community College, Houston Texas, United States, 010422, USA.
[3]Sir Syed Hospital and College of Medical Science Karachi, 74400, Karachi, Pakistan.
[4]Department Computer Science and Information Technology, Lund University Sweden, Lund, 22100, Sweden.
[5]Department of Computer Science & Information Technology, Thal University Bhakkar, 30000, Pakistan.
*Corresponding Author: Sayyid Kamran Hussain. Email: kamranshah.092@gmail.com.

Received: August 17, 2023 **Accepted:** November 29, 2023 **Published:** December 05, 2023

Abstract: Lung cancer is a very harmful form of cancer that is hard to find. It happens when cancerous cell grow in the airways. This type of cancer can be deadly. Doctors struggle to detect and treat it properly. They still don't know the exact cause or cure for cancer. If cancer is found early, it can be treated. It causes death in both men and women, so it's important to quickly and accurately examine any nodules. To detect lung cancer early, different techniques have been used. Finding cancer early can save lives and give patients a better chance of recovery. Technology plays a significant part in accurately identifying cancer. Researcher have proposed different approaches based on their studies. In recent years, machine learning algorithms have become powerful tools for helping detect lung diseases. Different algorithms have been used to study medical data and find patterns that show lung diseases. This paper reviews the utilization of ML methods for detecting lung cancer. Four methods—ANN (Artificial Neural Networks), SVM(Support Vector Machines), KNN(k-Nearest Neighbors), and Naive Bayes—were tested to see how well they detected lung cancer. KNN had the highest accuracy, with an average of 98.5%. The paper also looked at different datasets to find the most suitable one for detecting lung cancer. After reviewing multiple papers, the LIDC dataset was considered the best choice for this task. This review paper will assist researchers in efficiently reviewing relevant literature without having to refer to numerous papers.

Keywords: Lungs Classification; Machine learning; Diagnose; Respiratory system.

1. Introduction

In recent times, advancements in health and medicine have brought significant changes to modern well-being and disease prevention strategies.[1] One of the disease is cancer, a condition characterized by ab-normal tissue growth in the body.[2] Cancer comes in many different forms, with over a hundred identified types.[3] Specifically, lung cancer occurs when certain cells in the lungs undergo uncontrolled growth. The lungs have their own natural processes, such as coughing, which help clear out mucus, foreign objects, and irritants from the respiratory system.[4] Alveolar macrophages, specialized immune cells, also play a role in expelling harmful substances from the lungs, including germs and dust particles.[5] Additionally, mucociliary clearance, where tiny hair-like structures called cilia transport mucus and trapped particles out of the lungs, and pulmonary circulation, which removes carbon dioxide and supplies oxygen to the body, contribute to the self-cleansing ability of the lungs.[6] [7] Since the Lungs are vital for breathing and eliminating carbon di-oxide, early detection of cancerous cell within the lungs is crucial.[8] The world Health Organization(WHO)plays a very important role in combating lung cancer, as It is the primary cause of cancer-related fatalities worldwide. Through research, advocacy, and the public health initiatives, the WHO aims to enhance lungs cancer surveillance, diagnosis, and treatment to reduce the overall impact of the disease.

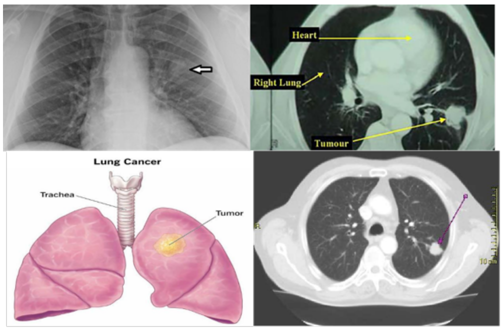

Figure 1. Lung cancer surveillance images

According to projections by the WHO, there were 10 million deaths from cancer and 19 million new cancer diagnoses in 2020. With an aging and growing population, as well as changes in lifestyle and other factors, these numbers are expected to continue rising in the coming years.

Lung cancer, accounting for more than 2 million newly registered cases or approximately 12% of overall cancer diagnoses, is the most prevalent cancer globally. Additionally, it is the primary cause of cancer-related fatalities, responsible for 18.2% of all cancer fatalities, with an estimated 2 million deaths attributed to lung cancer. According to The IARC, latest global cancer data Asia with the highest ratio of lungs cancer deaths as of 57% followed with Europe approximately 20%, America 14.5% and Africa 7 %.

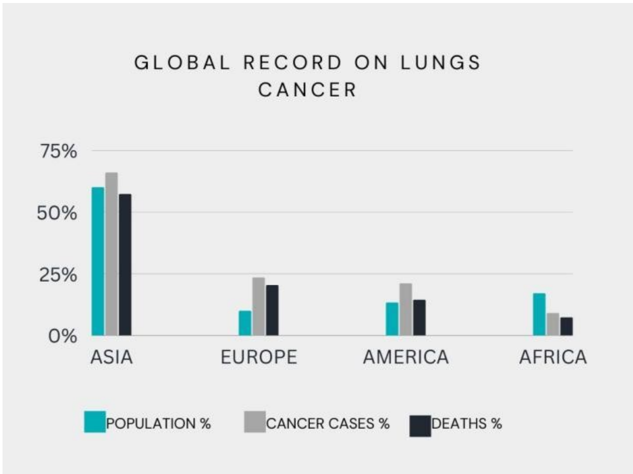

Figure 2. Global record of lung cancer

Since the phases can be read incorrectly due to the tumor's interconnected characteristics, lung cancer detection, and stage determination are both extremely important and challenging. There are three main categories for lung cancer: SCLC, LCT, and NSCLC [9]. SCLC (Small cell lung cancer) is the most severe and fast-growing type of lung cancer. It grows and spreads rapidly, frequently necessitating immediate and forceful therapy.[10] Only 15 % of lung cancer overall is NCLC. About 85.4% of lung cancer are NSCLC (non-small cell lung cancer) , which has a tendency to develop and disseminate more slowly than SC-lungs cancer

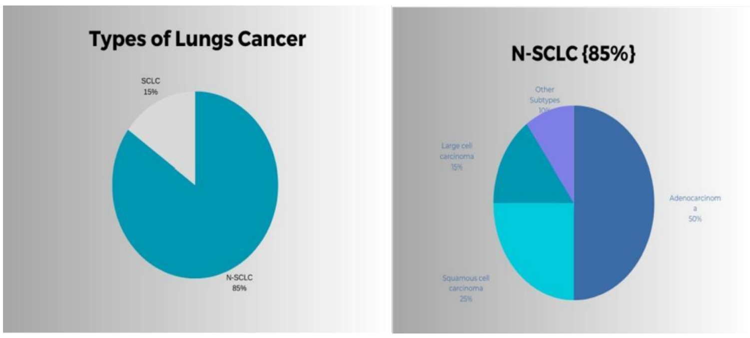

Figure 3. NSCLC comparison.

Options for treating NSCLC rely on a number of variables, including cancer's stage and molecular traits.[11] There are numerous contemporary algorithms and ways for finding lung cancer. LDCT (Low dose computed tomography), biopsies, (magnetic resonance imaging) MRI (positron emission tomography) PET, liquid biopsy, , X-rays and bone scans among other imaging modalities. The possibility of getting this cancer can be increased by a number of risk factors, including age, smoking, gender, family history, cannabis use, radon, and air pollution.[12]

Many scientific fields rapidly move in the direction of data-dependent methodology. In the area of medicine, image processing is essential because it allows healthcare professionals to analyze medical images quickly and correctly. machine learning is a form learning of system(Computers can learn without being expressly programmed thanks to the area of study known as machine- learning.) that enables system to learn from certain condition and make decisions by understand the world in relation to an idea's priority in the world.[13] machine learning is a subset of artificial intelligence that employs sophisticated neural-networks to process and learn from massive quantities of data. In order to acquire increasingly abstract representations of the data being examined, machine neural networks with numerous levels are used. The operation of the human brain are mimicked by a subgroup of machine learning systems called neural networks. They are made up of interconnected nodes, also known as "neurons," that process and send data by handling largely sizable, and labeled datasets.[14] machine learning algorithms are effective tools for medical image analysis because they can correctly recognize patterns and features in new pictures after being trained on a dataset of labeled images.[15] machine Learning applications for lung cancer early identification and diagnostics are becoming more popular as a result of this.[16]

The following table presents the %age, cancer cases of lungs diagnosed in the US by stage and their corresponding 5-year survival rates based on data from SEER 18.

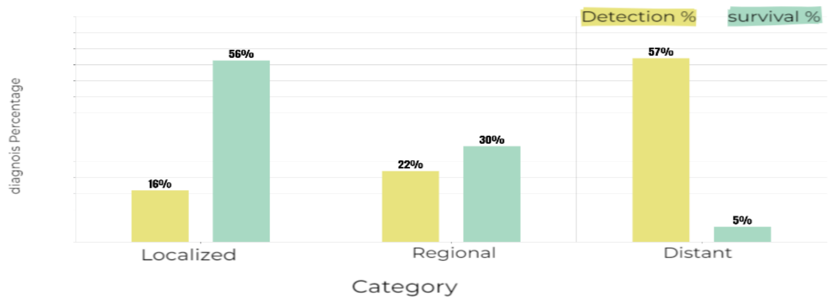

Figure 4. Lung cancer cases in US.

In this context, "localized" refers to lung cancer confined to the primary site, "regional" indicates the cancer has spread to regional lymph nodes, and "distant" signifies metastasized cancer. It's important to note that the exact %ages and survival rates are not specified in the provided information.

The main motive of this review paper is to conduct a extensive review of existing study on the use of conventional ML method, specifically SVM (Support vector machine), Naive bayes, ANN (artificial neural networks) in combination with the K-nearest neighbors algorithm (KNN) for the detection of lungs cancer. The paper aims to provide valuable insights into the advantages and limitations of utilizing machine learning techniques in diagnosing lungs conditions. By analyzing this research, readers can gain a deeper understanding of the benefits and drawbacks associated with the implementation of machine learning techniques in the analysis of lung-related conditions. The purpose of review papers is to give a concise and thorough overview of relevant studies in the field, allowing researchers to save time and effort by accessing a consolidated resource rather than having to refer to many different papers.

2. Materials and Methods

The main goal of conducting this systematic literature review was to identify and classify the most effective methods for lung cancer detection using machine learning. Systematic literature reviews entail acquiring and assessing existing studies using predetermined rating criteria. These reviews aid in determining the current state of knowledge and findings in a certain field. After the systematic literature review is done, all data gathered from primary sources is arranged and analyzed. This method produces a more cohesive, sensible, and dependable response to the research question at hand. Researchers can get meaningful insights and develop well-supported findings by analyzing and synthesizing current literature in a systematic manner. The methodical methodology of a systematic literature review improves the validity and dependability of study findings, laying a solid foundation for decision-making and future research in the field.

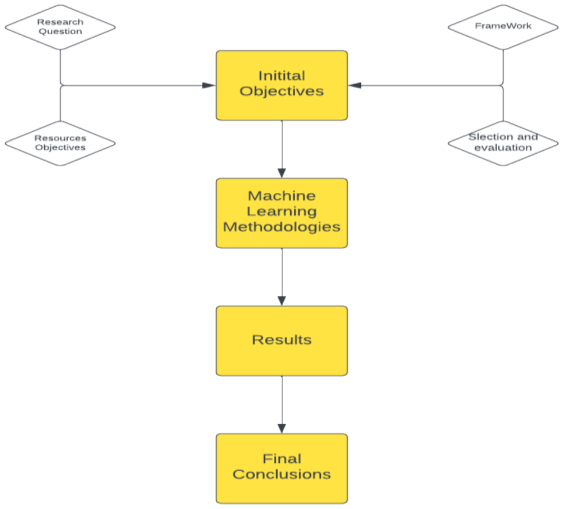

Figure 5. Flow chart of the paper

2.1 Research Objective

This article's main goal is to evaluate the improvements and advances made in lung disease diagnosis by applying machine learning techniques. Different approaches have developed and evolved throughout time. This paper examines recent developments in the field of lung disease detection research and highlights the progress made in these approaches. The paper offers a comprehensive overview of the real-world use of machine learning models for lung disease identification.

2.2 Research Questions

Here are questions which will cover in this review paper

- Best Machine learning technique in term of Accuracy?
- Best dataset available free?

2.3 Selection Criteria

The identification and selection of research articles that align with our research objective of lungs disease detection using machine learning methodologies is crucial for our analysis. Fig 1 provides the inclusion and exclusion criteria for this article, which facilitate the selection process. By applying these criteria, we can effectively analyze specific research articles that meet our requirements. This ensures that we focus on relevant studies and gather the most appropriate information for our research analysis.

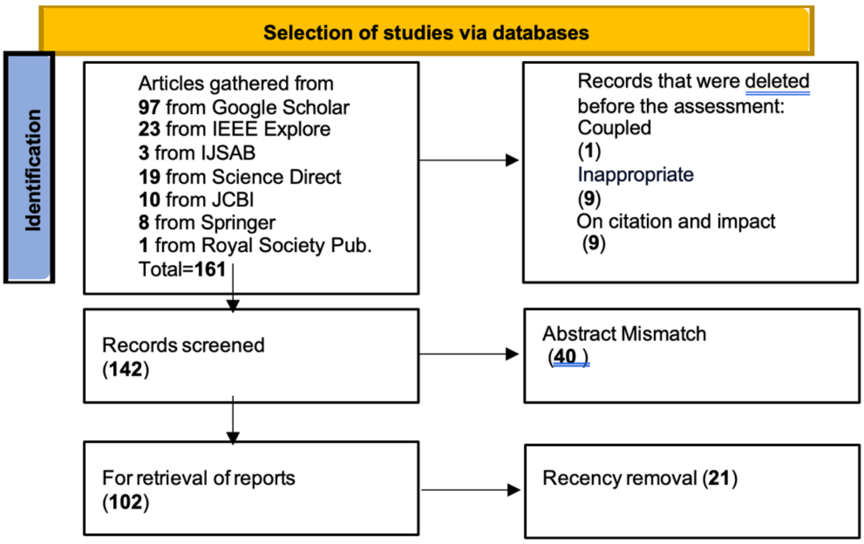

Figure 6. Prisma flow diagram of paper.

Our search identified 161 papers. Total of 60% of our papers are downloaded from Google Scholar, 14% from IEEE Explore, 2% from IJSAB, 12% from Science Direct, 6% from JCBI, 5% from Springer, 1%

from Royal Society Publications. 1 paper was screened out due to duplicate download. 9 did not fulfil inclusion criteria based on title. To verify the calibre of the work, We checked the reputation of the journal where the chosen paper was published and looked at how many times it has been cited. According to this filter, we removed another 9 papers from included papers. Then next it's time for screening, we have total of 142 papers left behind. After screening we study out the abstracts of 142 papers, out of 142 papers 40 was removed due to abstract mismatch from our study. When writing a review paper, it is common to prioritize recent and current research. Older papers are often given less importance or excluded altogether. The decision to remove papers based on their age depends on the specific requirements and scope of the review. Typically, the goal is to include research released in recent years to ensure review reflects the latest findings and advancements in the field. We removed 21 papers out of 102 according to filter based upon recency. After that we have 81 papers left for further studies and classification. We removed further 4 papers due to no imaging was used in research papers. When writing a review paper, it is common practice to exclude papers that do not provide significant or satisfactory results. The goal is to select papers that contribute important insights or outcomes to the field of study. If a paper lacks proper outcomes or fails to make a significant contribution, it may be considered less relevant and not included in the review. We removed 12 papers on this base. Every study accounts for the size of the dataset that was used. We remove another 10 papers due to ensure the inclusion of studies that offer more extensive and reliable results, research papers with small or insufficient dataset sizes may be excluded from the review. Finally we have 47 papers left behind to study and produce the effective results.

3. Search String

Criterion 1:

To identify relevant articles, the search query includes keywords related to lungs diseases and machine learning methodologies. The search query used is as follows:

("lungs disease" OR "lungs disease classification" OR "disease lungs" OR "lungs disease detection" OR "lungs disorder" OR "lungs care" OR "lungs infection") AND ("machine learning" OR "support vector machine" OR "svm")

Criterion 2:

Another search query is used to broaden the scope and capture articles related to lungs disorders and machine learning methodologies. The query is as follows:

("lungs" OR "lungs disorder" OR "lungs diagnose" OR "Pulmonary disease" OR "Pulmonary disease abnormality" OR "Pulmonary disease disorder" OR "Respiratory system disease" "Respiratory system disease abnormality") AND ("machine network" OR "dnn" OR "ml")

These criteria are applied to identify articles that focus on lungs diseases, lungs disorders, and their detection.

3.1 ML Techniques

The goal of the field of machine learning, or ML, is to teach computers how to learn and anticipate things without having to be explicitly programmed. It entails creating algorithms and models for data analysis, pattern recognition, and decision- or prediction-making based on that information. We examined SVM Naive Bayes, ANN, KNN, and other four machine learning algorithms in the framework of our review study. Every one of these algorithms takes a different approach to learning and predicting outcomes. We may use machine learning to obtain insights and deepen our understanding by utilizing these algorithms in our study.

3.2 SVM

SVM (Support Vector Machine) is a supervised learning technique type for classification analysis. It is particularly useful for datasets that are not easily separated in a linear manner, as it helps to minimize the mis-classification rat. In SVM, goal finding the point that is the minimum distance away from the classes and maximize the distance.[17] Figure 2 illustrates the design of SVM, where the Green and pink images represent distinct classes that are separated by a hyperplane. The margin and support vectors are associated with this separation, which play important roles in SVM, are appropriately labeled below.

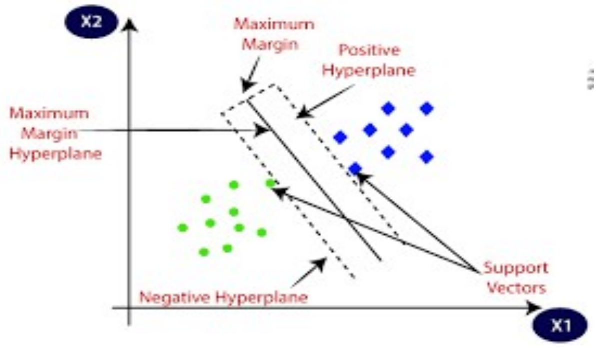

Figure 7. Key structure of SVM

3.3 ANN

The computer technique known as an artificial neural network, or ANN, is made up of networked processing units called neurons. When it comes to arranging information in response to outside inputs, these neurons are essential. There are two primary stages to utilizing ANNs: testing and training.

An ANN is taught to categorize fresh input using the patterns it has discovered from past data during the learning phase. During the testing phase, the network receives an input signal, processes it, and outputs the result. The study and calculation of the network produced this output.

ANN techniques have demonstrated their worth in a number of healthcare domains. They have been effectively used to the diagnosis and prognosis of lung cancer, breast cancer, and other illnesses.

3.4 Naive Bayes

The Naive Bayes algorithm employs the Bayes theorem as a classification technique. It makes the assumption that a dataset's characteristics are unrelated to one another. This implies that the presence or absence of one trait has no bearing on the presence or absence of another.

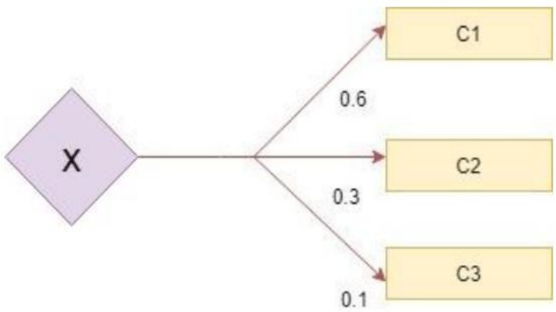

Figure 8. X data & class probability connection.

The Naive Bayes algorithm employs Bayes' theorem as a classification technique. This makes the assumption that the features of the dataset are unrelated to each other. This implies that the presence or absence of one quality has no effect on the presence or absence of the other.

3.5 KNN

K-Nearest Neighbors, or KNN, is an acronym for a well-liked machine learning method that is used to predict and categorize data. To locate related data points within a given collection is how it operates. The number of nearby points to consider while forming a forecast is represented by the variable "K". The algorithm looks at the K nearest neighbors based on similarity when a new data point is supplied, and it gives a label depending on what the majority of those neighbors are.

KNN's simplicity is one of its benefits. It may be applied to a variety of issues and makes no assumptions about the data. It can easily handle noisy data and manage several classes. Despite this, KNN may be costly and sluggish, particularly when dealing with big data sets. The choice of K is also important and can affect the performance of the algorithm. It is recommended to scale the features to ensure fairness. Overall, KNN is a versatile and easy-to-understand algorithm for classification tasks, providing a good balance between simplicity and accuracy.

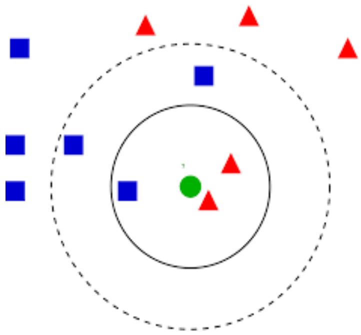

Figure 9. KNN model.

Table 1. Models with good accuracies

Algorithm Used	Dataset	Accuracy	Ref.
SVM	LIDC-IDRI	96.70 %	[18]
Naive Bayes	UCI	90%	[19]
KNN	(CT) images	99.2%	[20]
SVM	The Cancer Imaging Archive (TCIA)	79.7%	[21]
KNN	LIDC-IDRI	96%	[22]
ANN & others	LIDC-IDRI	90%	[23]
ANN	JSRT	82.43%	[24]
ANN	LIDC-IDRI	90%	[25]
SVM	50 CT pictures	60%	[26]
Naive Bayes	The LCDS	87%	[27]
ANN	LIDC-LDRI	96%	[28]
SVM	LIDC-IDRI	91.60%	[29]
KNN	Data World source	100%	[30]
SVM & Random Forest	CT scan images	94.5%	[31]
SVM	LDCT images	96.25%	[32]

Table 2. Average accuracies of models

Sr. No	Model	Accuracy %
1	SVM	84.45%
2	Naive Bayes	88.5%
3	KNN	98.4%
4	ANN	85.35%

4. Discussion

The use of machine learning algorithms for lung cancer early detection is the main topic of discussion in this review study. Early detection of lung cancer is a significant challenge and is associated with significant risk. The research underscores the significance of prompt identification, given its substantial influence on treatment results and patient survival rates.

With an emphasis on lung cancer specifically, the paper emphasizes the importance of machine learning algorithms in assisting in the diagnosis of lung disorders. The efficacy of four machine learning techniques—Artificial Neural Networks (ANN), Support Vector Machines (SVM), k-Nearest Neighbors (KNN), and Naive Bayes—in the identification of lung cancer was assessed. With an average accuracy rate of 98.5%, KNN showed the greatest accuracy among these techniques.

The paper also covers the significance of datasets for machine learning model evaluation and training in the context of lung cancer diagnosis. After a number of datasets from different articles were examined, it was determined that the LIDC dataset was best suited for this objective. The LIDC dataset offers a thorough compilation of imaging data related to lung cancer, facilitating the development and testing of precise and dependable models by researchers.

The limits and difficulties in the realm of lung cancer detection are also acknowledged in the debate. The precise etiology and full therapy of lung cancer are still unknown, despite the progress made in machine learning techniques. To improve the precision and application of machine learning algorithms in early lung cancer detection, more study is required.

This review study concludes by highlighting the promise of machine learning methods, especially KNN, in the early identification of lung cancer. The efficiency and dependability of these methods are further enhanced by the use of the LIDC dataset. The review's conclusions support current efforts to enhance lung cancer detection techniques and can direct further research in this crucial field.

5. Data Set

Many computer-based methods have been proposed for diagnosing lung cancer. To accurately evaluate their diagnostic performance and ensure reliable results, it is crucial to have access to a consistent and trustworthy dataset of cancer images. This section we will discuss and Examine the real-world datasets utilized to evaluate how well these suggested methods work. for detecting lung cancer. Understanding the characteristics and key information about these datasets is essential for industry professionals and scholars.

Table 3. Accuracies on different data sets

Reference	Name of Database	Samples used	Accuracy
[33]	NBIA-ELCAP	(113)	74%
[34]	LIDC-IDRI	(33)	90.91
[35]	LIDC	N/A	92.78
[36]	LIDC-IDRI	(8296)	82.3
[37]	Private	(120	90
[38]	LIDC	(110)	93.30

6. Review Analysis

They [39] combine image processing techniques with biomedical methods and data analysis to improve accuracy and precisely and early detection of lung cancer. They process the Competed Tomography scan images of the lungs, identify the important areas, and separate them. The Random Forest method is used to identify unique characteristics, while the SURF algorithm extracts The features extracted from enhanced images include entropy, correlation, power, and variance. These features are used for classifying the images into safe or cancerous categories. The dataset comprises CT scan images, and the SVM method and random forest algorithm are employed for the entire process. The SVM method achieves the better results, with an overall effectiveness of 94%, sensitivity of 74%, recall of 67%, and specificity of 78%

Nidhi research done in 2019 [40] introduced an self-operating system to detect cancer in lungs in its early stages. They utilized CT scan images from the CIAD in the format of D-ICOM. These format images underwent pre-processing, including techniques like Smoothening, Median Filtering, and Contrast Adjustment,, to enhance image quality & reduce noise. To segment the images, After converting the images into binary format, morphological opening operations were performed on them. In the feature extraction process, important attributes such as We assessed the size, perimeter, and eccentricity (roundness). A supervised learning classifier called SVM was used to classify the image as either Depending on these characteristics, normal or abnormal. According to the authors, the proposed methodology accurately detects cancer in the early stages.

In their study, Boban et al. [41] utilized machine learning algorithms to analyze 400 lung disease videos, specifically employing the MLP, KNN and SVM classifiers, with a focus on CT scan images. Evaluation of the classifiers' effectiveness after extracting features and comparing their accuracy. To eliminate irrelevant content from CT scan images, the GLCM was employed to select the most significant attribute. The

MLP classifier achieved 98% accuracy, SVM achieved 70.45% accuracy, and KNN achieved 99.2% accuracy in the classification process.

Maleki et al. [42] developed a technique called k-Nearest Neighbors, where they utilized a genetic algorithm to effectively select features, decrease the size of the dataset and accelerate the classifier. They conducted experiments to determine the optimal value of k, which improved the accuracy of given algorithm. after applying to a database for lung cancer, the proposed solution achieved 100% accuracy.

In their work, Banerjee the researcher[43] give a infrastructure for tumor detection, employing ANN, SVM and random forests as ML algorithms. The study revealed that artificial neural networks performed better in terms of Both area and texture-dependent characteristics were considered in the analysis.. When evaluating the accuracy of the suggested model, It was noticed that the accuracy of the model increased, but at the same time, the recall decreased. For region-based features, the accuracy for Random Forest, SVM, and ANN was 79%, 86%, and 92%, respectively. The achieved accuracy for texture-based features was 70% for Random Forest, 80% for SVM, and 96% for ANN.

In 2019, Ibrahim M. Nasser [46] presented a method for detecting the presence or absence of lung cancer using Artificial Neural Networks (ANN). The symptoms of the disease were utilized for diagnosis. The dataset used in the study is described in the table. About 97% of the time, the ANN model correctly predicted the existence of lung cancer. Additionally, after 1,418,105 training iterations, the model had a training error rate of less than 1%. The analysis indicated that "Age" was the factor that had the most significant impact on the results.

Lakshmanaprabu [47] developed a classification model for CT images based on an Optimal Deep Neural Network (ODNN) and Linear Discriminant Analysis (LDA). The model is designed to classify lung nodules, and prediction of lung cancer uses the Modified in Gravitational Search to optimise. The experimental analysis utilized a standard CT database consisting of fifty low dosage lung cancer computed tomography images. The developed algorithm was compared with other models already in use such as K-Nearest Neighbors (KNN), Support Vector Machine (SVM) and others. According to the experimental findings, the constructed model performed the best with an accuracy of about 95%, sensitivity of 96% and specificity of 94%.

7. Conclusion

To sum up, the purpose of this review study was to investigate the use of machine learning in the identification of lung cancer. The efficacy of four methods—Artificial Neural Networks (ANN), Support Vector Machines (SVM), KNN (k-Nearest Neighbors), and Naive Bayes—was assessed in this field. KNN outperformed the others in terms of precision, with an average precision rate of 98.5%.

To find the best dataset for lung cancer detection, a thorough analysis of other datasets was also carried out. The best alternative for this purpose was found to be the LIDC dataset after a number of studies were analyzed.These results demonstrate the potential of machine learning techniques to aid in the diagnosis of lung cancer, with the KNN demonstrating particularly encouraging outcomes. Furthermore, the LIDC dataset may be used to improve the detection systems' accuracy and dependability. All things considered, this review offers insightful information about the use of machine learning to the detection of lung cancer, suggesting possible directions for further study and advancement in this important area.

References

1. Alfredo Vellido. The importance of interpretability and visualization in machine learning for applications in medicine and health care. Neural computing and applications, 32(24):18069– 18083, 2020.
2. Jolanta Malyszko, Petra Tesarova, Giovambattista Capasso, and Anna Capasso. The link between kidney disease and cancer: complications and treatment. The Lancet, 396(10246):277–287, 2020.
3. Mesut To˘ga ¸c ar, Burhan Ergen, and Zafer C¨omert. Detection of lung cancer on chest ct images using minimum redundancy maximum relevance feature selection method with convolutional neural networks. Biocybernetics and Biomedical Engineering, 40(1):23–39, 202
4. Giovanni A Fontana. Before we get started: what is a cough? lung, 186(Suppl 1):3–6, 2008.
5. Nikita Joshi, James M Walter, and Alexander V Misharin. Alveolar macrophages. Cellular immunology, 330:86–90, 2018.
6. Janna C Nawroth, Anne M van der Does, Amy Ryan, and Eva Kanso. Multiscale mechanics of mucociliary clearance in the lung. Philosophical Transactions of the Royal Society B, 375(1792):20190160, 2020.
7. Anton Vonk Noordegraaf, Kelly Marie Chin, Fran¸cois Haddad, Paul M Hassoun, Anna R Hemnes, Susan Roberta Hopkins, Steven Mark Kawut, David Langleben, Joost Lumens, and Robert Naeije. Pathophysiology of the right ventricle and of the pulmonary circulation in pulmonary hypertension: an update. European Respiratory Journal, 53(1), 2019.
8. Elias Dritsas and Maria Trigka. Lung cancer risk prediction with machine learning models. Big Data and Cognitive Computing, 6(4):139, 2022.
9. Pravin Shende, Steffi Augustine, and Bala Prabhakar. A review on graphene nanoribbons for advanced biomedical applications. Carbon Letters, 30:465–475, 2020.
10. Charles M Rudin, Elisabeth Brambilla, Corinne Faivre-Finn, and Julien Sage. Small-cell lung cancer. Nature Reviews Disease Primers, 7(1):3, 2021.
11. Valeria Relli, Marco Trerotola, Emanuela Guerra, and Saverio Alberti. Abandoning the notion of non-small cell lung cancer. Trends in Molecular Medicine, 25(7):585–594, 2019.
12. Krishna Chaitanya Thandra, Adam Barsouk, Kalyan Saginala, John Sukumar Aluru, and Alexander Barsouk. Epidemiology of lung cancer. Contemporary Oncology/Wsp´o lczesna Onkologia, 25(1):45–52, 2021.
13. Jahanzaib Latif, Chuangbai Xiao, Azhar Imran, and Shanshan Tu. Medical imaging using machine learning and deep learning algorithms: a review. In 2019 2nd International conference on computing, mathematics and engineering technologies (iCoMET), pages 1–5. IEEE, 2019.
14. Nikolaus Kriegeskorte and Tal Golan. Neural network models and deep learning. Current Biology, 29(7):R231–R236, 2019.
15. Yoshua Bengio, Yann Lecun, and Geoffrey Hinton. Deep learning for ai. Communications of the ACM, 64(7):58–65, 2021.
16. Diego Ardila, Atilla P Kiraly, Sujeeth Bharadwaj, Bokyung Choi, Joshua J Reicher, Lily Peng, Daniel Tse, Mozziyar Etemadi, Wenxing Ye, Greg Corrado, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nature medicine, 25(6):954–961, 2019.
17. PR Radhika, Rakhi AS Nair, and G Veena. A comparative study of lung cancer detection using machine learning algorithms. In 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), pages 1–4. IEEE, 2019.
18. Murillo B Rodrigues, Raul Victor M Da Nobrega, Shara Shami A Alves, Pedro Pedrosa Reboucas Filho, Joao Batista F Duarte, Arun K Sangaiah, and Victor Hugo C De Albuquerque. Health of things algorithms for malignancy level classification of lung nodules. IEEE Access, 6:18592–18601, 2018.
19. Muhammad Imran Faisal, Saba Bashir, Zain Sikandar Khan, and Farhan Hassan Khan. An evaluation of machine learning classifiers and ensembles for early stage prediction of lung cancer.
20. Binila Mariyam Boban and Rajesh Kannan Megalingam. Lung diseases classification based on machine learning algorithms and performance evaluation. In 2020 International Conference on Communication and Signal Processing (ICCSP), pages 0315–0320. IEEE, 2020.

21. Mizuho Nishio, Mitsuo Nishizawa, Osamu Sugiyama, Ryosuke Kojima, Masahiro Yakami, Tomohiro Kuroda, and Kaori Togashi. Computer-aided diagnosis of lung nodule using gradient tree boosting and bayesian optimization. PloS one, 13(4):e0195875, 2018.
22. Murillo B Rodrigues, Raul Victor M Da Nobrega, Shara Shami A Alves, Pedro Pedrosa Reboucas Filho, Joao Batista F Duarte, Arun K Sangaiah, and Victor Hugo C De Albuquerque. Health of things algorithms for malignancy level classification of lung nodules. IEEE Access, 6:18592–18601, 2018.
23. Lea Marie Pehrson, Michael Bachmann Nielsen, and Carsten Ammitzbøl Lauridsen. Automatic pulmonary nodule detection applying deep learning or machine learning algorithms to the lidc-idri database: a systematic review. Diagnostics, 9(1):29, 2019.
24. O¨zge Gu¨naydin, Melike Gu¨nay, and O¨znur S¸engel. Comparison of lung cancer detection algorithms. In 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT), pages 1–4. IEEE, 2019.
25. Lea Marie Pehrson, Michael Bachmann Nielsen, and Carsten Ammitzbøl Lauridsen. Automatic pulmonary nodule detection applying deep learning or machine learning algorithms to the lidc-idri database: a systematic review. Diagnostics, 9(1):29, 2019.
26. N Bhaskar and TS Ganashree. Lung cancer detection with fpcm and watershed segmentation algorithms. In Advances in Decision Sciences, Image Processing, Security and Computer Vision: International Conference on Emerging Trends in Engineering (ICETE), Vol. 1, pages 687–695. Springer, 2020.
27. Yasemin Gultepe. Performance of lung cancer prediction methods using different classification algorithms. Lung, 2:1–760, 2021.
28. Nikita Banerjee and Subhalaxmi Das. Prediction lung cancer–in machine learning perspective. In 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA), pages 1–5. IEEE, 2020.
29. Yutong Xie, Yong Xia, Jianpeng Zhang, Yang Song, Dagan Feng, Michael Fulham, and Weidong Cai. Knowledge-based collaborative deep learning for benign-malignant lung nodule classification on chest ct. IEEE transactions on medical imaging, 38(4):991–1004, 2018.
30. Negar Maleki, Yasser Zeinali, and Seyed Taghi Akhavan Niaki. A k-nn method for lung cancer prognosis with the use of a genetic algorithm for feature selection. Expert Systems with Applications, 164:113981, 2021.
31. Kyamelia Roy, Sheli Sinha Chaudhury, Madhurima Burman, Ahana Ganguly, Chandrima Dutta, Sayani Banik, and Rayna Banik. A comparative study of lung cancer detection using supervised neural network. In 2019 International Conference on Opto-Electronics and Applied Optics (Optronix), pages 1–5. IEEE, 2019.
32. Dakhaz Mustafa Abdullah, Adnan Mohsin Abdulazeez, and Amira Bibo Sallow. Lung cancer prediction and classification based on correlation selection method using machine learning techniques. Qubahan Academic Journal, 1(2):141–149, 2021.
33. Hiram Madero Orozco, Osslan Osiris Vergara Villegas, Humberto de Jesu´s Ochoa Dom´ınguez, and Vianey Guadalupe Cruz S´anchez. Lung nodule classification in ct thorax images using support vector machines. In 2013 12th Mexican International Conference on Artificial Intelligence, pages 277–283. IEEE, 2013.
34. Henry Krewer, Benjamin Geiger, Lawrence O Hall, Dmitry B Goldgof, Yuhua Gu, Melvyn Tockman, and Robert J Gillies. Effect of texture features in computer aided diagnosis of pulmonary nodules in low-dose computed tomography. In 2013 IEEE International Conference on Systems, Man, and Cybernetics, pages 3887–3891. IEEE, 2013.
35. Leonardo Barros Nascimento, Anselmo Cardoso de Paiva, and Arist´ofanes Corrˆea Silva. Lung nodules classification in ct images using Shannon and Simpson diversity indices and SVM. In Machine Learning and Data Mining in Pattern Recognition: 8th International Conference, MLDM 2012, Berlin, Germany, July 13-20, 2012. Proceedings 8, pages 454–466. Springer, 2012.
36. Giovanni da Silva, Arist´ofanes Silva, Anselmo de Paiva, and Marcelo Gattass. Classification of malignancy of lung nodules in ct images using convolutional neural network. In Anais do XVI Workshop de Inform´atica M´edica, pages 2481–2489. SBC, 2016.
37. Bhawna Gupta and Shamik Tiwari. Brain tumor detection using curvelet transform and support vector machine. International Journal of Computer Science and Mobile Computing, 3(4):1259–1264, 2014.
38. Jinsa Kuruvilla and K Gunavathi. Lung cancer classification using neural networks for ct images. Computer methods and programs in biomedicine, 113(1):202–209, 2014.

39. Kyamelia Roy, Sheli Sinha Chaudhury, Madhurima Burman, Ahana Ganguly, Chandrima Dutta, Sayani Banik, and Rayna Banik. A comparative study of lung cancer detection using supervised neural network. In 2019 International Conference on Opto-Electronics and Applied Optics (Optronix), pages 1–5. IEEE, 2019.
40. Nidhi S Nadkarni and Sangam Borkar. Detection of lung cancer in ct images using image processing. In 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), pages 863–866. IEEE, 2019.
41. Binila Mariyam Boban and Rajesh Kannan Megalingam. Lung diseases classification based on machine learning algorithms and performance evaluation. In 2020 International Conference on Communication and Signal Processing (ICCSP), pages 0315–0320. IEEE, 2020.
42. Negar Maleki, Yasser Zeinali, and Seyed Taghi Akhavan Niaki. A k-nn method for lung cancer prognosis with the use of a genetic algorithm for feature selection. Expert Systems with Applications, 164:113981, 2021.
43. Nikita Banerjee and Subhalaxmi Das. Prediction lung cancer–in machine learning perspective. In 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA), pages 1–5. IEEE, 2020.
44. DendiGayathri Reddy, Emmidi Naga Hemanth Kumar, Desireddylohithsaicharan Reddy, and P Monika. Integrated machine learning model for prediction of lung cancer stages from textual data using ensemble method. In 2019 1st International Conference on Advances in Information Technology (ICAIT), pages 353–357. IEEE, 2019.
45. O¨zge Gu¨naydin, Melike Gu¨nay, and O¨znur S¸engel. Comparison of lung cancer detection algorithms. In 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT), pages 1–4. IEEE, 2019.
46. Ibrahim M Nasser and Samy S Abu-Naser. Lung cancer detection using artificial neural network. International Journal of Engineering and Information Systems (IJEAIS), 3(3):17–23, 2019.
47. SK Lakshmanaprabu, Sachi Nandan Mohanty, K Shankar, N Arunkumar, and Gustavo Ramirez. Optimal deep learning model for classification of lung cancer on ct images. Future Generation Computer Systems, 92:374–382, 201