ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

AN ANALYSIS ON ADVANCES IN LUNG CANCER DIAGNOSIS WITH MEDICAL IMAGING AND DEEP LEARNING TECHNIQUES: CHALLENGES AND OPPORTUNITIES

VAHIDUDDIN SHARIFF[1], CHIRANJEEVI PARITALA[2], KRISHNA MOHAN ANKALA[3]

[1]Research Scholar, Jawaharlal Nehru Technological University Kakinada, Department of CSE, Kakinada, India
[2]Professor, Amrita Sai Institute of Science and Technology, Department of CSE, Bathinapadu, India
[3] Professor, UCEK, Jawaharlal Nehru Technological University Kakinada, Department of CSE, , Kakinada, India

[1] shariff.v@gmail.com, [2] mailmeparitala@gmail.com, [3] krishna.ankala@gmail.com

ABSTRACT

This abstract provides an overview of numerous studies on the identification and diagnosis of lung cancer using medical imaging and deep learning. However, techniques like the VGG16 (Visual geometry group) model, SSD (Social Ski-Driver), SVM (Support Vector Machines), CNN (Convolutional Neural Network), and CSA (crow search algorithm) have demonstrated encouraging results in accurately identifying and classifying lung cancer from CT images. Accuracy still needs to be improved. Pre-processing operations including edge identification, picture resampling, and segmentation are crucial for improving the visual appeal of input photos. The use of computer-aided diagnosis (CAD) systems can significantly increase the efficacy of diagnostic classifiers and pre-processing methods. For the identification of lung cancer, the use of deep learning models and meta-heuristic-based optimisation techniques can produce accurate and durable diagnosis models. Effective lung cancer treatment depends on early detection. The limits of the current histology-based diagnostic methods are also covered in the paper, as well as the potential of molecular biomarkers for classification. In categorising the morphology of lung cancer, selecting the optimal course of action, and forecasting the outcomes of systemic therapy for non-small cell lung cancer, machine learning algorithms have demonstrated success. A larger dataset is needed for the veracity of these conclusions, but the paper claims that doctors are increasingly employing machine learning to better understand their patients and create tailored treatment regimens.

Keywords: *Computer-Aided Diagnosis, Support Vector Machines, Convolutional Neural Network, Visual Geometry Group, Social Ski-Driver, Crow Search Algorithm.*

1. INTRODUCTION

Lung cancer usually strikes both men and women due to the lungs' unchecked cell proliferation. This leads to a major breathing issue in the chest's exhalation and inhalation regions. According to the World Health Organization, tobacco use and passive smoking are the main causes of lung cancer [1]. When compared to all other kinds of cancer, the morbidity and mortality rate of lung cancer is currently steadily rising [2].An X-ray image of the lung obtained during a lung mammography is a frequent diagnostic tool for detecting lung calcifications, tumours, and swellings. These are all symptoms of preliminary lung cancer. Nevertheless, it might be challenging to recognize these symptoms from the lung imaging. Additionally, erroneous interpretation of these images results in a dangerously inaccurate diagnosis. Think about the situation of a false negative diagnosis, in which a lung cancer in its early stages is mistakenly diagnosed as a typical instance [3].The death rate may be decreased by adopting early lung cancer forecasting.

Worldwide, lung cancer is the most common cancer-related reason for death, and optimal

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

outcomes for therapy depend greatly on early identification. Lung nodules, which may be an indication of lung cancer, can be found using medical imaging techniques including computed tomography (CT) scans. Deep learning techniques have recently made significant strides, and this has showed promise for lung cancer diagnostic efficiency and accuracy using medical imaging data. This review paper seeks to give an overview of the current developments in lung cancer diagnosis utilising medical imaging and deep learning techniques, as well as the difficulties and potential benefits of adopting these methods.

Previous research has focused on a variety of methods, including the use of various deep learning models (e.g., VGG16, CNN, AlexNet) and meta-heuristic optimisation techniques (e.g., CSA) for lung cancer detection utilising medical imaging and deep learning techniques. These investigations have shown encouraging outcomes in correctly diagnosing and categorising lung cancer using CT scans. The use of larger datasets has been advised to further validate the findings; however accuracy improvement has been a barrier throughout these investigations.

Our work, on the other hand, attempts to present an overview of various studies on lung cancer diagnosis using medical imaging and deep learning, with a focus on the difficulties encountered and the possible advantages of employing these approaches. In this section, we go through the value of pre-processing operations (such image denoising, edge identification, and segmentation) and how to use CAD systems to improve diagnostic classifiers. Our study emphasises the need for larger datasets to support findings and the growing use of machine learning by physicians to better understand their patients and design individualised treatment plans.

Evaluate the performance and accuracy of different deep learning models, including VGG16, SSD, SVM, CNN, and CSA, for identifying and classifying lung cancer from CT images.

Investigate the impact of pre-processing operations, such as edge identification, picture resampling, and segmentation, on improving the quality of input photos for lung cancer diagnosis using medical imaging and deep learning techniques.

The next section will discuss the current challenges in using mammography to identify lung cancer. While Section 4 will examine the shortcomings in current research approaches, Section 3 will examine the importance of datasets in identifying lung cancer. The conclusion and next steps for this study will be covered in Section 5 finally.

2. CURRENT CHALLENGES IN LUNG CANCER DETECTION WITH MAMMOGRAPHY

Mammograms are typically used to detect cancer, according to Pramanik, Payel, et al. [4]. The problem is solved in two stages by Pramanik, Payel, and others. In the first step, the renowned VGG16 model [29] is used to obtain deep features from the input mammograms. The next step is to get the optimal features subset. To do this, they use the Social Ski-Driver (SSD) [30] technique, a meta-heuristic technique coupled with Adaptive Beta Hill Climbing-based local search. The best feature subset is used to do the classification by the K-nearest neighbours (KNN) classifier [5].96.07% accuracy is provided by the suggested model and the flow is shown in figure 1.

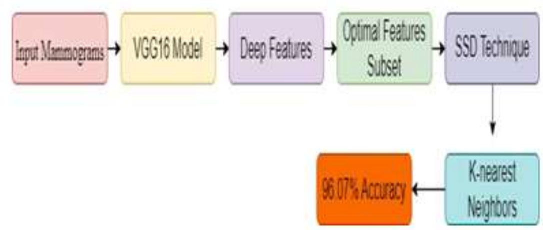

Figure 1: Process for Cancer Detection using Mammograms, VGG16, SSD, and KNN Classifier

The CT scan photos of the patients were categorised into normal and abnormal by Rahane et al. [6]. Segmenting the aberrant photos allows the tumour region to be the main focus. Classification performed using image-extracted features. The efficient method to recognise cancer of the lung and its stages also tries to generate more exact results by utilising SVM [7] and image processing techniques.

Continuous surveillance was used by Palani and Venkatalakshmi [12] to forecast the emergence of lung cancer. This was accomplished via fuzzy cluster-linked augmentation with categorization. It is crucial for producing effective picture identification. Instead, they used the fuzzy -means clustering approach [31] to better distinguish between the properties of the transitional area and those of the malignant lung picture. The Otsu thresholding approach was used in this study to separate the transition zone from the malignant cell's representation. To optimise segmentation display, the right edge picture is also employed in conjunction with the morphological, thinning

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

method, gradual classification to be done, a novel incremental classification technique is integrated with the present Association Rule Mining (ARM), conventional decision tree (CDT), and CNN [32]. Common photos from the database were used for the procedures. The results of the study show an improvement in prediction the system's precision.

The accuracy of diagnosing lung cancer disease still needs to be improved, according to Pawar, Vikul J et al.'s [13] examination of CT image segmentation. Despite the fact that their objectives were successfully met through the efficient application of numerous techniques like thresholding and classification methods like ANN, SVM, [33] and others, it is noted that further research is required. To increase accuracy and development, they used computer diagnostic systems (CAD). In order to meet the goals of the Medical Imaging sector by using Image Processing technology, a CAD system of this kind must focus on the reliability of pre-processing techniques, segmentation methods, noise reduction methods, and effective diagnosis classifier, as illustrated in figure2.

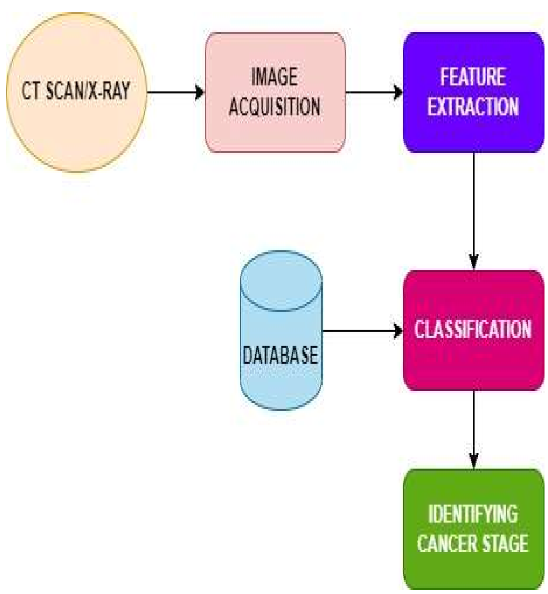

Figure:2 The processes for using CT and X-ray images in an image processing technique for diagnosing lung disease [13]

The relevance of automated lung illness identification is discussed by Shakeel, P. Mohamed et al. (2019) [14]. This is a difficult task since noise signals are present in the pictures taken in the image capture process. The paper goes on to discuss how important lung cancer pre-processing is for enhancing the quality of input pictures. Edge detection, lung image resampling, lung image augmentation, and image denoising are some of the processes in this pre-processing [34, 35]. Denoising an image is a vital pre-processing step that gets rid of the noise while, to the maximum extent possible, keeping the edges along with other full features. The paper focuses on enhancing lung image quality and lung cancer diagnosis by lowering misclassification. The noise from the lung CT scans acquired from the Cancer Imaging Archive (CIA) dataset was successfully reduced using the weighted mean histogram equalisation technique. This reduced image noise, enhancing the image's quality. The impacted region's spectral properties were recovered after the injured region was segmented using the improved robust clustering technique (IPCT). These traits were analysed using a deep learning immediately trained neural network for predicting lung cancer. Utilising simulation results from MATLAB, the system's effectiveness was evaluated. The method guaranteed 98.42% accuracy with a 0.038-percentage-point minimal classification error.

Deep learning models are discussed in Subramanian, R. Raja, et al.'s (2020) [15] discussion of their usage in the identification and classification of lung cancer. The study emphasises that lung cancer is one of the worst illnesses in the world and that successful treatment depends on early identification. Lung cancer is frequently found using blood test results and computed tomography (CT) scans. For a precise diagnosis, CT scans must be categorised as normal or abnormal.The study's suggested model for detecting lung cancer is trained using a Convolutional neural network (CNN) [16] and makes use of the AlexNet model [17] and the softmax layer shown in figure2. The suggested model is used in the study to reach an accuracy level of 99.52% utilising a dataset of CT pictures in jpg format. The findings imply that the suggested model may be applied as a reliable and long-lasting diagnostic model for lung cancer detection.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

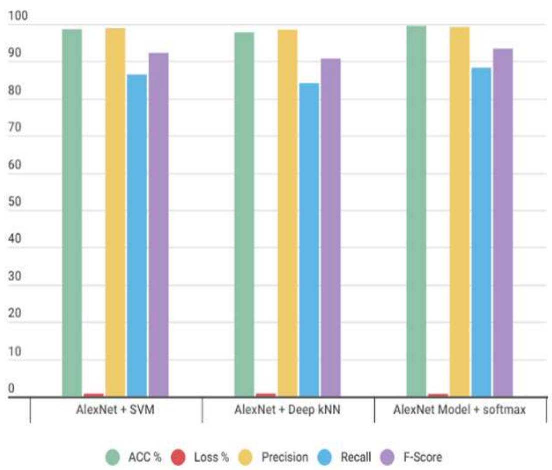

Figure3: Algorithm inputs and results from observation [15]

The application of meta-heuristic-based optimisation methods, particularly the crow search algorithm (CSA) [19], for detecting lung cancer in CT scans is covered by Alagarsamy, Saravanan, et al. (2021) [18]. Due to the intricate structures shown in CT scans, the author emphasises the difficulty in predicting lung cancer and the critical role that early identification plays in saving lives. Additionally, the review discusses the shortcomings of manual segmentation methods and the demand for automated approaches to increase precision.The crow search algorithm is described in the research as an effective technique for detecting lung cancer in CT scans. To properly pinpoint the cancer and healthy tissues, the system performs clustering and a flight process to look for comparable pixels. According to the review, the CSA algorithm outperforms more traditional methods like FCM and PSO in terms of sensitivity, 'MSE, PSNR, and DOI values. Using CSA, identifying the tumour and normal tissue takes only a brief amount of time.

A summary of a study that suggests a new method for diagnosing lung cancer utilising image processing and artificial intelligence from CT scan pictures is given in Xu, Yeguo et al. (2022). The Alexnet network is used in the study to diagnose both healthy and malignant cases, and a modified version of Satin Bowerbird Optimisation Algorithm is used to construct the Alexnet architecture and choose its characteristics optimally. The study's findings demonstrate that the suggested approach, which has a 95.96% accuracy rate and a higher harmonic mean value than other methods with similar F1 scores, is an effective tool for diagnosing lung cancer.The suggested method's best test recall scores (98.06%) also point to a greater rate of retrieving pertinent examples for the pictures. In general, the study finds that applying the suggested methodology as a fresh deep-learning-based methodology may offer more accuracy and can address the issue of choosing the right hyper parameters for the deep-learning-based methodology procedures for the targeted scenario. The importance of this work involves the development of an effective tool for lung cancer detection using CT images, which can significantly affect the chances of a successful cure.

Deep learning models, meta-heuristic optimisation methods, and computer-aided diagnostic (CAD) systems have all been the subject of recent research on lung cancer detection utilising medical imaging techniques. The research have used methods to increase the precision of lung cancer detection, including feature extraction, picture segmentation, and classification. The authors go over a number of methods and procedures for identifying and evaluating lung cancer utilising imaging technology. These publications all share the use of deep learning models and machine learning methods for image processing and analysis.

Using CT and mammography images for lung cancer detection and diagnosis, the writers of this topic mostly explore the use of different image processing and deep learning approaches. The trials' goals include enhancing lung cancer early detection and diagnostic accuracy. These researches have made use of a variety of methods, including the VGG16 model, the Social Ski-Driver (SSD) methodology, SVM, fuzzy cluster-linked augmentation with classification, and deep learning models like CNN and CSA.Studies have shown how crucial pre-processing methods such image denoising, edge identification, lung image resampling, and segmentation are for increasing diagnostic precision. Regardless of the success of these approaches, the studies highlight the need for additional developments in the precision of diagnosis and dependability of pre-processing strategies, segmentation methods, noise reduction techniques, and efficient diagnosis classifiers. In general, the papers included in this subject provide a potential method for applying deep learning and image processing to identify and diagnose lung cancer early.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

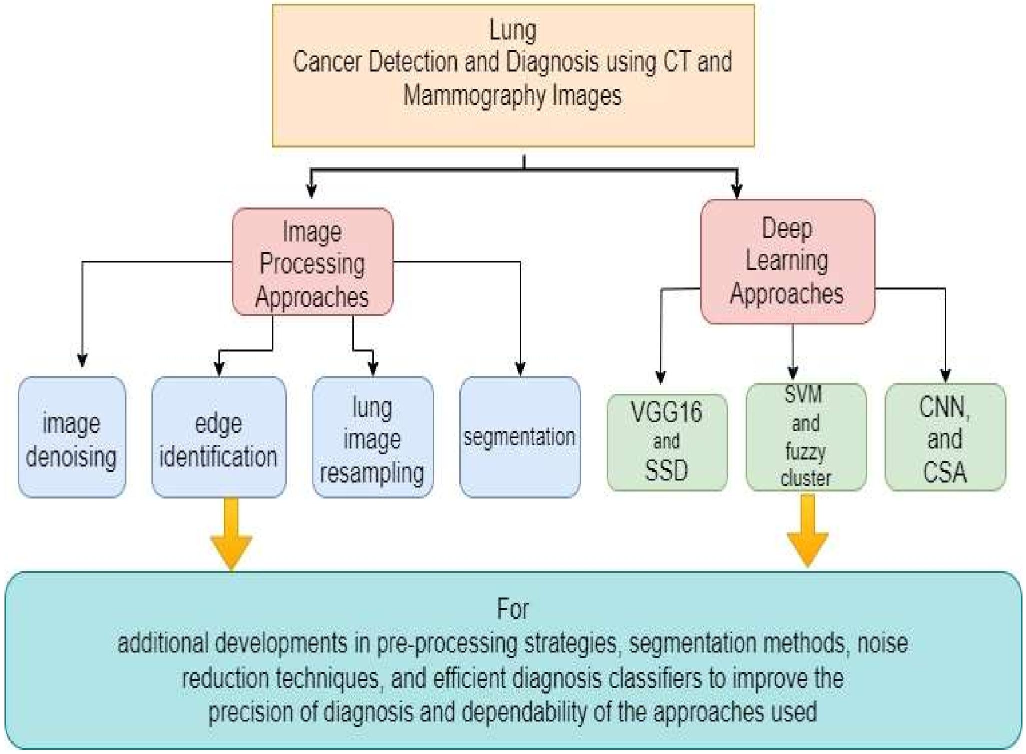

Figure 4: Process Flow Diagram for Enhancing Lung Cancer Patient Outcomes with Machine Learning using CT and mammography images

Table1: Comparative Analysis of Advantages, Research Gaps and Disadvantages in Lung Cancer Detection Studies with mammography

S No	Author	year	Advantages	Research Gap	Disadvantages
1	Pramanik, Payel, et al. [4].	2022	High efficiency and accuracy in cancer detection.	Dataset diversity, model comparison	False positives/negatives, lack of interpretability.
2	Rahane et al.[6]	2018	SVM is used to produce findings that are more accurate, improving the results for patients through early diagnosis.	There isn't any contrast to current state-of-the-art techniques.	Limited use of medical data
3	Palani and Venkatalakshmi [12]	2019	Increased system accuracy in forecasting the development of lung cancer.	Challenges of real-world implementation not discussed	Potential side effects not mentioned
4	Pawar, Vikul J et al.'s [13]	2020	Effective implementation of a variety of methodologies and computer diagnostic tools.	For accuracy and progress to expand, more research is necessary.	Lung cancer diagnosis precision still has to be increased.
5	Shakeel, P. Mohamed et al. [14]	2019	Effectively lowered picture noise and enhanced robust clustering method for segmenting the wounded region.	Small sample size used for evaluation, scalability requires investigation.	X-ray or MRI images not included; only applies to CT scan images.
6	Subramanian, R. Raja, et al. [15]	2020	The model that was suggested can be used as a trustworthy and durable diagnostic tool.	Limited dataset size for model training and testing.	Lack of consideration for ethical and societal implications of AI-based lung cancer detection

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

7	Alagarsamy, Saravanan, et al. [18]	2021	Traditional approaches are slower and less accurate than the CSA algorithm.	Compared to cutting-edge techniques.	Could miss small tumours.
8	Xu, Yeguo et al. [21]	2022	Explains how to choose hyper parameters for approaches based on deep learning.	Interpretability of proposed method unclear	Relies solely on CT-scan images

3. THE ROLE OF DATASETS IN LUNG CANCER DETECTION

Despite cancer of the lung cannot be completely prevented, its risk can be reduced, according to Radhika, P. R., et al. (2019) [8]. As a result, enhancing patient lifespan requires early detection of lung cancer. The sheer number of cigarette smoking is directly negatively connected with the incidence of lung cancer. The prediction of lung cancer was studied using a variety of classification algorithms, including Naive Bayes (NB) [9], SVM [7], Decision trees (DT) [10], and Logistic Regression (LR) [11]. SVM exceeds all other classification techniques in terms of accuracy for these particular datasets.

The importance of correctly categorising lung cancer subtypes for efficient therapy and prognosis, as well as the difficulties of existing histology-based diagnosis techniques, are discussed in Cai, Zhihua, et al.'s (2015)[22] study. The study emphasises the value of molecular biomarkers for classification and reviews earlier research that identified certain microRNAs, genomic mutations, copy number variations, and gene expression profile data as possible biomarkers for lung cancer subtypes. The review talks about the shortcomings of earlier research that attempted to differentiate between NSCLC and SCLC. The review concludes by introducing the current work, which suggests an ensemble-based feature selection methodology to find a panel of 16 DNA methylation markers for identifying LADC, SQCLC, and SCLC.The study estimates accuracy of 86.54% and 84.6% in cross-validation with one leave-out and solitary data set test trials, respectively. It is suggested that these results may help with clinical diagnosis and treatment. The research on the effectiveness of machine learning methods for dividing cancer of the lung into categories based on levels of gene expression is reviewed in the article. The introduction gives a general overview of the high incidence and death rates of lung cancer across the world and emphasises how crucial it is to correctly identify the histological type of cancer for efficient therapy. The article describes how gene expression levels may be utilised as categorization features and how machine learning algorithms can be employed as support tools to analyse massive volumes of patient data for diagnosis.

The Matthews correlation coefficient (MCC) is a criterion for binary and non-binary classifications that Maxim D. Podolsky et al. (2016)[23] use to evaluate the effectiveness of the different algorithms for machine learning used in the evaluation. Four publicly available datasets from Brigham and Women's Hospital, University of Michigan, Dana-Farber Cancer Institute, and University of Toronto were studied for the study. In the findings section, a summary of the data sets and their associated algorithmic efficacy is provided. The support vector machine technique delivered the best outcomes for the Brigham and Women's Hospital and Dana-Farber Cancer Institute datasets. The University of Michigan data set showed maximum theoretical performance for all approaches, with one notable exception of the C4.5 decision tree (CDT). It showed the best results for the University of Toronto dataset. The discussion section emphasises the danger of false-positive or false-negative results of differentially expressed genes caused by the noise and dispersion of processed data. The work generally demonstrates the potential application of machine learning algorithms for tasks like classifying lung cancer morphology and various assessments of gene expression level, among others.

In Rabbani, Mohamad, et al. (2018) [24], a review of the current state of the literature on the application of ML in the treatment of non-small cell lung cancer (NSCLC) is provided. It draws attention to the difficulties doctors have when combining and analysing enormous volumes of data from numerous sources to create individualised treatment programmes for patients with NSCLC. Additionally, it highlights how machine learning algorithms may enhance workflow and patient outcomes all along the patient's journey. The paper then focuses on the particular use of machine learning algorithms for systemic medicines in NSCLC therapy selection and outcome prediction.The review includes citations from various papers that demonstrate how machine learning algorithms may be utilised to predict early mortality in NSCLC patients after

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

curative intent chemoradiation and identify characteristics that predict treatment outcomes. The review comes to its conclusion by noting that machine learning is quickly becoming a vital tool for doctors to better understand their patients and that, with continued integration of these tools into the clinical care continuum of NSCLC, patients will eventually receive a consistent, prompt, and personalised treatment plan.

A Weight Optimised Neural Network with Maximum Likelihood Boosting (WONN-MLB) strategy for reliable detection of Lung Cancer Disease (LCD) in large data is suggested by ALzubi, Jafar A., et al. (2019) [25]. To reduce classification time and increase diagnostic accuracy, the suggested method consists of two steps, namely feature selection and ensemble classification. Integrated Newton Raphson's Maximum Likelihood and Minimum Redundancy (MLMR) pre-processing model is used to choose the key features. The patient is classified using chosen attributes in the second step using the Boosted Weighted Optimised Neural Network Ensemble Classification method. Using the Thoracic Surgery Dataset, the suggested strategy is assessed and contrasted with more established approaches.The experimental findings show that, in comparison to traditional methodologies, the suggested strategy achieves a better false positive rate, forecast accuracy, and reduced latency. The proposed technique is successful in increasing the LCD diagnostic accuracy and reducing the false positive rate, and further testing with larger datasets is necessary, according to the paper's conclusion.

By using a hybrid classification approach that combines Support Vector Machine with Feed-Forward Back Propagation Neural Network, Nanglia, Pankaj, et al. (2021) [26] seeks to increase the accuracy of lung cancer detection. Kernel Attribute Selected Classifier is the name of the suggested method, which preprocesses the dataset, collects features using the SURF approach, optimises the features using genetic algorithms, and conducts classification using FFBPNN. A total classification accuracy of 98.08% is achieved using a batch of 500 photos, with 75% of the data utilised for training and 25% for classification.According to the literature study, SVM and NN are the two most widely utilised machine learning and neural network algorithms for diagnosing lung cancer. Pre-processing, feature extraction, and optimisation have all been the subject of several research, some of which have utilised hybrid feature extraction methods. Overall, the suggested hybrid algorithm for lung cancer detection yields encouraging results.

The serious issue of lung cancer, which is the main cause of cancer-related fatalities globally, is discussed by Anil Kumar, C., et al. (2022)[27]. Effective treatment of lung cancer depends on the early identification of its symptoms. The study suggests harnessing recent advancements in computational intelligence to build a durable prototype model for the treatment of lung cancer. The work optimises the detection procedure from the lung cancer dataset using a machine learning model based on support vector machines (SVMs). Using the Python programming language and the SVM classifier, lung cancer patients are categorised according to their symptoms. The University of California, Irvine library's cancer datasets are used to assess the SVM model's efficacy.The study's conclusions are encouraging, and the accuracy rate of the suggested model is compared to those of the SVM and SMOTE techniques already in use. The article also highlights further works that use supervised machine learning algorithms to forecast the development and progression of cancer using various input and data samples. The evaluation highlights the importance of the mutation test, which identifies the most effective patient treatments. The research shows that the suggested SVM model performs better than conventional classification techniques. Finally, the review emphasises how crucial it is to clean up the data and fill in any missing values before running it through analysis.

Humayun, Mamoona, et al. (2022)[28] provide a deep-learning-based approach to the detection and recognition of lung tumours using medical imaging. The study uses a deep neural network as an extraction of features technique in a computer-aided diagnosing (CAD) system to aid in the detection of lung ailments. The suggested model consists of three steps: localization, classification, and data augmentation. A transfer learning (TL) method is employed to train the classifier in order to address the problem of limited data in the evaluation of medical imagesThe suggested method offers a non-invasive diagnostic tool that is effective and has fewer parameters than the most modern models. The suggested design's effectiveness is assessed using established performance measures. On this dataset, VGG 16, VGG 19,[36] and Xception for the 20 epoch structure all exhibit excellent performance. Preprocessing is a wonderful bridge for creating a trustworthy model that ultimately helps in scenario forecasting because it incorporates the interface at a quicker stage for any model. The

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

accuracy of VGG 16, VGG 19, and Xception are, respectively, 98.83%, 98.05%, and 97.4% at the 20th epoch. The research indicates that the proposed model can be used to predict lung cancer in the future and that it delivers decent outcomes.

The authors mentioned above are primarily interested in various uses of machine learning algorithms for lung cancer detection, classification, and therapy. In addition to highlighting the difficulties encountered by doctors in integrating and analysing enormous volumes of patient data, the studies stress the significance of lung cancer early diagnosis in improving patient outcomes. The articles also discuss several machine learning methods for classifying and predicting lung cancer subtypes and treatment results, such as Naive Bayes, Support Vector Machines, Decision Trees, Logistic Regression, and Neural Networks.The articles also cover the possibility of biological biomarkers as diagnostic tools, including gene expression profiling, genetic alterations, and microRNAs, as well as the significance of feature selection in enhancing classification accuracy. In general, the authors show how machine learning algorithms might be used to enhance lung cancer patient outcomes, diagnosis, and care. In the publications mentioned above, numerous methods are discussed for increasing the precision of lung cancer classification and diagnosis utilising biomarkers and machine learning algorithms. To find more efficient biomarkers and increase the precision of classification algorithms, research is still required.The difficulties of processing vast volumes of patient data, particularly noisy and dispersed data must also be addressed. For better patient outcomes, future research may concentrate on creating algorithms that are more reliable and effective in processing and analysing patient data as well as incorporating machine learning techniques into the clinical care continuum. Another crucial area for future study is the requirement for more thorough evaluation of these strategies' clinical efficacy utilising bigger and more varied datasets.

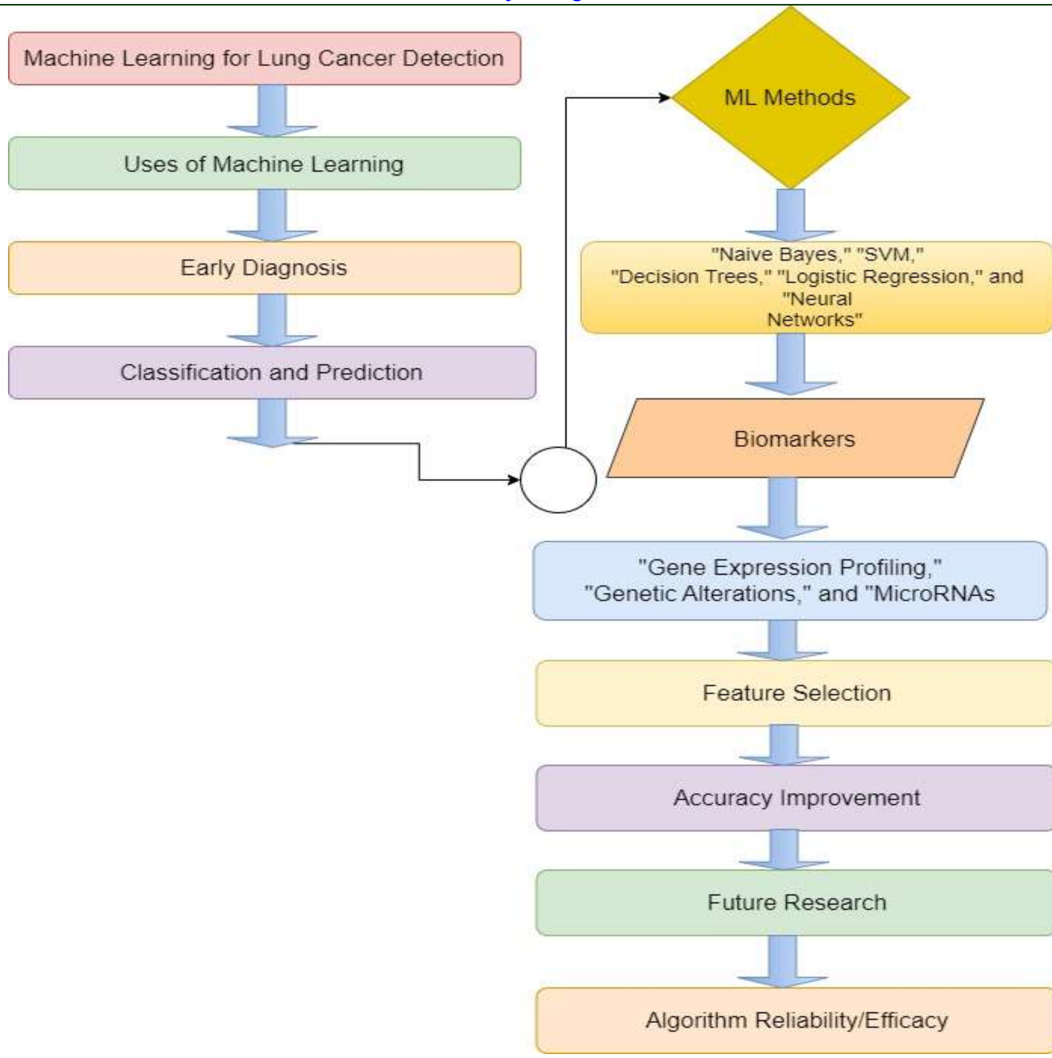

Figure5: Process Flow Diagram for Enhancing Lung Cancer Patient Outcomes with Machine Learning using Dataset.

Table2: Comparative Analysis of Advantages, Research Gaps and Disadvantages in Lung Cancer Detection Studies with Dataset

Sno.	Author Name	Year	Advantages	Research Gap	Disadvantages
1.	Radhika, P. R., et al. [8]	2019	The best prediction accuracy for lung cancer is shown using SVM.	Clearly outlining restrictions and recommending fresh ideas.	Restricted dataset, characteristics, and generalizability. No original strategy
2	Cai, Zhihua, et al.'s [22]	2015	Lung cancer subtype classification is now more accurate because to molecular biomarkers.	The classification of lung cancer subtypes requires comparison with other approaches.	Minimal sample size; insufficient information regarding demographics.
3	Podolsky, Maxim D., et al. [23]	2016	The best results were obtained using the support vector machine technique.	Limited study area; analysis has to be expanded.	Dependence on publicly accessible datasets, lack of patient data, and absence of performance evaluations against that of human experts.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4	Rabbani, Mohamad, et al. [24]	2018	Individualised treatment plans for patients can be made using ML.	Clinical assessment and cutting-edge imaging techniques for NSCLC identification.	Data integrity, clarity, and possibility for bias in medical judgement.
5	ALzubi, Jafar A., et al.[25]	2019	Reduces false positive rate	Testing with a larger dataset, interpretability of the ensemble technique.	Minimal applicability to other medical disciplines due to computational complexity.
6	Nanglia, Pankaj, et al.[26]	2021	Can possibly lead to better results for patients by giving more precise diagnoses.	In compared with modern algorithms.	Huge training set is necessary, although only a few photos are used.
7	Anil Kumar, C., et al.[27]	2022	The suggested model's accuracy rate is high.	Existing difficulties and deficiencies in lung cancer detection.	Small dataset, no computational information, symptom-based classification.
8	Humayun, Mamoona, et al.[28]	2022	High precision was attained in the 20th period using VGG 16, VGG 19, and Xception.	Examination of several datasets with demographic data included.	Dependence on CT images, massive training data demand, and absence of clinical data integration.

4. DISCUSSIONS

The ability of deep learning algorithms to improve lung cancer diagnosis has been demonstrated. Before these techniques are extensively used in clinical practise, there are a number of restrictions and research gaps that must be filled. The need for more varied datasets that can represent the whole range of illness presentation, including changes in anatomy, physiology, and demography, is one of the key difficulties.

To improve the accuracy of the models, it is also crucial to include additional medical data, such as patient history, clinical test findings, and genetic information.The assessment of robustness to noise and artefacts, which can have an impact on how well deep learning algorithms work, is an additional crucial factor. Because it can be challenging to comprehend how the algorithms decide what to do, these models' lack of interpretability is also a serious concern. The implementation of these techniques in clinical practise may be hampered as a result, which can make it difficult to establish trust between healthcare providers and patients.

The use of machine learning in medical decision-making has ethical and legal consequences as well. For instance, problems about algorithm bias, informed consent, and data privacy may exist. To make sure that the use of these techniques is fair and safe for all patients, these problems must be resolved.

The experiments covered in this article offer important insights into the potential of machine learning for cancer detection and treatment despite these drawbacks. The goal of future research should be to overcome these constraints and create machine learning methods for lung and breast cancer diagnosis and treatment that are more precise and efficient. This may entail the addition of more sophisticated imaging characteristics, validation in separate cohorts, and assessment in actual clinical settings. The ethical and legal ramifications of employing machine learning models for medical decision-making must also be taken into account if these technologies are to be used responsibly.

This survey study reviews works on VGG16, SSD, SVM, CNN, CSA, pre-processing operations, and the effects of CAD systems to address the diagnosis of lung cancer using medical imaging and deep learning approaches.

Evaluated studies on deep learning with medical imaging for lung cancer diagnosis. Relevance, approach quality, outcome accuracy, and usefulness to enhance diagnosis were taken into account.

Clear aims, the inclusion of pertinent information on problems and datasets, a well-structured organization, and a thorough discussion of various studies on lung cancer diagnosis using medical imaging and deep learning are some of the qualities of the abstract. The shortcomings, however, include a lack of particular references, scant information on deep learning models, broad assertions about constraints, a lack of a concluding statement, and some content repetition. The quality of the abstract might be improved by giving more precise citations, going into greater detail about model properties, addressing limits with examples, and making a concluding statement.

5. CONCLUSION AND FUTURE SCOPE

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

In this work, we reviewed numerous studies on deep learning and imaging for lung cancer detection and diagnosis. There is still a need to increase the overall accuracy of algorithms such as the VGG16, SSD, SVM, CNN, and CSA, which have demonstrated promising results in properly diagnosing lung cancer from CT scans. The visual appeal and diagnostic effectiveness of the input photo are greatly improved by pre-processing operations such edge identification, picture resampling, and segmentation. Diagnostic classifiers and pre-processing techniques could be considerably improved by using computer-aided diagnosis (CAD) systems.

When comparing the results of our research to those of earlier studies, we found that machine learning algorithms and meta-heuristic optimisation techniques have been effective in categorising the morphology of lung cancer, choosing the best treatment options, and forecasting the results of systemic therapy for non-small cell lung cancer. We do admit, though, that a larger dataset is required to confirm the accuracy of our findings. The growing use of machine learning by doctors to better understand their patients and create individualized treatment plans, in our opinion, illustrates the potential benefits of these methods in clinical practice.

Despite these limitations, the research demonstrated how machine learning and deep learning models might be used to identify and classify lung cancer. By resolving problems with data diversity, feature selection, interpretability, and ethical considerations, future studies can increase the accuracy and reliability of lung cancer diagnosis. A comprehensive CAD system including cutting-edge image processing methods and molecular biomarkers may be a powerful tool for identifying lung cancer in its earliest stages and planning a customized course of therapy.

Our analysis concludes by highlighting the advancements made in the use of machine learning and medical imaging tools for lung cancer screening. However, to fully realize the potential of these techniques in enhancing lung cancer diagnostic and treatment results, careful assessment and additional study are required. We envision a future in which machine learning plays a critical role in revolutionizing lung cancer management and having a significant impact on patient outcomes. This will be possible with ongoing breakthroughs and wider clinical validation.

To improve the accuracy and reliability of lung cancer diagnosis, the feature scope can include creating new machine learning algorithms and deep learning models, as well as better pre-processing techniques such image denoising, edge recognition, lung image resampling, and segmentation. By gathering a larger sample size and high-quality data for the algorithms' training and validation, it may also concentrate on building larger datasets. Creating a computer-aided diagnosis (CAD) system to help physicians identify and categorize lung cancer from medical imaging is another feature scope. Machine learning algorithms can also be used to investigate the potential of molecular biomarkers for classification and the best course of lung cancer treatment.

REFERENCES:

[1] Patra, Radhanath. "Prediction of lung cancer using machine learning classifier." Computing Science Communication and Security: First International Conference, COMS2 2020, Gujarat, India, March 26–27, 2020, Revised Selected Papers 1. Springer Singapore, 2020.

[2] Patra, R. (2020). Prediction of Lung Cancer Using Machine Learning Classifier. In: Chaubey, N., Parikh, S., Amin, K. (eds) Computing Science, Communication and Security. COMS2 2020. Communications in Computer and Information Science, vol 1235. Springer,Singapore.https://doi.org/10.1007/978-981-15-6648-6_11

[3] Khamparia, Aditya, et al. "Diagnosis of Breast Cancer Based on Modern Mammography Using Hybrid Transfer Learning." Multidimensional Systems and Signal Processing, vol. 32, no. 2, Springer Science and Business Media LLC, Jan. 2021, pp. 747–65. Crossref, https://doi.org/10.1007/s11045-020-00756-7.

[4] Pramanik, Payel, et al. "Deep Feature Selection Using Local Search Embedded Social Ski-Driver Optimization Algorithm for Breast Cancer Detection in Mammograms." *Neural Computing and Applications*, vol. 35, no. 7, 2022, pp. 5479–5499., https://doi.org/10.1007/s00521-022-07895-x.

[5] Chang, Yi-Chun, et al. "A learning style classification mechanism for e-learning." Computers & Education 53.2 (2009): 273-285.

[6] Rahane, Wasudeo, et al. "Lung Cancer Detection Using Image Processing and Machine Learning HealthCare." *2018 International Conference on Current Trends Towards Converging Technologies (ICCTCT)*,

IEEE, Mar. 2018. *Crossref*, https://doi.org/10.1109/icctct.2018.8551008.

[7] Huang, Shaobo and Ning Fang. “Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models.” Comput. Educ. 61 (2013): 133-145

[8] Radhika, P. R., et al. “A Comparative Study of Lung Cancer Detection Using Machine Learning Algorithms.” IEEE International Conference on Electrical, Computer and Communication Technologies, 2019, https://doi.org/10.1109/icecct.2019.8869001.

[9] Putra, RyandaSatria, et al. "The Application of Naïve Bayes Classifier Based Feature Selection on Analysis of Online Learning Sentiment in Online Media." JurnalTransformatika 20.1 (2022): 44-56.

[10] Rizvi, Saman, Bart Rienties, and Shakeel Ahmed Khoja. "The role of demographics in online learning; A decision tree based approach." Computers & Education 137 (2019): 32-47.

[11] Hazra, Animesh, NanigopalBera, and AvijitMandal. "Predicting lung cancer survivability using SVM and Logistic Regression Algorithms." International Journal of Computer Applications 174.2 (2017): 19-24.

[12] D. Palani and K. Venkatalakshmi, “An IoT based predictive modelling for predicting lung cancer using fuzzy cluster based segmentation and classification,” Journal of Medical Systems, vol. 43, no. 2, p. 21, 2019.

[13] Pawar, Vikul J. “Lung Cancer Detection System Using Image Processing and Machine Learning Techniques.” International Journal of Advanced Trends in Computer Science and Engineering, vol. 9, no. 4, The World Academy of Research in Science and Engineering, Aug. 2020, pp. 5956–63. https://doi.org/10.30534/ijatcse/2020/260942020.

[14] Shakeel, P. Mohamed, MohdAboobaider Burhanuddin, and Mohamad Ishak Desa. "Lung cancer detection from CT image using improved profuse clustering and deep learning instantaneously trained neural networks." Measurement 145 (2019): 702-712.

[15] Subramanian, R. Raja, et al. "Lung Cancer Prediction Using Deep Learning Framework." International Journal of Control and Automation 13.3 (2020): 154-160.

[16] Srinivasulu, Asadi, et al. "Advanced lung cancer prediction based on blockchain material using extended CNN." Applied Nanoscience (2021): 1-13.

[17] Agarwal, Aman, KritikPatni, and D. Rajeswari. "Lung cancer detection and classification based on alexnet CNN." 2021 6th International Conference on Communication and Electronics Systems (ICCES). IEEE, 2021.

[18] Alagarsamy, Saravanan, et al. "Prediction of lung cancer using meta-heuristic based optimization technique: Crow search technique." 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS). IEEE, 2021.

[19] Surrender, P. "Diagnosis of lung cancer using hybrid deep neural network with adaptive sine cosine crow search algorithm." Journal of Computational Science 53 (2021): 101374.

[20] Mohamed, TehnanIbrahemAlhasseen. Optimized deep learning model for early detection and classification of lung cancer on CT images. Diss. 2022.

[21] Xu, Yeguo, Yuhang Wang, and NavidRazmjooy. "Lung cancer diagnosis in CT images based on Alexnet optimized by modified Bowerbird optimization algorithm." Biomedical Signal Processing and Control 77 (2022): 103791.

[22] Cai, Zhihua, et al. "Classification of lung cancer using ensemble-based feature selection and machine learning methods." Molecular BioSystems 11.3 (2015): 791-800.

[23] Podolsky, Maxim D., et al. "Evaluation of machine learning algorithm utilization for lung cancer classification based on gene expression levels." Asian Pacific journal of cancer prevention 17.2 (2016): 835-838.

[24] Rabbani, Mohamad, et al. "Role of artificial intelligence in the care of patients with nonsmall cell lung cancer." European journal of clinical investigation 48.4 (2018): e12901.

[25] ALzubi, Jafar A., et al. "Boosted neural network ensemble classification for lung cancer disease diagnosis." Applied Soft Computing 80 (2019): 579-591.

[26] Nanglia, Pankaj, et al. "A hybrid algorithm for lung cancer classification using SVM and Neural Networks." ICT Express 7.3 (2021): 335-341.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[27] Anil Kumar, C., et al. "Lung cancer prediction from text datasets using machine learning." BioMed Research International 2022 (2022).

[28] Humayun, Mamoona, et al. "A transfer learning approach with a convolutional neural network for the classification of lung carcinoma." Healthcare. Vol. 10. No. 6. MDPI, 2022.

[29] Pang, Shanchen, et al. "VGG16-T: a novel deep convolutional neural network with boosting to identify pathological type of lung cancer in early stage by CT images." International Journal of Computational Intelligence Systems 13.1 (2020): 771.

[30] Chatterjee, Bitanu, et al. "Late acceptance hill climbing based social ski driver algorithm for feature selection." IEEE Access 8 (2020): 75393-75408.

[31] Han, Yanfang, and Pengfei Shi. "An improved ant colony algorithm for fuzzy clustering in image segmentation." Neurocomputing 70.4-6 (2007): 665-671.

[32] Zio, Enrico. "Prognostics and Health Management (PHM): Where are we and where do we (need to) go in theory and practice." Reliability Engineering & System Safety 218 (2022): 108119.

[33] Wong, Wai-Tak, and Sheng-Hsun Hsu. "Application of SVM and ANN for image retrieval." European Journal of Operational Research 173.3 (2006): 938-950.

[34] Perumal, S., and ThambusamyVelmurugan. "Preprocessing by contrast enhancement techniques for medical images." International Journal of Pure and Applied Mathematics 118.18 (2018): 3681-3688.

[35] Abdulmajeed, Ashraf Abdulmunim, and Nada Nimat Saleem. "Pre-processing and enhancement techniques for COVID-19 X-ray and CT-scan medical images." AIP Conference Proceedings. Vol. 2394. No. 1. AIP Publishing LLC, 2022.

[36] Junaidi, Apri, et al. "Image classification for EGG incubator using transfer learning of VGG16 and VGG19." 2021 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT). IEEE, 2021.