International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 06 | June - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Comprehensive and Comparative Framework for Lung Cancer Detection Using CNN

Dr.R.SATISH, Associate professor, ECE Department, Sir C.R.Reddy College of Engineering - Eluru ,India-534007,Email:satishus@gmail.com.
L.NANDINI, ECE Department, Sir C.R.Reddy College of Engineering.
M.PRASANNA KUMAR, ECE Department, Sir C.R.Reddy College of Engineering
K.RAMA SUDHA, ECE Department, Sir C.R.Reddy College of Engineering.
G.CHINMAYEE NAGA SAI, ECE Department, Sir C.R.Reddy College of Engineering.

Abstract

Lung cancer is one of the most lethal cancer types; thousands of peoples are infected with this type of cancer, and if they do not discover it in the early stages of the disease, then the chance of surviving of the patient will be very poor. For the suggested reasons above and to help in overcoming this terrible, early diagnosis with the assistance of artificial intelligence procedures most needed. Through this research, a Computer-aided system introduced for detecting lung cancer in a dataset by using a convolutional neural network technique for helping with the diagnosis of the patient's cases: normal or abnormal. The CNN with PCA approach as a promising solution to address these challenges. By combining the strengths of CNNs for image analysis and PCA for feature extraction, the project seeks to develop a robust and efficient lung cancer detection system. The proposed method is compared with the baseline method (SVM). The method Support Vector Machine achieved accuracy upto 75.0%. Finally The results also show that convolutional neural networks(CNN) outperforms the other Machine Learning algorithms. The proposed model gives high accuracy upto 98.548%.

1.0 INTRODUCTION

Lung cancer remains one of the most prevalent and deadliest forms of cancer worldwide, presenting a significant public health challenge. Timely and accurate detection is crucial for improving patient outcomes, as early-stage diagnosis drastically enhances treatment effectiveness and survival rates. In recent years, the advent of deep learning techniques has revolutionized medical image analysis, offering promising avenues for enhancing lung cancer detection.

This introduction sets the stage for proposing a comprehensive and competitive framework for lung cancer detection, leveraging the power of deep learning methodologies. By integrating state-of-the-art algorithms, data preprocessing techniques, and innovative model architectures, this framework aims to significantly advance the accuracy, efficiency, and scalability of lung cancer detection systems.

There are 2123 people who have lung cancer from two genders. This number represents about 8.31% of the total infections in the country. This portion indicates a small increase as compared to the ratio of the past year, which is about 8.1%. The rate of lung cancer represents approximately 13.27% of the total cancer cases, and this shows that lung cancer is the leading type in males. Also, there is a rise when compares to the ratio recorded in 2015, which approximately reaches 12.7%**.** For females, lung cancer not the leading cancer type, it ranked fifth between other cancer types, in 2016 there is only 638 woman who has this disease, there are 638 women who diagnosed with lung cancer at 2016, this represents about 4.44% from the total cancer types infections.

Artificial intelligence not only used in the area of lung cancer diagnosing, but it applied in all fields of biomedical engineering such as: diagnosis of breast cancer ,diagnosis of Heart disease, also diagnosing and classification of diabetes. Various methods are convenient for diagnosing lung cancer, particularly MRI, isotope, Xray, and CT. X-ray chest radiography and Computer Tomography (CT) are the two well-known imaging modalities that are commonly utilized in the identification of different lung diseases.

The aim of our study is to implement a CAD system used as an assistant to doctors while deciding and diagnosing lung cancer, this system used for detecting and classifying the lung cancer cases if it normal, benign, or malignant with high accuracy. This done by applying convolutional neural network technique to a data set of lung cancer CT scans collected and diagnosed at the Iraqi hospitals.

.

1.0 LITERATURE REVIEW

After doing lots of literature works in related area for selection of proposed work. after going through literature from books, research papers and standard websites we come up with conclusion that available methods are good enough but limitation regarding accuracy and performance is not efficient for considered CT images. Available methods are quite good so we did not make any changes in the methods, we have improvised the accuracy and performance of an image using Learning methods with deep learning techniques by using convolutional neural network algorithm.

Welch HG, Schwartz LM, Woloshin S. in [1],

The association between changes in 5-year survival and changes in mortality and incidence measured using simple correlation coefficients (Pearson and Spearman).From 1950 to 1995, there was an increase in 5-year survival for each of the 20 tumor types. The absolute increase in 5-year survival ranged from 3% (pancreatic cancer) to 50% (prostate cancer). During the same period, mortality rates declined for 12 types of cancer and increased for the remaining 8 types. There was little correlation between the change in 5-year survival for a specific tumor and the change in tumor-related mortality (Pearson r=.00; Spearman r=-.07). On the other hand, the change in 5-year survival was positively correlatedwiththechange in the tumor incidence rate (Pearson r=+. 49; Spearman r=+.37). Although 5-year survival is a valid measure for comparing cancer therapies in a randomized trial, our analysis shows that changes in 5-year survival over time bear little relationship to changes in cancer mortality.

AsunthaA, SrinivasanA in [2],

The main objective of this work is to detect the cancerous lung nodules from the given input lung image and to classify the lung cancer and its severity. To detect the location of the cancerous lung nodules, this work uses novel Deep learning methods. This work uses best feature extraction techniques such as Histogram of oriented Gradients (HOG), wavelet transform-based features, Local Binary Pattern (LBP), Scale Invariant Feature Transform (SIFT) and Zernike Moment. After extracting texture, geometric, volumetric and intensity features, Fuzzy Particle Swarm Optimization (FPSO) algorithm is applied for selecting the best feature. Finally, these features are classified using Deep learning. A novel FPSOCNN reduces computational complexity of CNN. An additional valuation is performed on another dataset coming from Arthi Scan Hospital which is a real-time data set. From the experimental results, it is shown that novel FPSOCNN performs better than other techniques

Nie L, Zhang L, Yang Y, Wang M, Hong R, Chua T-S in [3],
A large and growing body of literature has investigated the disease progression problem. However, far too little attention to date has been paid to jointly consider the following three observations of the chronic disease progression: 1) the health statuses at different time points are chronologically similar; 2) the future health statuses of each patient can be comprehensively revealed from the current multimedia and multimodal observations, such as visual scans, digital measurements and textual medical

histories; and 3) the discriminative capabilities of different modalities vary significantly in accordance to specific diseases. In the light of these, we propose an adaptive multimodal multi-task learning model to co-regularize the modality agreement, temporal progression and discriminative capabilities of different modalities.

Siegel, Rebecca, Naishadham, Deepa, Jemal, Ahmedin in [5],

During the most recent 5 years for which there are data (2005-2009), delay-adjusted cancer incidence rates declined slightly in men (by 0.6% per year) and were stable in women, while cancer death rates decreased by 1.8% per year in men and by 1.5% per year in women. Overall, cancer death rates have declined 20% from their peak in 1991 (215.1 per 100,000 population) to 2009 (173.1 per 100,000 population). Death rates continue to decline forall 4 major cancer sites (lung, colorectum, breast, and prostate). Over the past 10 years of data (2000-2009), the largest annual declines indeathrateswere forchronicmyeloidleukemia(8.4%), cancersofthe stomach (3.1%) and colorectum (3.0%), and non- Hodgkin lymphoma (3.0%). The reduction in overall cancer death rates since 1990 in men and 1991 in women translates to the avoidance of approximately 1.18 million deaths from cancer, with 152,900 of these deaths averted in 2009 alone.

Republic of Iraq , Ministry of Health\Environment, Board. IC. In [6,7],

"These alarming figures highlight the need for WHO, the Ministry of Health and all partners to address cancer management as a key public health priority programme in the country. The cancer registration data for 2017–2018 provides essential data that will contribute to the development of effective prevention and treatment plans for the disease," said Dr Adham Ismail, WHO Representative in Iraq.To ensure accurate acquisition and management of data at both the central and governorate levels in Iraq, WHO recently trained seven master trainers and staff working in cancer institutions on the updated CanReg 5 program, an open-source tool to input, store, check and analyse cancer registry data.

Nasser, Ibrahim MAbu-Naser, S. S in[8],

In thispaper, we developed an Artificial Neural Network(ANN) fordetect theabsenceorpresence of lung cancer in human body. Symptoms were used to diagnose the lung cancer, these symptoms such as Yellow fingers, Anxiety, Chronic Disease, Fatigue, Allergy, Wheezing, Coughing, Shortness of Breath, Swallowing Difficulty and Chest pain. They were used and other information about the person as input variables for our ANN. Our ANN established, trained, and validated using data set, which its title is "survey lung cancer". Model evaluation showed that the ANN model is able to detect the absence or presence of lung cancer with 96.67 % accuracy.

Taher, FatmaSammouda, Rachid in [9],

The early detection of lung cancer is a challenging problem, due to the structure of the cancer cells, where most of the cells are overlapped with each other. This paper presents two segmentation methods, Hopfield Neural Network (HNN) and a Fuzzy C-Mean (FCM) clustering algorithm, for segmenting sputum color images to detect the lung cancer in its early stages. The thresholding algorithm succeeded in extracting the nuclei and cytoplasm regions. Moreover, it succeeded in determining the best range of thresholding values. In this study, we used 1000 sputum color images to test both methods, and HNN has shown a better classification result than FCM, the HNN succeeded in extracting the nuclei and cytoplasm regions.

Eskandarian P,BagherzadehJ in [10],

Lung cancer is one of the most lethal cancer types; thousands of peoples are infected with this type of cancer, and if they do not discover it in the early stages of the disease, then the chance of surviving of the patient will be very poor. . The proposed model gives high accuracy ups to 93.548%. The other performance metrics comes with high values such as 95.714% for sensitivity and 95% for Specificity.

2.0 METHODOLOGY

The methodology of working of our project involves in use of deep learning algorithms and mainly the CNN is used to different purposes like detection in our project and Classifiers are used to predict the type of cancer based on the manual inputs given by the user and the CNN used in our project extracts useful information and analyzes the CT images to detect the presence of any type of cancer. we have followed the below steps for working. Data Gathering and Preprocessing we have taken datasets for prediction and for detection. we have taken the CT dataset from Kaggle which consists of 138 CT images.

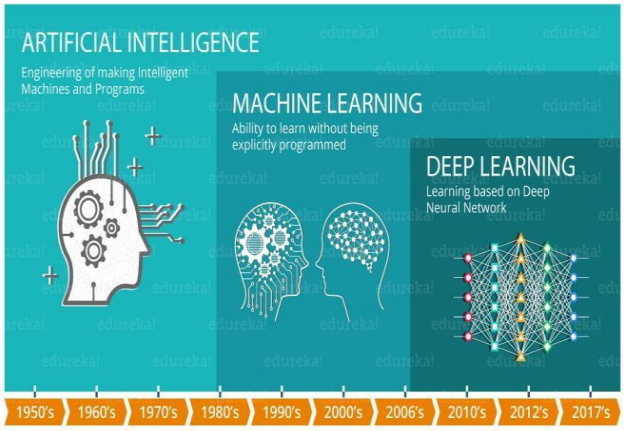

Fig 1. Deep Learning structure

A dataset is given to the deep learning technique. The dataset contains more than 100 images. It contains normal and abnormal lung disease images. With these images we will train the algorithm and test the algorithm.After loading the dataset we will split the data set into two parts one is for training the algorithm another one is testing the algorithm.SVM(support vector machine) algorithm is used to train the data. It will train and give accuracy of the algorithm lesser than CNN (Convolutional nueral network).CNN(Convolutional neural network) is used to train with the data. It will give more accuracy than machine learning algorithms. Approximately 90% above.In this module we will give input as the image to the algorithm. The algorithm will preduict either it is normal or abnormal.We will show accuracies of the algorithms in the form of graph.

Data collection and preprocessing:

The first step is to load the data set that will be used to train the algorithm. This involves collecting a dataset of CT scans. These scans may include images of healthy lungs as well as lungs affected by disease such as cancer. The images are typically pre-processed to standardize their resolution, orientation, and contrast, ensuring consistency across the dataset.This data set should contain examples of the type of data that the algorithm will be used to classify or predict.The next step is to pre process the data. This may involve cleaning the data, removing outliers, and transforming the data into a format that is compatible with the algorithm.

Train data:

The pre processed data is then split into two sets: a train set and a test set. The train set is used to train the algorithm, and the test set is used to evaluate the performance of the trained algorithm. Python provides a rich ecosystem of deep learning libraries, such as TensorFlow, Keras, PyTorch, and scikit-learn, that make training deep learning models accessible and efficient.

Train algorithm:

The train set is then used to train the algorithm. This involves adjusting the parameters of the algorithm so that it can accurately classify or predict the data in the train set. Train the selected model using the training dataset. During training, the model learns to map the input grayscale features to the corresponding lung cancer labels.

Testing

After training, the performance of the CNN is evaluated using a dataset. This dataset contains CT scans that were not seen by the model during training. The model's predictions are compared against the ground truth labels to assess its accuracy. Additionally, the model may be tested on an independent test set to further evaluate its generalization capability.

Predict results:

The trained algorithm is then used to predict the results for the data in the test set. The performance of the algorithm is then evaluated by comparing the predicted results to the actual results in the test set.

Algorithm:

1) Convolutional Nueral Network:

Convolutional Neural Networks (CNNs) play a vital role in the detection and classification of lung cancer from medical images such as X-rays, CT scans, and MRI scans. Here's how CNNs are typically utilized in the lung cancer detection process. The Convolutional Neural Network (CNN) plays a pivotal role in the proposed method for detecting lung cancer in CT images. Leveraging a densely connected lassifying lung images as either normal or abnormal. This approach is crucial for early detection, significantly impacting survival rates for individuals with lung cancer. The utilization of a dataset comprising 138 lung images, with a division for training and testing, underscores the robustness of the model. The experimental results validate the efficacy of the CNN-based system, demonstrating commendable accuracy. The integration of deep learning techniques, specifically the MobileNet model, enhances

the model's ability to discern intricate patterns in CT images associated with lung pathology. In the realm of medicinal image processing, this CNN-driven method stands as a promising avenue for advancing early detection of lung cancer, contributing to improved prognoses for affected individuals.

4.0 RESULTS AND DISCUSSION

In our project, we are classifying the Lung Image Classification, with the help of deep learning. Lung diseases are indeed the lung-affecting diseases which impair the respiratory mechanism. Lung cancer has been one of the leading causes of mortality in humans worldwide. Early detection can enhance survival chances amid humans. If the condition is diagnosed in time. the average survival rates for people with lung cancer rise from 14 to 49 percent While computed tomography (CT) is far more effective than X-ray, a thorough diagnosis includes multiple imaging approaches to support each other.For the classification of the lung image as normal or malignant, a densely connected convolution neural network (CNN) algorithm was used.

In fig 4.1 On this screen, when the user submitted an image of a cancer image, the prediction also identified it as a abnormal.

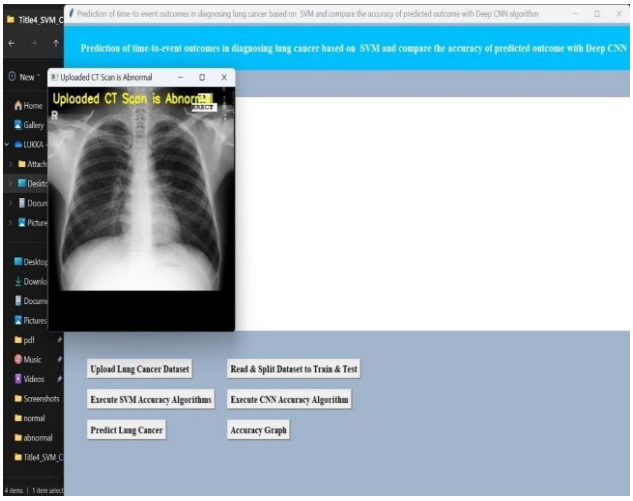

Fig 4.1 Output window

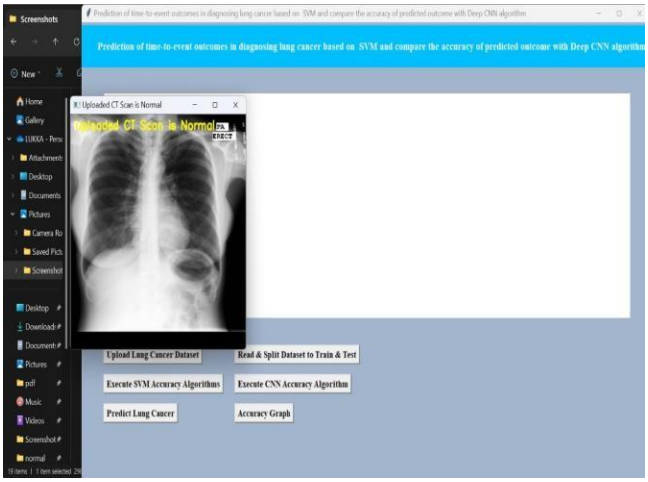

Fig 4.2. Output window

In fig 4.2 On this screen, when the user submitted an image of a normal image, the prediction also identified it as a normal.

GRAPHICAL REPRESENTATION:

Graphical representations are used to calculate the accuracy of CT scan images for detecting lung cancer because visualizing the data helps us analyze and interpret the results more effectively. Graphical representations, such as charts or graphs, allow us to see patterns, trends, and anomalies in the data. By visually comparing the CT scan images of patients with and without lung cancer, we can identify specific characteristics or abnormalities that may indicate the presence of cancerous cells.

Fig 6. shows the graptical representation of the accuracy between the learning methods CNN and SVM where x-axis represents the classifiers and y-axis represents the accuracy.

ACCURACY COMPARISION

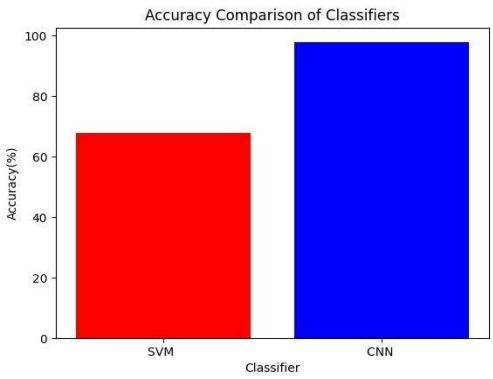

Fig 4.3 Accuracy comparision of CNN and SVM

QUANTITATIVE ANALYSIS

We use quantitative analysis in lung cancer detection using deep learning with CNN (Convolutional Neural Networks) because it allows us to extract meaningful numerical features from CT scan images. CNNs are particularly effective in analyzing medical images because they can automatically learn and identify patterns, textures, and structures that may be indicative of lung cancer. By applying quantitative analysis techniques, we can measure and quantify various characteristics of the CT scan images, such as the size, shape, density, and texture of potential cancerous lesions. These numerical features can then be used as inputs to the CNN model, which learns to classify the images as either cancerous or non-cancerous. Quantitative analysis provides a more objective and standardized approach to lung cancer detection, as it removes subjective interpretations and biases that may occur with manual analysis.

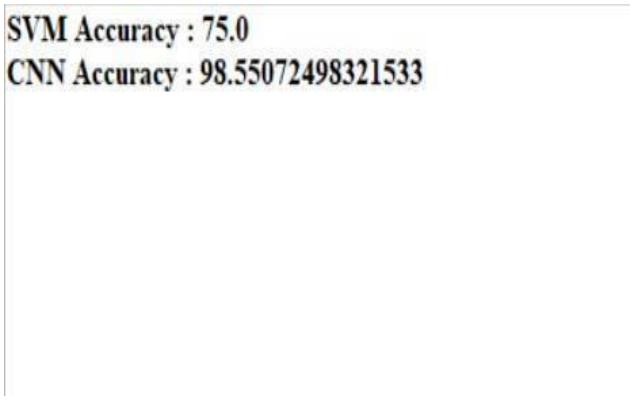

Fig 6.5 Accuracy of CNN and SVM

5.0 CONCLUSION:

This proposed study tries to defeat the problems faced in the early detection of lung cancer nodules before it gets worst. For this purpose, this study develops an effective computer-aided diagnosis scheme for early detecting of this lethal cancer. Chest tomography scans have been employed here as data input to the proposed model. This study's goal was to improve a CNN deep learning model able to detecting and classifying lung cancer nodules successfully. The obtained model gives high accuracy reaches 93.548% while applying on the dataset collected. The use of quantitative analysis techniques has allowed for the extraction of meaningful numerical features from the images, enabling the CNN model to classify them as cancerous or non-cancerous. This objective and standardized approach removes subjectivity and biases that may occur with manual analysis, leading to more reliable and consistent results.

6.0 REFERENCES:

[1] Welch HG, Schwartz LM, Woloshin S. Are increasing 5-year survival rates evidence of success against cancer? Jama. 2000;283(22):2975-8.

[2] Asuntha A, Srinivasan A. Deep learning for lung Cancer detection and classification Multimedia Tools Applications. 2020:1-32.

[3] Nie L, Zhang L, Yang Y, Wang M, Hong R, Chua T-S, editors. 2015, Beyond doctors: Future health prediction from multimedia and multimodal observations. Proceedings of the 23rd ACM international conference on Multimedia;

[4] American Lung Association, URL www.lung.org.

[5] Siegel, Rebecca, Naishadham, Deepa, Jemal, Ahmedin. Cancer statistics, 2013. a cancer journal for clinicians. 2013;63(1):11-30.

[6] Republic of Iraq, 2016, Ministry of Health\Environment, Board. IC. Annual Report Iraqi Cancer Registry 2016.

[7] Republic of Iraq, Ministry of Health\Environment, 2015., Board. IC. Annual Report Iraqi Cancer Registry 2015.

[8] Nasser, Ibrahim M Abu-Naser, S. S. 2019, Lung Cancer Detection Using Artificial Neural Network.

International Journal of Engineering Information

[9] Taher, Fatma Sammouda, Rachid., editors. 2011, Lung cancer detection by using artificial neural network and fuzzy clustering methods. 2011 IEEE GCC Conference and Exhibition (GCC);: IEEE.

[10] Eskandarian P, Bagherzadeh J, editors. 2015, Computer-aided detection of Pulmonary Nodules based on SVM in thoracic CT images. 7th Conference on Information and Knowledge Technology (IKT); 2015: IEEE

[11] Ganesan S, Subashini T, Jayalakshmi K, editors. 2014, Classification of X-rays using statistical moments and SVM. 2014 International Conference on Communication and Signal Processing;: IEEE.

[12] Parveen SS, Kavitha CJIJoCA. 2014, Classification of lung cancer nodules using SVM Kernels.;95(25).
[13]Thamilselvan P, Sathiaseelan J. 2016, Detection and classification of lung cancer MRI images by using enhanced k nearest neighbor algorithm. Indian Journal of Science Technology.;9(43):1-7.
[14] Kurkure M, Thakare A, editors. 2016, Lung cancer detection using genetic approach 2016 International Conference on Computing Communication Control and automation .
[15] H. F. Maqbool, Hafiz F. Maqbool, H. Farhan MaqboolFuzzy Rule Based Diagnostic System to Detect the Lung Cance
[16] P. Chitra; K. Srilatha; F.V. Jayasudha; P. Brindha; K. Vijaya; S. Sneha 2023, Lung Cancer Detection Using Classification .
[17] T. Aggarwal, A. Furqan and K. Chest CT Kalra, "Feature extraction and LDA based classification of lung nodules in scan images", *2015 International Conference On Advances In Computing. Communications And Informatics, 2015.*
[18] G. Xiuhua, S. Tao and Zhigang, "Prediction Models for Malignant Pulmonary Nodules Based-on Texture Features of CT Image", *CT imaging and analysis theory and applications, 2011*.
[19] Kaori Togashi,Richard Kinh Gian Do,Mizuho Nishio,Rikiya Yamashita, Convolutional neural networks: an overview and application in radiology, IEEE article.
[20] Mohammad Mustafa Taye , Theoretical Understanding of Convolutional Neural Network: Concepts, Architectures, Applications, Future Directions.
[21] shaza Dawood Ahmed Rihan,Mohammed Anbar,Basim Ahmad Alabsi, CNN :Dual Convolutional Neural Network Approach for Feature Selection and Attack Detection on Internet of Things Networks.
[22] Enam A. Khalil,Furat Y. Mohsen,Muayed S Al-Huseiny,Hamdalla F. Al-Yasriy, Diagnosis of Lung Cancer Based on CT Scans Using CNN.
[23] Aditi Kajala,V K Jain, Diagnosis of Breast Cancer using Machine Learning Algorithms
[24] Mamatha Sai Yarabarla,Lakshmi Kavya ,RaviA. Sivasangari, Breast Cancer Prediction via Machine Learning.
[25] Ali Bou Nassif,Heba Yahia Youssef,Chaimaa Boukhatem, Heart Disease Prediction Using Machine Learning.