iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** II **Month of publication:** February 2022

DOI: https://doi.org/10.22214/ijraset.2022.40245

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue II Feb 2022- Available at www.ijraset.com

Literature Survey for Lung Cancer Analysis and Prediction

Gayathri Devi Nagalapuram[1], Varshashree D[2], Vansika Singh[3], Dheeraj D[4], Donal Jovian Nazareth[5], Dr. Savitha Hiremath[6]

[1, 2, 3, 4, 5]*Department of Computer Science and Engineering, Dayananda Sagar University, Bengaluru*

[6]*Associate Professor, Department of Computer Science and Engineering, Dayananda Sagar University, Bengaluru*

Abstract: ***Lung cancer is one of the most common and deadly cancers worldwide which can be cured only if it is discovered at an early stage. Lung cancer can be diagnosed using various technologies, including MRI, isotopes, X-rays, and CT. One of the most effective ways to fight cancer is to discover it early enough to significantly improve the patient's chances of survival which can be done by the means of Artificial Intelligence. The proposed approach uses past medical records to determine if the patient has lung cancer. The CT scans are analyzed by a Convolutional Neural Network (CNN) model to determine the stage of cancer. Finally, the suggested model would forecast the patient's estimated medical insurance costs. Machine learning (ML) and Deep Learning (DL) approaches will be used to train and test the models by utilizing open-source datasets.***

Keywords: Lung Cancer, Computed Tomography (CT) scans, Machine Learning (ML), Convolutional Neural Network (CNN), GoogLeNet.

I. INTRODUCTION

Lung cancer has repeatedly been identified as one of the deadliest diseases in the history of mankind. It is also one of the most frequent malignancies and one of the leading causes of mortality. According to the World Health Organization (WHO), lung cancer causes around 7.6 million deaths worldwide each year. Furthermore, the number of people affected by cancer is expected to continue to rise, reaching around 17 million by 2030. In order to curb the rise, early detection is important.

There are many reasons behind cancer, ranging from behavioral traits such as high body mass index, tobacco and alcohol usage to physical carcinogens, such as exposure to ultraviolet rays and radiation, including certain biological and genetic carcinogens. However, the cause may vary from one patient to another. Common cancer symptoms are pain, fatigue, nausea, persistent cough, breathing difficulties, weight loss, muscle pain, bleeding, bruising, and many more. Then again, neither of these symptoms are exclusive to cancer, nor are all of them apparent in every patient. As a result, it is hard to determine the presence of cancer without a thorough diagnostic procedure such as Computed Tomography (CT) scan, Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET) scan, ultrasound, or biopsy. In many cases, the victims show little to no symptoms at the early stages, and when symptoms become apparent, more often than not, it is already too late.

This initiative aims to present a web application which allows people prone to lung cancer to keep a check on their symptoms and predict if they have cancer following which Deep Learning methods are used to determine the stage of cancer node and Machine Learning algorithms are used to predict medical insurance cost that can be availed for the patient. This system is meant to make it easy for the initial diagnosis for patients especially, in areas where medical care is not easily available and/or expensive hence, more people will check themselves regularly.

CT scans are used to determine the stages as it is a non-invasive method and the image can be uploaded on the web application with ease. CNN is the proposed model to be used for stage classification because of its high accuracy while dealing with images. It compares the image piece by piece and is extremely powerful.

Overall, this initiative shall help in quicker diagnosis and can be emulated for other diagnostic fields in Medicine.

II. RELATED WORK

Cancer causes about one in six deaths every year [1][2] and lung cancer stands at the top of all of this as it is responsible for 1.76 million deaths up to 2016. [1] Early detection of cancer can provide a suitable treatment to not just prolong life but also save a patient’s life and hence increase the survival rate. [1][2][3][4]

©IJRASET: All Rights are Reserved

The journal paper [1] published by Muthazhagan B, Ravi T, Rajinigirinath D (**2021**) states that with the aid of current lung cancer prediction technologies, predicting and detecting lung cancer at an early stage is a difficult challenge. An early lung tumor prediction might extend a person's life by one to five years.

They created a Support Vector Machine based classification model which provided about 98% prediction accuracy in a small amount of time.

However, the images were merely classified into 'abnormal' or 'normal' and did not take into account the various stages [Stage 0 – stage IV] which is what this project aims to improve on.

The paper [2] proposed by Masud M, Sikder N, et al. (**2021)** uses a CNN based model for classifying the image into one of five kinds: colon adenocarcinomas, benign colonic tissues, lung adenocarcinomas, lung squamous cell carcinomas and benign lung tissues.

While a peak accuracy of 96.33% has been achieved in the classification, the authors state that two out of five classes can have much improved performance with further experimentation. The dataset used is Histopathological and Histopathology is the microscopic examination of a biopsy which is an invasive process. Our approach prefers to work on CT scans which is a non-invasive mechanism to detect cancer.

Sajja T, Devarapalli R, et al. (**2019)** [3] published a paper which worked on detecting lung cancer using the pre-trained CNN model called Google-Net. The deployed 60% of all neurons in the drop out layers to prevent overfitting and achieved a simplified and sparse network for classifying the CT images into benign or malignant. The model still requires testing on various dropout ratios to check for better performance accuracy. Our approach aims to construct a simplified CNN model to classify cancer along with providing medical information costs.

Tripathi P, Tyagi S, et al. (**2019)** [4] published a paper in which they attempt to detect lung cancer using four different segmentation techniques of image processing.

They conclude that marker-controlled watershed segmentation provides the most accurate results. Through the comparative analysis, it is found that CT scans tend to provide the best chance at detecting cancer and should be the preferred means to do the same. Hence, we shall use Deep Learning on CT scans to classify the various stages.

Nasrullah Nasrullah et al. (**2019)** [5] study focuses on developing a model that can detect cancerous nodules using CT images. They opt to employ 3D CNN after some research because of its proven performance in image analysis. To further identify the condition as benign or malignant, they use 3D MixNet to extract nodule features, which are then classified using Gradient Boosting Machine (GBM). The proposed model was validated using the free response receiver operating characteristic (FROC) evaluation matrix to obtain a FROC score of 94.21%. The suggested model outperformed all other models in terms of computational cost and desired output accuracy.

Siddharth Bhatia et al. (**2019)** [6] present a method for detecting lung cancer using deep residual learning. They offer a series of preprocessing strategies for extracting cancer-vulnerable lung features using UNet and ResNet models. They examine the likelihood of predicting carcinogenic CT scans by comparing the effectiveness of classifiers such as Random forest and XGBoost. When the authors combine the two classifiers, they get the greatest accuracy of 84%. The constraint in this case is that the best achievable accuracy may have been higher.

Suren Makaju et al. (**2018)** [7] made a comparison of many probable cancer detection approaches and ranked them in order of effectiveness. They decide to upgrade that model to achieve even higher accuracy by selecting the current best approach from their survey of articles.

The Median and Gaussian filters were used in the pre-processing stage, and the data was then segmented using the Watershed algorithm. They went on to use support vector machines to identify diagnosed cancerous nodules as benign or malignant. This upgraded model outperformed the previous best model by 5.4%, with an accuracy rate of 92 %. The model's sole flaw is that it does not differentiate between cancer stages (I to IV).

Inspired by the AlphaGo system, Ali I et al. (**2018)** [8] developed a deep learning algorithm that takes a CT image and perceives it as a collection of states, producing a classification of whether or not a malignant nodule is present. They employ a Reinforcement Learning algorithm that improves with time and with more data.

Their research shows that the model's training data has a high accuracy of 99.1%, however the validation data has a low accuracy of 64.4 %. The model appears to be overfitted as a result of this. The authors suggest that because this is the only flaw, the constraint can be solved with more data.

III.ANALYSIS

Table I: Analysis of Various Methodologies

Sl. No.	Author Name and Year	Title of Paper	Methodology	Limitations/ Conclusions
1	Muthazhagan B, Ravi T, Rajinigirinath D- 2021	An Enhanced Computer-assisted Lung Cancer Detection Method Using Content-Based Image Retrieval and Data Mining Techniques [1]	Support Vector Machine image classification algorithm	The malignancy is classified as 'Normal' and 'Abnormal', not as Stages 1-4
2	Masud M, Sikder N, et al. - 2021	A Machine Learning Approach to Diagnosing Lung and Colon Cancer Using a Deep Learning-Based Classification Framework [2]	3 Digital Image Processing techniques with CNN	Dataset uses microscopic cells images rather than CT/MRI scans
3	Sajja T, Devarapalli R, Kalluri H- 2019	Lung Cancer Detection Based on CT Scan Images by Using Deep Transfer Learning [3]	A deep neural network based on Google-Net	Overfitted data causing the need for max dropout ratio
4	Tripathi P, Tyagi S, Nath M- 2019	Comparative Analysis of Segmentation Techniques for Lung Cancer Detection [4]	comparative analysis- image segmentation techniques	marker-controlled watershed segmentation provides more accurate results.
5	Nasrullah N, Sang J, Alam MS, Mateen M, Cai B, Hu H - 2019	Automated Lung Nodule Detection and Classification Using Deep Learning Combined with Multiple Strategies [5]	Two 3D CNN with CMixNet architectures	3D CMixNet had better accuracy feature exploitation than other models compared with.
6	Bhatia S, Sinha Y, Goel L – 2019	Lung Cancer Detection: A Deep Learning Approach [6]	deep residual networks with XGBoost and Random Forest classifiers and ensemble	The highest accuracy was 84% using an ensemble of both models tried which still a comparatively low accuracy
7	Makaju S, Prasad PW, et al. - 2018	Lung Cancer Detection using CT Scan Images [7]	Watershed algorithm with SVM	Classification of different stages of cancer is not done
8	Ali I, Hart GR, et al. - 2018	Lung Nodule Detection *via* Deep Reinforcement Learning [8]	Reinforcement learning algorithm	The model is overfit as training accuracy obtained was 99.1% whereas the testing accuracy was 64.4%

IV. DESCRIPTION OF PROJECT

The proposed project is a web application with the main web page comprising three buttons. The first button directs the user to the lung cancer analysis page, the second button to the lung cancer prediction page, and the third button takes the user to the insurance analysis page. The analysis pages contain interactive graphs made by Plotly. The visualizations provide a better understanding of the datasets. The lung cancer prediction page comprises a form that takes user symptoms as input of patient symptoms. Here, The Random Forest Classifier is trained with a dataset containing previous patients' symptoms records and is used to classify whether or not the patient has lung cancer. If the user does not have lung cancer, then it directs the user to a page displaying a message that the user is healthy. But, if the patient has lung cancer, then it directs the patient to another page. This page comprises the lung cancer stage classification button and the Medical insurance estimation form.

On clicking the lung cancer stage classification button, the user has to upload a CT Scan image to know whether it is a Normal, Benign, or Malignant case. The Convolutional Neural Network is trained with class weights using the IQ-OTHNCCD lung cancer dataset to perform this classification. Further, if it is a malignant case, it displays the lung cancer stage on a scale of 1 to 4. The GoogLeNet model which is trained using the LUNA16 dataset and LIDC-IDRI dataset is used to predict and display the Malignant stage of lung cancer. The datasets contain CT scan images and the patient's lung cancer stage and other patient details.

The Medical insurance estimation form takes user details like age, gender, region as input and provides the estimated lung cancer treatment costs. Here, a dataset containing previous patients records is used to train the Random Forest Regressor to predict the costs. Finally, to provide access to everyone, the proposed flask web application is deployed on Heroku. By using this web application, all the users can get results accurately, in less time effortlessly.

V. CONCLUSION

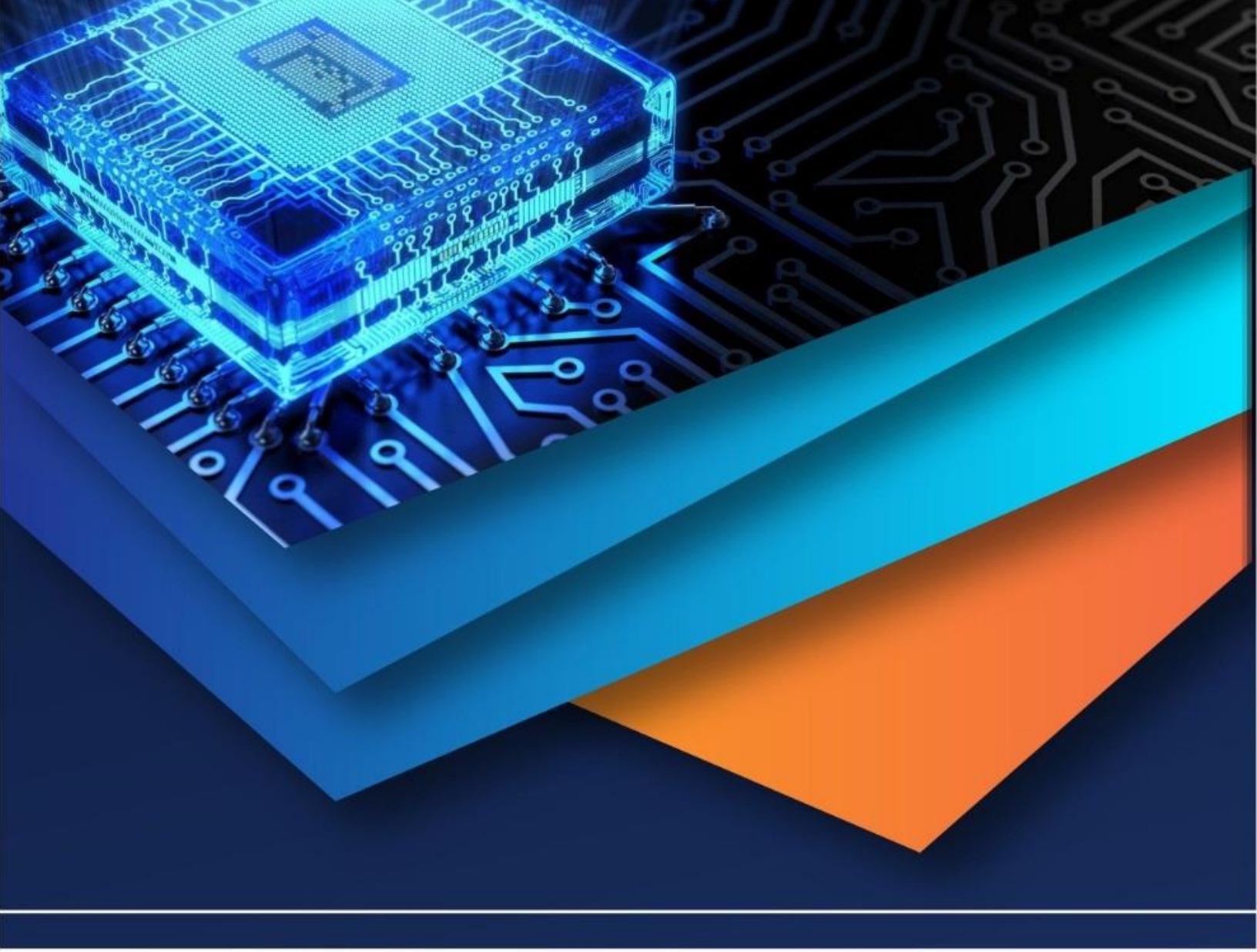
A technique for quick detection of lung cancer using CT scans was proposed in this research. We discovered that the random forest classifier and regressor gave more accurate results than other algorithms when we examined different machine learning techniques from the survey. In addition, Convolutional Neural Networks with class weights provide reliable results for categorizing CT scan pictures by overcoming unbalanced data issues. When compared to previous transfer learning models, the GoogLeNet model successfully predicts the malignant tumor level in less time.

As a result, the proposed model overcomes the disadvantages and provides an application in which the user may obtain information ranging from a basic level, such as symptom prediction. Then it allows lung cancer stage classification using CT scans, and presents insurance costs. The user may obtain all of the information they want from a single application in less time.

REFERENCES

[1] Muthazhagan B, Ravi T, Rajinigirinath D. An enhanced computer-assisted lung cancer detection method using content-based image retrieval and data mining techniques. Journal of Ambient Intelligence and Humanized Computing. 2020 Jun 2:1-9.
[2] Masud M, Sikder N, Nahid AA, Bairagi AK, AlZain MA. A machine learning approach to diagnosing lung and colon cancer using a deep learning-based classification framework. Sensors. 2021 Jan;21(3):748.
[3] Sajja T, Devarapalli R, Kalluri H. Lung Cancer Detection Based on CT Scan Images by Using Deep Transfer Learning. Traitement du Signal. 2019 Oct;36(4):339-44.
[4] Tripathi P, Tyagi S, Nath M. A comparative analysis of segmentation techniques for lung cancer detection. Pattern Recognition and Image Analysis. 2019 Jan;29(1):167-73.
[5] Nasrullah N, Sang J, Alam MS, Mateen M, Cai B, Hu H. Automated lung nodule detection and classification using deep learning combined with multiple strategies. Sensors. 2019 Jan;19(17):3722.
[6] Bhatia S, Sinha Y, Goel L. Lung cancer detection: a deep learning approach. InSoft Computing for Problem Solving 2019 (pp. 699-705). Springer, Singapore.
[7] Makaju S, Prasad PW, Alsadoon A, Singh AK, Elchouemi A. Lung cancer detection using CT scan images. Procedia Computer Science. 2018 Jan 1;125:107-14.
[8] Ali I, Hart GR, Gunabushanam G, Liang Y, Muhammad W, Nartowt B, Kane M, Ma X, Deng J. Lung nodule detection via deep reinforcement learning. Frontiers in oncology. 2018 Apr 16;8:108.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)