



iJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 Issue: VI Month of publication: June 2022

DOI:

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

A Web Application for Lung Cancer Analysis and Detection using Deep Learning Approach

Gayathri Devi Nagalapuram¹, Varshashree D², Vansika Singh³, Dheeraj D⁴, Donal Jovian Nazareth⁵, Dr. Savitha Hiremath⁶

^{1, 2, 3, 4, 5}Department of Computer Science and Engineering, Dayananda Sagar University, Bengaluru

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

Abstract: Lung cancer is one of the most common and deadly cancers worldwide. One of the most effective ways to fight cancer is to discover it early enough to improve the patient's chances of survival. The Discovery of lung cancer at an early stage helps in reducing its risk. Various technologies like MRI, isotopes, X-rays, and CT scans are used for diagnosis of lung cancer. The studying of lung nodules helps a doctor to determine if the patient is malignant. These nodules sometimes have a chance of growing undetected by the naked eye. In this project, Lung cancer stage is detected with the help of patient details, symptoms and CT scans by using Machine learning and Deep learning algorithms with open-source datasets. The proposed approach uses Machine learning algorithms to study past medical records and determine if the patient has lung cancer. Deep learning models are used to analyze the CT scans to determine the stage of lung cancer. The major goal of this project is to find nodules as small as 3 mm to detect cancer stage accurately. Finally, the machine learning model calculates the patient's estimated medical insurance costs. This project is useful for the early detection of lung cancer in individuals and can help them in overcoming these health conditions. The effectiveness of cancer prediction systems helps the people to know their cancer risk with low cost and it also helps the people to take the appropriate decision based on their cancer risk status.

Keywords: Cancer, U-Net, Deep Learning, Machine Learning (ML), Convolutional Neural Network (CNN)

I. INTRODUCTION

Lung cancer is repeatedly identified as one of the deadliest diseases in the history of humankind. It has been one of the most common cancers and one of the top causes of death. Lung cancer kills over 7.6 million people globally each year, as per the World Health Organization (WHO). The number of cancer patients is predicted to climb further, reaching roughly 17 million by 2030.

According to the Centers for Disease Control and Prevention (CDC), individuals who smoke tobacco are 15 to 30 times more likely than non-smokers to develop or succumb from lung cancer. Lung cancer in non-smokers can be triggered by radon, passive smoking, air pollution, or other reasons. Worksite exposure to asbestos, diesel fumes, or other pollutants can also lead to lung cancer in nonsmokers. Typical symptoms include coughing (often with blood), chest discomfort, wheezing, and loss of weight. However, these symptoms do not generally appear until the malignancy has advanced.

Cancer is caused by a variety of factors, ranging from behavioral factors such as a high BMI, tobacco and alcohol use, to physical carcinogens through exposure to UV rays and radiation, as well as some biological and hereditary carcinogens. However, the aetiology may differ from one victim to the next. Soreness, exhaustion, nausea, chronic cough, breathing difficulty, weight loss, muscular pain, bleeding, bruises, and other symptoms are frequent in cancer patients. However, none of these symptoms is unique to cancer, nor are they all present in every patient. CT scans that look for nodules in the lung are frequently used to detect lung cancer. Lung nodules (or masses) are little abnormal areas that can be found during a chest CT scan. By examining the nodules, a doctor can determine whether or not this scan is malignant.

Doctors can evaluate nodules larger than 7 millimeters in diameter, and physicians frequently instruct patients to wait to see if the nodule develops. If it does not develop, the nodule is harmless. There is a greater likelihood that the nodule may go undiscovered. As a result, determining cancer without a comprehensive diagnostic method, such as a Computed Tomography (CT) scan, Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET) scan, ultrasound, or biopsy, is difficult. In most cases, victims exhibit little to no indications in the early stages, and by the time symptoms emerge, it is typically too late.

The goal of this study is to look at alternative ways for detecting and recognizing lung cancer in its early stages. As a consequence, lung cancer can be treated before it progresses to the point where therapy is no longer viable.

The focus of this work is solely on the detection and prediction of lung cancer and its stages.

II. LITERATURE SURVEY

Every year, cancer kills around one in every six people [1-2], with lung cancer topping the list with 1.76 million fatalities in 2016. [1] Early cancer identification can give a proper therapy to not only prolong but even save a person's life, hence increasing the survival rate. [1]-[4].

Deep residual learning has been used by Siddharth Bhatia et al. (2019) [6] to detect lung cancer. They provide a set of preprocessing algorithms for obtaining cancer-vulnerable lung characteristics from images using UNet and ResNet models. They compare the efficiency of classifiers such as Random forest and XGBoost in predicting carcinogenic CT images. The authors achieve the highest accuracy of 84% when they mix the two classifiers.

In this study (2019) [9], Ibrahim M. Nasser et al. created an Artificial Neural Network (ANN) to identify the presence or absence of lung cancer in the human body. Symptoms such as yellow fingers, anxiety, chronic disease, and others were utilized to identify lung cancer. They were employed as input variables to their ANN, along with additional information of the patient. The ANN model was shown to be 96.67% accurate in detecting the absence or presence of lung cancer.

Wookjin Choi et al. (2018) [10] attempted to address various shortcomings identified in the existing literature in their article. They used hierarchical clustering to identify discrete radiomic characteristics before building a support vector machine (SVM) model with just two key features chosen using a least absolute shrinkage and selection operator (LASSO). They created this model using two CT radiomic characteristics to indicate the malignancy of pulmonary nodules. It provided an 84.6% accuracy rate.

Gayathri Devi Nagalapuram et al. (2022) [OUR] offered a review of several approaches employed in numerous practical experiments in their research. They determined that all techniques might go beyond simply determining if a patient's status is normal or not. They also indicate that the majority of accuracies for a healthcare endeavor are relatively mediocre.

III. PROPOSED DESIGN

The fundamental purpose of this study was to review past medical information in order to discover lung cancer, forecast if the patient has lung cancer using CT scans, and then determine the location of the cancer nodule in the scan. Estimating medical insurance costs provides further assistance.

The proposed design was broadly divided into 6 stages:

- 1) *Lung Cancer Detection based on Symptoms:* The Random Forest Classifier was trained on a dataset including previous patients' symptom records and was used to determine whether or not the patient has lung cancer.
- 2) *Lung Cancer Classification using CT Scans:* The IQ-OTHNCCD lung cancer dataset is used to train the Convolutional Neural Network. This CNN model categorizes CT scans as normal, benign, or malignant.
- 3) *Lung Cancer Nodule detection using Deep Learning:* UNET model trained using the LUNA 16 dataset for nodule detection. The mhd files in the dataset are converted to png files for training the UNet. Also, mask extraction is performed to train the CNN to identify nodules and variety of preprocessing steps are performed to segment out the ROI (the lungs) from the surrounding regions of bones and fatty tissues.
- 4) *Medical Insurance Cost Prediction:* An insurance dataset was used to train the Random Forest Regressor to estimate a patient's insurance costs if the individual has lung cancer.
- 5) *Data Analysis using Plotly:* Plotly interactive graphs were used on analytical web pages. The visuals improve comprehension of the datasets.
- 6) *Develop a Web Application:* All of the above procedures are combined to produce a web application for improved user engagement. The program aids in the detection of lung cancer in real time and the assessment of medical insurance costs.

The proposed project is a web application with the main web page comprising four buttons. The first button directs the user to the lung cancer detection based on symptoms, the second button to the lung cancer classification based on CT scans page, and the third button takes the user to the lung cancer nodule detection using deep learning. Finally, the fourth button directs the user to the medical insurance cost prediction. A flowchart of the proposed design is shown in Fig. 1.

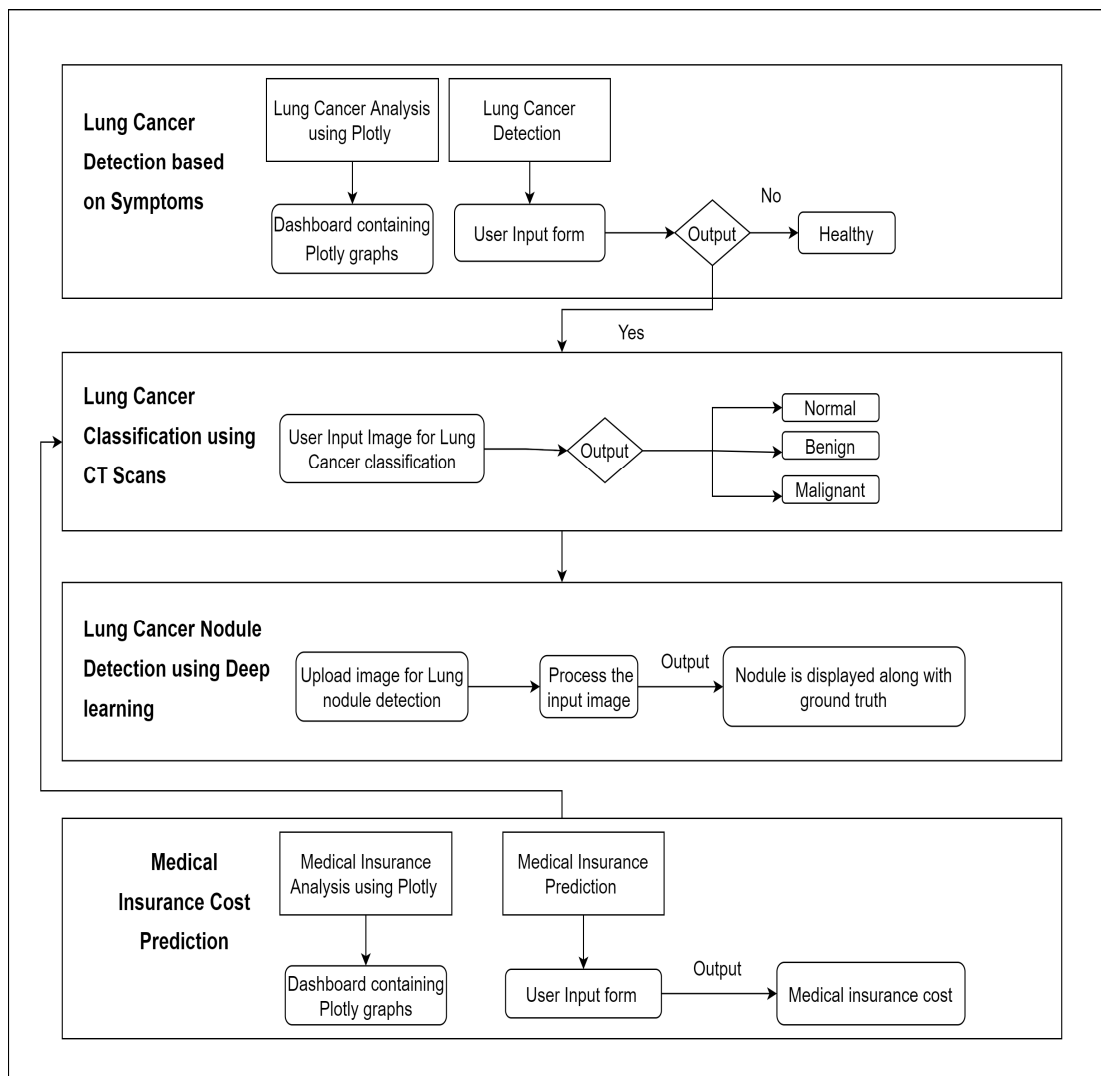


Fig.1 Proposed Workflow

The lung cancer detection based on symptoms page contains two buttons. On clicking the first button, the interactive graphs made by Plotly are displayed. The visualizations provide a better understanding of the datasets. The lung cancer detection button comprises a form that takes user symptoms as input and classifies whether or not the patient has lung cancer. If output is not healthy then the user is directed to lung cancer type classification page. If it is healthy, then the user is directed to a page containing the message and a button to go to the type detection page.

The lung cancer type detection page contains a message that patient maybe at risk and an image upload button where the patient can upload the image and then click submit button.

The trained CNN model predicts whether the output is malignant, benign or normal case. The output page will display the CT scan image, the type of cancer case and a button to direct the user to nodule detection page.

The lung cancer nodule detection page contains a button to upload the CT scan images. On clicking the upload button, the trained UNET model will predict the malignant nodule and directs the user to the output page which displays the nodule in the CT scan.

The Medical insurance page contains two buttons. The first button directs the user to the plotly dashboard which provide insights about the dataset and the second button takes user to the estimation form which takes user details. The trained RFR model estimates lung cancer treatment costs and displays it on the output page when user clicks the submit button. This output page also contains a button to navigate user to type detection page.

This proposed flask web application provides accurate results to all the users in less time effortlessly.

IV.IMPEMENTATION/METHODOLOGY

The project has different modules implemented with different methodologies as shown in Fig. 2 Each module and its implementation will be explained in depth in the sections that follow.

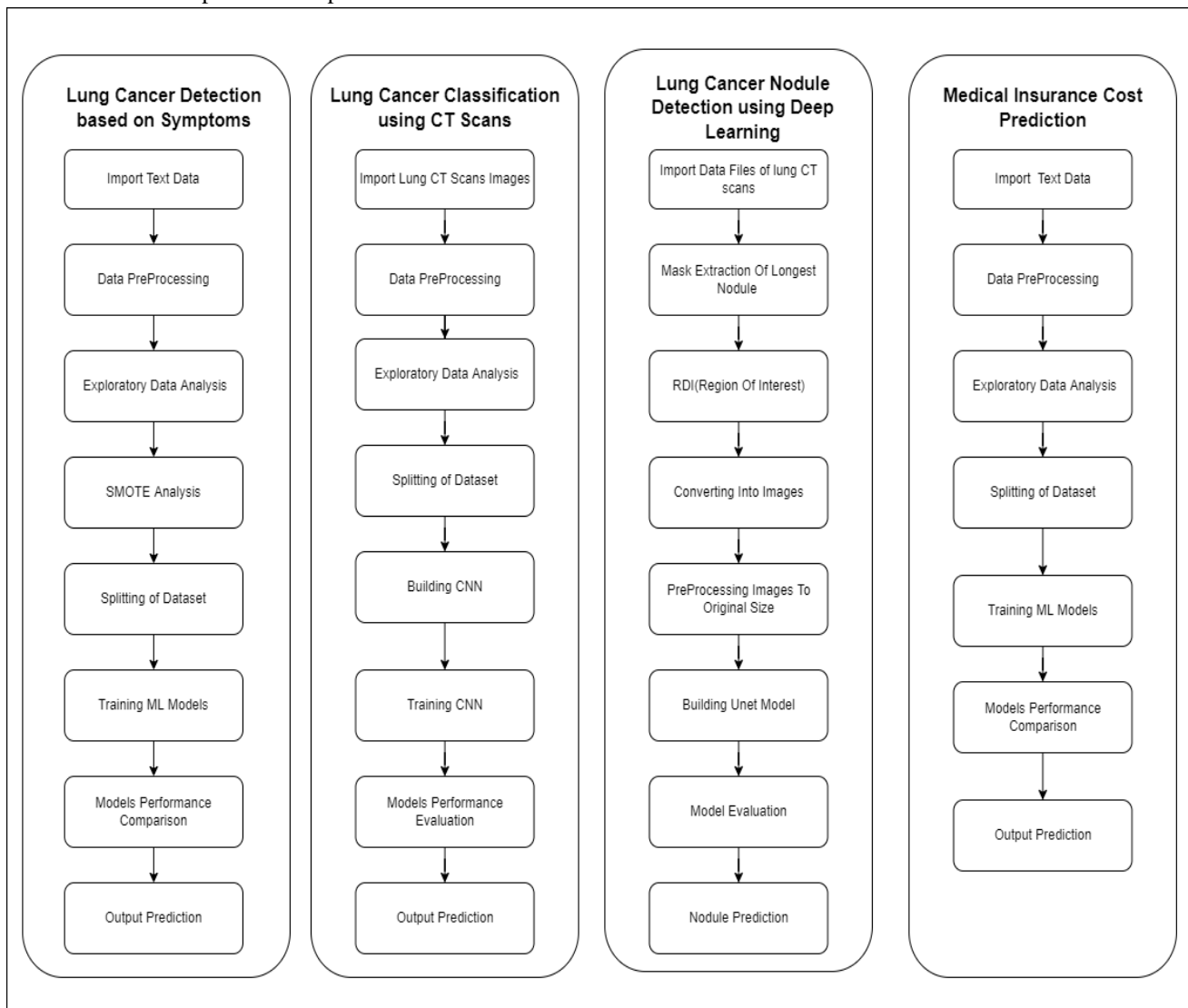


Fig.2 Flowchart of Implementation

A. Lung Cancer Detection Based on Symptoms

Lung cancer can be caused by a variety of symptoms and habits. We construct models to predict whether or not a patient has lung cancer using user data.

- 1) *Data Analysis:* A comprehensive data analysis is required to determine the type of data with which we are working. The Survey Lung Cancer dataset was obtained from the *data.world* website. The data is gathered through the website's online lung cancer prediction system by collecting user inputs. The users who visited this site implemented this throughout the month of August 2013. The data set contains 309 entries with 16 variable columns.
- 2) *Data Processing:* This step is where certain alterations to the data are made. Some adjustments included renaming columns and encoding categorical data. Resampling was employed to balance the dataset since it was uneven, with 87.4% of people having lung cancer and the remaining 12.6% people not having lung cancer. To expand the number of instances in the dataset in a balanced manner, the Synthetic Minority Oversampling Approach (SMOTE) was employed as an oversampling technique.
- 3) *Data Exploration:* Determining the relationship between distinct variables. Among all features, Anxiety and Yellow Fingers show the highest correlation whereas Peer pressure and Gender show the least.

Certain inferences are reached to aid our understanding, such as:

- Males outnumber females in number
- Males have a 0.4 older mean age than females
- Male smokers outweigh female smokers
- Females are more likely to have yellow fingertips
- Females frequently exhibit anxiety symptoms
- Females face higher peer pressure
- Chronic illness is also more prevalent among women

Further investigation was conducted with regard to all factors. The gathered findings were utilized to develop a dashboard page, which is depicted in the succeeding sections.

- 4) *Model Building*: The generation of different training and testing samples allows us to evaluate model performance. As a result, we divide our modelling dataset into training and testing samples using the scikit-learn library's train test split() method. Following data splitting, the train data is sent to several models for training. The models utilized are KNN: K-Nearest Neighbors, RFC: Random Forest Classifier, SVC: Support Vector Classifier, DTC: Decision Tree Classifier. The models are built to forecast once they have been trained. The predicted values are compared to the validation data to determine the accuracy of each model. The RFC model had the highest accuracy and was chosen as the best model to predict the existence of cancer based on the user's symptoms.

B. Medical Insurance Cost Prediction

The insurance dataset is used to train a model that can forecast the approximate cost of insurance that a cancer patient will require. The forecast will be based on the information provided by the user.

- 1) *Data Analysis*: A comprehensive data analysis is required to determine the type of data we are working with. The Medical Insurance Dataset was obtained from Kaggle and had 1338 records with 7 variable columns.
- 2) *Data Processing*: This step is where some adjustments to the data are made. Categorical data was encoded.
- 3) *Data Exploration*: Studying the features reveals that Smokers and Charges have the highest correlation whereas Smoking and Age, and Smoking and Number of Children have the least.
- 4) *Model Building*: The generation of different training and testing samples allows us to evaluate model performances. As a result, we divide our modelling dataset into training and testing samples, to which models were implemented using about 1070 samples for training and 268 samples for testing.

Following data splitting, the train data is sent to several models for training. The models utilized are as follows:

- Linear Regression
- Random Forest Regressor
- Decision Tree Regression
- Lasso Regression

To assess the correctness of each model, the predicted values are compared to the validation data. The RFR model was chosen as the best model to anticipate insurance premiums based on user input since it had the highest accuracy.

C. Lung Cancer Classification Using CT scans

Lung cancer can be caused by a variety of symptoms and habits. We construct models to predict whether or not a patient has lung cancer using user data.

- 1) *Data Analysis*: The lung cancer dataset from Iraq-Oncology Teaching Hospital/National Center for Cancer Diseases (IQ-OTH/NCCD) was gathered in the aforementioned specialty hospitals over a three-month period in autumn 2019. It comprises CT images of patients with lung cancer at various stages as well as healthy participants. Oncologists and radiologists from these two hospitals marked the IQ-OTH/NCCD slides. The collection comprises 1190 pictures depicting CT scan slices from 110 different cases. These instances are classified as either normal, benign, or malignant. 40 of these instances have been classed as malignant, 15 as benign, and 55 as normal. Gender, age, educational achievement, location of residence, and living situation differ across the 110 cases. An example of benign, malignant and normal case is given in Fig. 3.

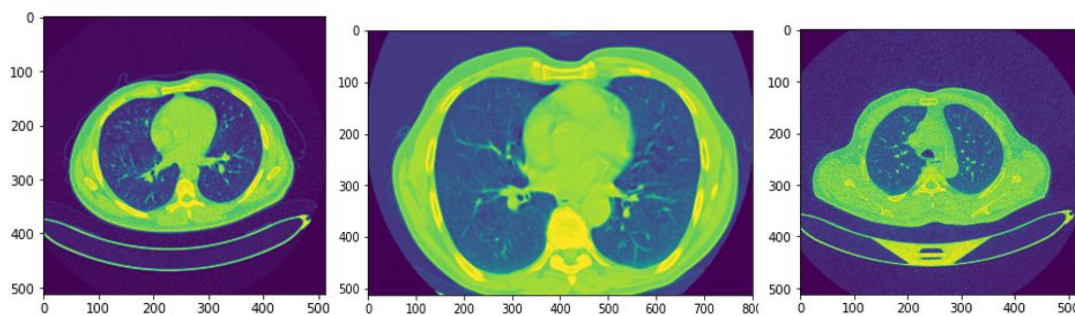


Fig. 3 Benign, Malignant and Normal Cases

- 2) *Data Processing and Exploration:* The visual data is jumbled and then displayed using colormaps. Using one hot encoding, the data is normalized, transformed, and encoded.
- 3) *Model Building:* The data is subsequently divided into two sets: training and testing. The data split includes photos from all three classes in both sets. A Convolutional Neural Network was built to predict the right cell classifications from photos. For feature extraction from photos, we employed three Conv2D layers with MaxPool2D layers in between. ReLU is the activation function employed. The output layer has just three neurons with SoftMax activation functions, which correspond to the three types of tumors (Benign, Malignant, and Normal). The model is then built using RMSprop as the optimizer and Categorical Crossentropy as the loss function. We will train the model with the class weights for three epochs. The training was terminated in the third epoch when a decent accuracy of roughly 92% was established.

D. Lung Cancer Nodule Detection Using Deep Learning

- 1) *Data Analysis:* Scans with a slice thickness of more than 2.5 mm were eliminated from the dataset. There are 888 CT scans in all. Annotations were collected throughout a two-phase annotation procedure by four experienced radiologists and are also available in the LIDC/IDRI database. Each radiologist labelled lesions as non-nodule, nodule less than 3 mm, or nodule greater than 3 mm. The specifics of the annotation procedure may be found in their publication. All nodules greater than or equal to 3 mm that are acknowledged by at least three out of four radiologists constitute the challenge reference standard. Irrelevant findings are annotations that are not included in the reference standard (non-nodules, nodules less than 3 mm, and nodules annotated by just one or two radiologists). All CT scans and annotations are contained in subset0, which is a zip file. The annotations used as a reference standard for the 'nodule detection' track are stored in a csv file. CT images are saved in MetaImage (mhd/raw) format in subset0. The pixel data is kept in a separate raw binary file for each mhd file. One discovery per line is contained in the annotation file, which is a csv file. The scan's SeriesInstanceUID, the x, y, and z positions of each discovery in world coordinates, and the matching dimension in mm are all listed on each line. There are 1186 nodules in the annotation file.
- 2) *Data Processing and Exploration:* It is necessary to minimise the domain size before entering the CT images into the U-net architecture in order to obtain more accurate findings. To segment out the ROI (the lungs) from the surrounding areas of bones and fatty tissues, a number of preprocessing processes are used. These include the following:
 - a) Binary Thresholding
 - b) Selecting the two largest connected regions
 - c) Erosion to separate nodules attached to blood vessels
 - d) Dilation to keep nodules attached to the lung walls
 - e) Filling holes by dilation
 - f) Converting the mhd files to png

The image file's index is collected, and the index image's directory is saved. The image is then opened, enlarged, and transformed to grayscale, after which its index is saved. Along with the grayscale picture, the mask of the image corresponding to the index is also saved. After that, masks may be read simply specifying the mask's directory. Finally, the mask image is preprocessed by shrinking and normalising the pixel value before being stored at the same index position in the output array as the pre-processed mask image. The picture is stored in X[n], while the mask is stored in y[n].

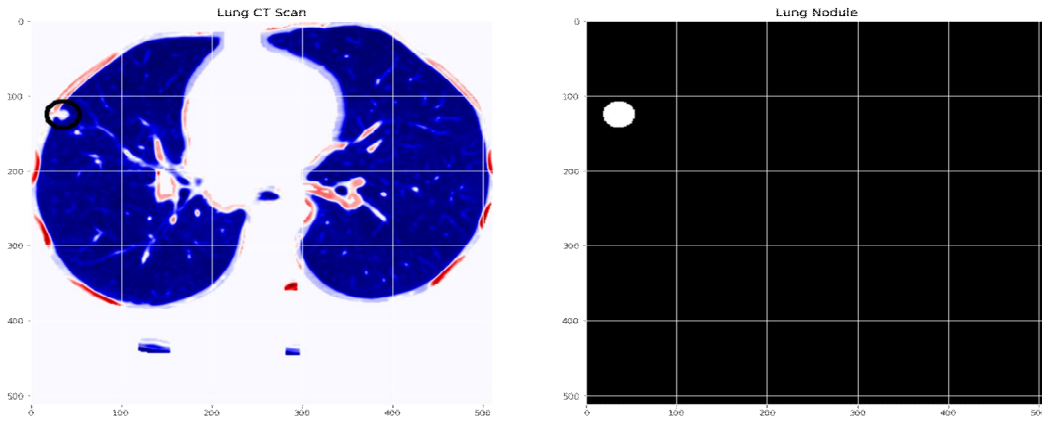


Fig. 4 CT Scan and its corresponding Lung Nodule mask Visualization

Fig. 4 shows the plot of the pre-processed sample image and the mask of the picture exhibiting the nodule.

- 3) *Model Building*: The base model and the custom layer that accepts that base model input whose output is subsequently transferred to the UNet model anticipate an input shape of 512x512x1. The output of the UNet model is then sent to further ConvNet layers with ReLU activation followed by a flatten layer and two dense layers. After that, the output is reshaped to 512X512. Finally, we've used the base model to construct a model that receives input (inp) and outputs (x out). The model is then built using Adam as an optimizer and Binary Crossentropy as a loss function. With the class weights, we'll train the model for three epochs. The training was halted at the tenth epoch since a decent accuracy of around 98\% had been achieved.

V. EXPERIMENTATION AND RESULTS

A. Lung Cancer Detection Based on Symptoms

As previously indicated, many models were used, and the finest of them were picked to be employed. Table I shows the train and test accuracies for the various models utilized for lung cancer detection based on symptoms.

TABLE I
ACCURACY OF MODELS FOR SYMPTOMS BASED DETECTION

Model	Train_Accuracy	Test_Accuracy
RFC	98.020312	0.969136
SVC	96.296296	0.969136
DTC	98.876574	0.919753
KNN	94.444444	0.907407

The Random Forest Classifier performed the best.

B. Medical Insurance Cost Prediction

Table II shows score for the various models utilized for medical insurance prediction.

TABLE II
ACCURACY OF MODELS FOR INSURANCE PREDICTION

Model	Train_Accuracy	Test_Accuracy
Random Forest Regression	88.91	86.29
Decision Tree Regression	88.02	86.18
Lasso Regression	74.61	76.18
Linear Regression	74.61	76.18

Random Forest Regressor has yielded the best performance.

C. Lung Cancer Classification Using CT Scans

The following experimental results were obtained after utilizing the developed CNN model to predict a random picture. The first value is the class of the object with the highest value in the array. The class's category label is the following value. As seen in Fig. 5, the picture was correctly anticipated.

```
[0.04567198 0.5397512 0.41457677]
[0.09867036 0.6228367 0.27849287]
[0.01539725 0.90645313 0.07814966]]
1
Malignant cases
```

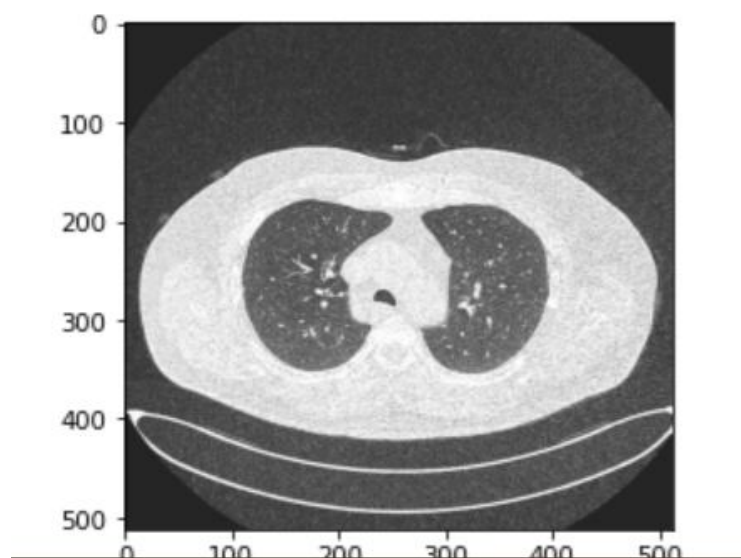


Fig.5 Prediction of randomly chosen image

A data frame with anticipated value and equality to its actual value is developed with further experimentation.

The model makes modest errors on just portion of the data if the accuracy is high and the loss is low, which is the optimum condition. The CNN model's loss and accuracy (Fig. 6) are shown in the charts below. The CNN Model yields an accuracy of 92.42% and a miss class of 7.58%

The performance of the CNN model is also evaluated using different performance metrics such as precision, recall, and f1-score, each of which provide an accuracy of 92%

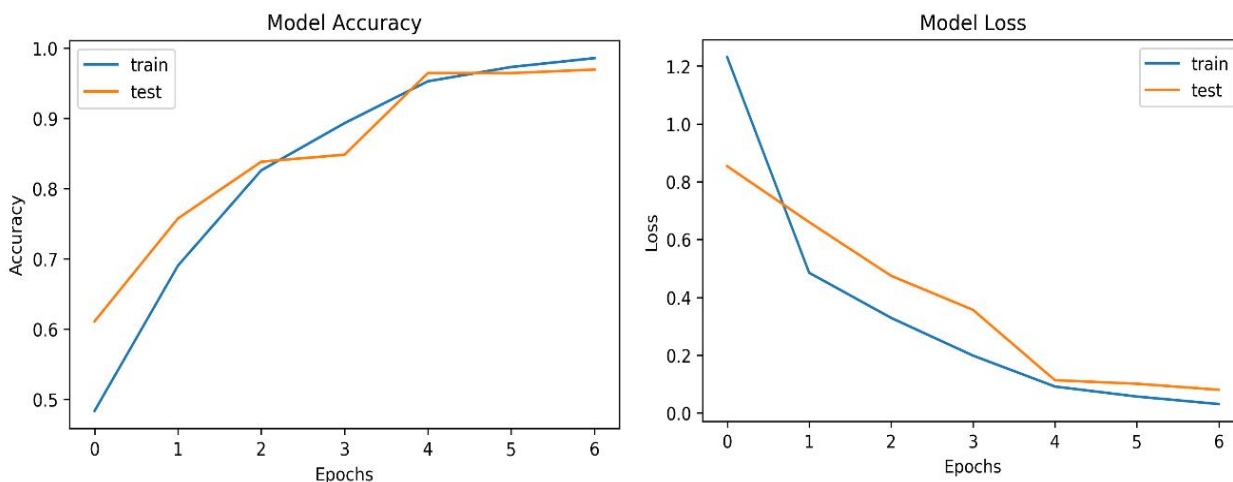


Fig.6 CNN model Accuracy and Loss

D. Lung Cancer Nodule Detection using Deep Learning.

A learning curve is a graph representing model learning performance as a function of time or experience. In machine learning, learning curves are a common diagnostic tool for algorithms that learn progressively from a training dataset. After each update during training, the model may be tested on the training dataset and a holdout validation dataset, and graphs of the measured performance can be constructed to demonstrate learning curves.

Examining model learning curves during training may help detect learning issues, such as an underfit or overfit model, as well as whether the training and validation datasets are sufficiently representative. The learning curve for lung cancer nodule identification is seen in Fig. 7. The curve is an example of a good fit

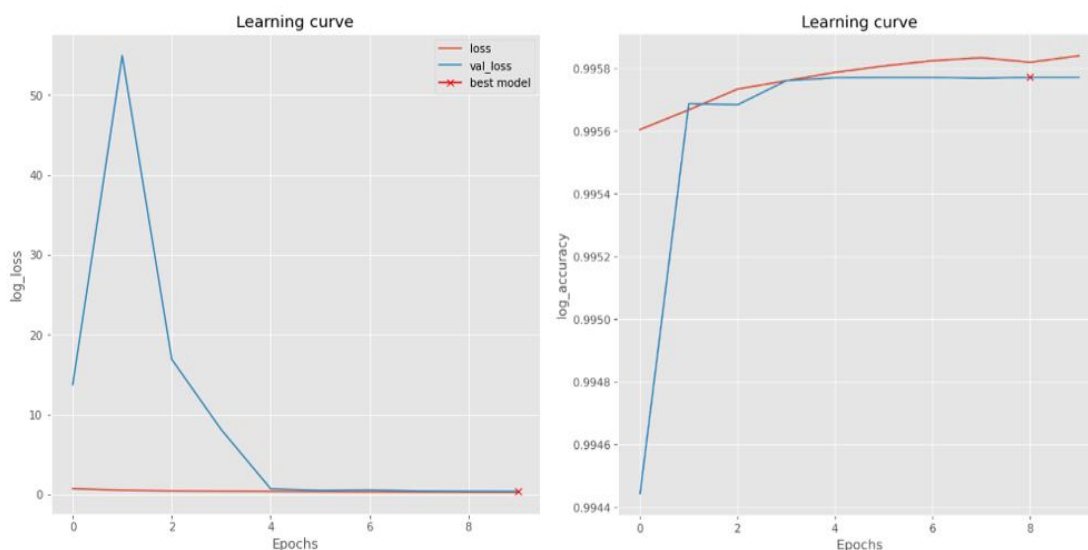


Fig.7 The Learning Curve for lung cancer nodule detection

Fig. 8 shows the output image that includes the Lung CT Scan, Ground Truth, Predicted Nodule and Predicted Nodule Binary.

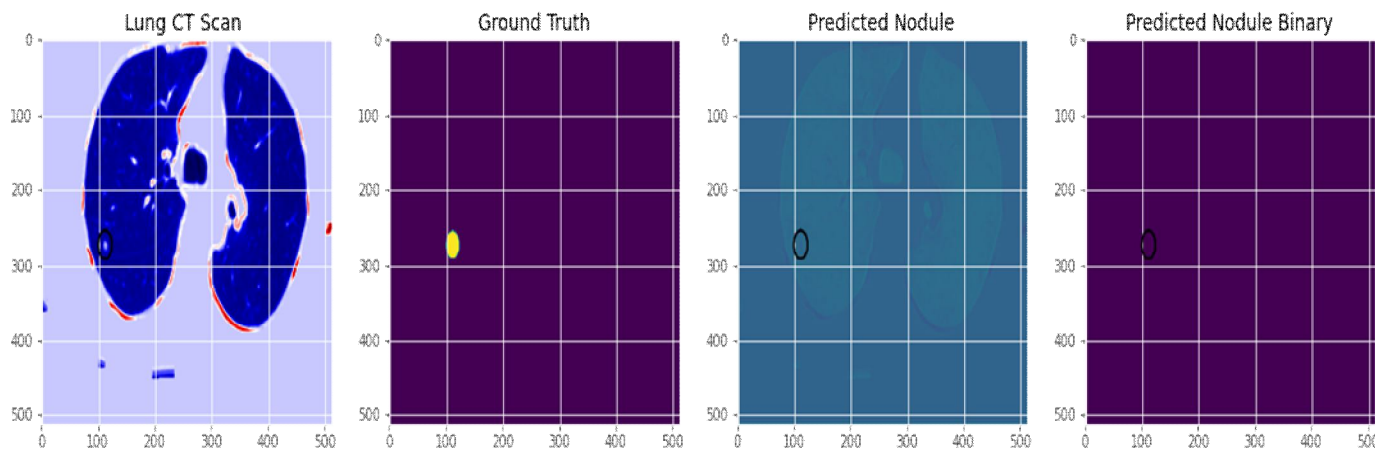


Fig.8 Output Image of detected nodule

E. Web App Integration Using Flask

A web application is created with the help of flask web framework to simplify the usage of the proposed project to all users the app. To test the proposed web application in a real-world, test data and images of CT scans are selected based on the convenient use of the applications and several experiments were conducted to test the model's robustness. On performing real-time testing on the proposed system using different sets of people, good results were achieved.

F. Comparison of Proposed Models

The accuracy of the proposed model is compared with different models from previously done experiments as shown in Table III. The proposed models outperformed the others in terms of performance.

TABLE III
COMPARISON OF PROPOSED MODELS

Modules	Ref	Model Used	Accuracy
Lung Cancer Nodule Detection	[19]	ResNet50 + SVM RBF	93.19
	[18]	CNN	92.63
	[20]	DFCNet	89.52
	[8]	CNN	64.40
	[10]	SVM-LASSO	84.60
		Proposed UNet	98.00
Lung cancer classification using CT scans	[21]	3D MixNet	88.83
	[21]	3D MixNet + GBM	90.57
	[6]	Ensemble of UNet + Random Forest and ResNet+XGBoost	84.00
	[7]	SVM	86.67
	[22]	Fine-tuning Conv4 of AlexNet	85.21
		Proposed CNN	92.42
Lung cancer prediction on the basis of symptoms	[9]	ANN	96.67
		Proposed RFC	96.91
Medical Insurance Cost Prediction	[23]	Stochastic Gradient Boosting	85.82
	[23]	XGBoost	85.36
	[24]	RFR	85.00
		Proposed RFR	86.29

VI.CONCLUSION

The goal of this project was to create a method for overcoming the challenges of lung cancer by utilizing Machine Learning and Deep Learning techniques to predict the presence of cancer in the lungs using medical records, as well as interpret CT images to accurately identify nodules with diameters as small as 3mm at a low cost and in less time. Along with that, to be able to make this technology available to everyone in the form of a web application. All of the objectives were met, as evidenced by the following outcomes. Random Forest Classifier with 96.9% accuracy for symptom-based recognition and Random Forest Regressor with 86.3% accuracy for predicting medical insurance costs. The CNN model, which was created to analyze CT images, The accuracy of the CNN model used to analyze CT images was 92.42%. Finally, the UNet model designed to detect nodules on CT scans performed excellently, with a 98% accuracy rate. The developed strategy is, in general, highly dependable for users.

REFERENCES

- [1] Muthazhagan, B., Ravi, T., & Rajiniginath, D.: An enhanced computer-assisted lung cancer detection method using content-based image retrieval and data min-ing techniques. *Journal of Ambient Intelligence and Humanized Computing*, 2:1-9, 2020.
- [2] Masud, M., Sikder, N., Nahid, A. A., Bairagi, A. K., & AlZain, M. A.: A machine learning approach to diagnosing lung and colon cancer using a deep learning-based classi-fication framework. *Sensors*, 21(3):748, 2021.
- [3] Sajja, T., Devarapalli, R., & Kalluri, H.: Lung Cancer Detection Based on CT Scan Images by Using Deep Transfer Learning. *Traitement du Signal*, 36(4):339-44, 2019.
- [4] Tripathi, P., Tyagi, S., & Nath, M.: A Comparative Analysis of Segmentation Techniques for Lung Cancer Detection. *Pattern Recognition and Image Analysis*, 29. 167-173, 2019.
- [5] Nasrullah, N., Sang, J., Alam, M. S., Mateen, M., Cai, B., & Hu, H.: Automated lung nodule detection and classification using deep learning combined with multiple strategies. *Sensors*, 19(17):3722, 2019.
- [6] Bhatia S, Sinha Y, Goel L.: Lung cancer detection: a deep learning approach. *InSoft Computing for Problem Solving*. Springer, Singapore 699-705 (019
- [7] Makaju, S., Prasad, P. W. C., Alsadoon, A., Singh, A. K., & Elchouemi, A.: Lung cancer detection using CT scan images. *Procedia Computer Science*, 125, 107-114, 2018.

- [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*, 16:8:108, 2018.
- [9] Nasser, I. M., & Abu-Naser, S. S.: Lung cancer detection using artificial neural net-work. *International Journal of Engineering and Information Systems (IJEAIS)*, Mar;3(3):17-23, 2019.
- [10] Choi, W., Oh, J. H., Riyahi, S., Liu, C. J., Jiang, F., Chen, W., ... & Lu, W.: Radiomics analysis of pulmonary nodules in low-dose CT for early detection of lung cancer. *Medical physics*, 45(4):1537-49, 2018.
- [11] Kadir, T., & Gleeson, F.: Lung cancer prediction using machine learning and advanced imaging techniques. *Translational lung cancer research*, 7(3):304, 2018.
- [12] Raoof, S.S., Jabbar, M.A., & Fathima, S.A.: Lung Cancer prediction using machine learning: A comprehensive approach. In: 2nd International conference on innovative mechanisms for industry applications (ICIMIA). IEEE, 2020.
- [13] Xie, Y., Meng, W. Y., Li, R. Z., Wang, Y. W., Qian, X., Chan, C., Yu, Z. F., Fan, X. X., Pan, H. D., Xie, C., Wu, Q. B., Yan, P. Y., Liu, L., Tang, Y. J., Yao, X. J., Wang, M. F., & Leung, E. L.: Early lung cancer diagnostic biomarker discovery by machine learning methods. *Translational oncology*, 14.1: 100907, 2021.
- [14] Singh, G.A., & Gupta, P.: Performance analysis of various machine learning-based approaches for detection and classification of lung cancer in humans. *Neural Computing and Applications* 6863-6877, 2018.
- [15] Shin, H., Oh, S., Hong, S., Kang, M., Kang, D., Ji, Y. G., ... & Choi, Y.: Early-stage lung cancer diagnosis by deep learning-based spectroscopic analysis of circulating exosomes. *ACS nano*, 14(5), 5435-5444, 2020.
- [16] Hosny, A., Parmar, C., Coroller, T. P., Grossmann, P., Zeleznik, R., Kumar, A., ... & Aerts, H. J.: Deep learning for lung cancer prognostication: a retrospective multi-cohort radiomics study. *PLoS medicine*, 15.11, 2018.
- [17] Lakshmanaprabu, S. K., Mohanty, S. N., Shankar, K., Arunkumar, N., & Ramirez, G.: Optimal deep learning model for classification of lung cancer on CT images. *Future Generation Computer Systems*, 92: 374-382, 2019.
- [18] de Carvalho Filho, A. O., Silva, A. C., de Paiva, A. C., Nunes, R. A., & Gattass, M.: Classification of patterns of benignity and malignancy based on CT using topology-based phylogenetic diversity index and convolutional neural network. *Pattern Recognition*, 81, 200-212 (2018)
- [19] da Nóbrega, R. V. M., Rebouças Filho, P. P., Rodrigues, M. B., da Silva, S. P., Dourado Júnior, C. M., & de Albuquerque, V. H. C.: Lung nodule malignancy classification in chest computed tomography images using transfer learning and convolutional neural networks. *Neural Computing and Applications*, 32(15), 11065-11082, 2020.
- [20] Masood, A., Sheng, B., Li, P., Hou, X., Wei, X., Qin, J., & Feng, D.: Computer-assisted decision support system in pulmonary cancer detection and stage classification on CT images. *Journal of biomedical informatics*, 79, 117-128, 2018.
- [21] Sang, J., Alam, M. S., & Xiang, H.: Automated detection and classification for early stage lung cancer on CT images using deep learning. In *Pattern Recognition and Tracking XXX (Vol. 10995, p. 109950S)*. International Society for Optics and Photonics, 2019.
- [22] Shan, H., Wang, G., Kalra, M. K., de Souza, R., & Zhang, J.: Enhancing transferability of features from pretrained deep neural networks for lung nodule classification. In *Proceedings of the 2017 International Conference on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, 2017.
- [23] Hanafy, Mohamed.: Predict Health Insurance Cost by using Machine Learning and DNN Regression Models. *International Journal of Innovative Technology and Exploring Engineering*. Volume-10. 137, 2021.
- [24] Iqbal, J., Hussain, S., AlSalman, H., Mosleh, M. A., & Sajid Ullah, S.: A Computational Intelligence Approach for Predicting Medical Insurance Cost. *Mathematical Problems in Engineering*, 2021.



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)