LUNG SENSE PRO: PREDICTIVE MODELLING FOR LUNG CANCER ANALYSIS USING MACHINE LEARNING

A PROJECT REPORT

Submitted by,

BATHALA PRAVEEN (723920104009)

CHUNCHU MANOJ (723920104015)

SASIDEVAN P (723920104049)

In partial fulfillment for the award of the degree

Of

BACHELOR OF ENGINEERING

In

COMPUTER SCIENCE AND ENGINEERING

ARJUN COLLEGE OF TECHNOLOGY

COIMBATORE – 642 120

ANNA UNIVERSITY: CHENNAI 600 025

MAY 2024

ANNA UNIVERSITY: CHENNAI 600 025

BONAFIDE CERTIFICATE

Certified that this Report titled "**LUNG SENSE PRO: PREDICTIVE MODELLING FOR LUNG CANCER ANALYSIS USING MACHINE LEARNING**" is the bonafide work of **BATHALA PRAVEEN (723920104009), CHUNCHU MANOJ (723920104015), SASIDEVAN P (723920104049),** who carried out the work under my supervision. Certified further that to the best of my knowledge the work reported here in does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

SIGNATURE
Mr. S. SATHEESH M.E.,
HEAD OF THE DEPARTMENT,
Assistant Professor**,**
Department of CSE**,**
Arjun College of Technology**,**
Coimbatore -642 120

SIGNATURE
Ms. R. LATHA PRIYADHARSHINI M.E.,
SUPERVISOR,
Assistant Professor.
Department of CSE,
Arjun College of Technology,
Coimbatore -642 120

Submitted for the university project viva-voice held on _______________

INTERNAL EXAMINER **EXTERNAL EXAMINER**

ACKNOWLEDGEMENT

We owe our sincere and heartful thanks to our chairman **Thiru. R. SURIYANARAYANAN,** and also we extend our profound thanks to our Secretary **Dr. R. SURESH KUMAR M.E., Ph.D.** for their exuberance in motivating young minds.

Our deepest gratitude and thanks to our motivator and Principal **Dr. N. JANAKI MANOHAR M.E., Ph.D.** who always helping us whenever we approach him during the course of our project.

We would also like to express our profound thanks to our Head of the Department **Mr. S. SATHEESH M.E.** Assistant Professor, Department of CSE, whose thoughtful words, advise and help to complete our project successfully.

We would also like to express our unbounded gratefulness to our Project Coordinator **Mr. S. SATHEESH M.E.** Assistant Professor, Department of CSE, for his constant support and offered facilities for the completion of the project.

Our sincere gratitude and unplumbed thanks to our beloved Project Guide **Ms. R. LATHA PRIYADHARSHINI M.E.** Assistant Professor, Department of CSE, for her constant encouragement, Valuable Guidance and constructive criticism in making this project a successful one.

We express our sincere thanks to all **Faculty Members and Skilled Assistants** of Computer Science and Engineering Department and our lovable **Friends** for their help and wishes for the successful completion of this project.

Finally, yet importantly, we would like to express our indebtedness to our beloved **Parents** for their affectionate blessing co-operation at all stages of this academic venture and also our well-wishers.

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 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 the CT scans to determine the type 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. All of these functionalities are combined and provided in the form of a web application. This project is useful for the early detection of lung cancer in individuals and can help them in overcoming these health conditions.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVATIONS

AI	ARTIFICIAL INTELLIGENCE
ML	MACHINE LEARNING
DL	DEEP LEARNING
CNN	CONVOLUTION NEURAL NETWORK
CT Scan	COMPUTE TOMOGRAPHY SCAN
MRI	MAGNETIC RESONANCE IMAGING
PET	POSITRON EMISSION TOMOGRAPHY
3D	3 DIMENSIONAL
CMixNet	CUSTOMIZED MIXED LINK NETWORK
KNN	K-NEAREST NEIGHBOURS
RFC	RANDOM FOREST CLASSIFIER
SVC	SUPPORT VECTOR CLASSIFIER
DTC	DECISION TREE CLASSIFIER
LR	LINEAR REGRESSION
RFR	RANDOM FOREST REGRESSOR

CHAPTER 1
INTRODUCTION

Lung cancer is repeatedly identified as one of the deadliest diseases in the history of humankind. 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. The number of people affected by cancer is expected to continue to rise, reaching around 17 million by 2030. Early detection can aid in treatment.

1.1 STATEMENT OF PROBLEM

Lung cancer is difficult to diagnose since symptoms only appear in the latter stages, and it is really hard to save a person's life at this stage. A single scan can provide up to 500 sections and it takes an experienced radiologist about 2–3.5 minutes to observe each section. Hence, a better mechanism is required. The proposed project aims to develop a web application which can eye help in the initial stages of lung cancer diagnosis and can be scaled up further for other diseases using Deep Learning. The proposed model can detect lung cancer at initial stage based on the patient’s symptoms and can interpret CT (Computed tomography) images to identify nodules with diameters as tiny as 3mm which are unlikely to be identified by a radiologist with the naked eye.

1.2 SIGNIFICANT OF THE STUDY

The abstract highlights a groundbreaking study in lung cancer detection and treatment. Utilizing machine learning and deep learning algorithms, it focuses on early detection, crucial for boosting patient survival rates. By integrating patient details, symptoms, and CT scan data, the study enables precise assessments and personalized treatment plans. Notably, it targets the detection of small lung nodules, as tiny as 3 mm, aiding in early-stage diagnosis when treatments are most effective. Moreover, a machine learning model estimates medical insurance costs, offering practical benefits for informed healthcare decisions and potentially easing financial burdens. The development of a web application streamlines diagnostics, benefiting both healthcare professionals and patients. Overall, this study's innovative use of technology and comprehensive approach holds promise for enhancing patient outcomes and healthcare efficiency in lung cancer management.

1.3 SCOPE OF STUDY

The study's scope includes revolutionizing early lung cancer detection using machine learning and deep learning algorithms, targeting small nodules as tiny as 3 mm. It also estimates medical insurance costs and develops a user-friendly web application for streamlined diagnostics. The study aims to empower patients with informed healthcare decisions, improve outcomes, and alleviate the burden of lung cancer on healthcare systems.

1.4 LIMITATION OF THE STUDY

a) **Limited Generalizability:** The study's findings may lack broad applicability if the dataset used for training is biased or lacks diversity, potentially leading to suboptimal performance for patients from different demographics or regions.

b) **Diagnostic Accuracy Concerns:** Despite advancements, there's still a risk of diagnostic inaccuracies, including false positives and false negatives. This could result in misdiagnoses or missed opportunities for early intervention, underscoring the need for ongoing refinement and validation of the algorithms.

1.5 OBJECTIVE

The main objective of this project is to forecast the existence of lung cancer using medical records and interpreting CT images, employing advanced machine learning and deep learning techniques. This project seeks to enhance early detection rates and improve patient outcomes through timely intervention and personalized healthcare approaches.

CHAPTER 2

LITERATURE SURVEY

TITLE: Support Vector Machine based Lung Cancer Prediction

AUTHOR: Muthazhagan B, et al.

YEAR: 2021

DESCRIPTION: In their 2021 study, Muthazhagan B. et al. pioneered a novel approach to lung cancer prediction utilizing Support Vector Machine (SVM) technology. With an outstanding accuracy rate of 98%, their model showcased remarkable efficacy in distinguishing between 'abnormal' and 'normal' lung images. By harnessing the power of image classification, the SVM-based system represents a significant leap forward in the realm of early cancer detection. While the model's current iteration lacks the ability to differentiate between specific cancer stages, its success underscores the immense potential of machine learning in transforming healthcare practices, paving the way for more precise diagnoses and improved patient outcomes.

TITLE: CNN-Based Lung Cancer Classification

AUTHOR: Masud M, Sikder N, et al.

YEAR: 2021

DESCRIPTION: This study introduced a Convolutional Neural Network (CNN) model achieving an impressive accuracy of 96.33% in classifying lung cancer into five distinct categories: colon adenocarcinomas, benign colonic tissues, lung adenocarcinomas, lung squamous cell carcinomas, and benign lung tissues. While demonstrating considerable success, the authors aimed to further enhance performance, particularly for two of the five classes. This research signifies a significant step forward in leveraging deep learning techniques for comprehensive lung cancer classification. This achievement not only underscores the potential of Convolutional Neural Networks (CNNs) in medical image analysis but also highlights the importance of continuous refinement in algorithmic performance for precise diagnostic accuracy. By addressing specific challenges within the classification task, such as improving differentiation between the targeted categories, this study paves the way for more refined and tailored approaches in lung cancer diagnosis, ultimately contributing to more effective treatment strategies and better patient outcomes.

TITLE: Lung Cancer Detection using Google-Net

AUTHOR: Sajja T, Devarapalli R, et al.

YEAR: 2019

DESCRIPTION: In their groundbreaking 2019 study, Sajja T, Devarapalli R, et al. embarked on an ambitious quest to revolutionize lung cancer detection, employing the formidable Google-Net architecture as their cornerstone. Their research odyssey navigated through the intricate maze of optimizing dropout ratios, illuminating the pivotal role of such meticulous fine-tuning in fortifying the accuracy and dependability of detection systems. By harnessing the cutting-edge capabilities of neural network architecture, the study not only underscored the evolving landscape of optimization techniques but also emphasized their indispensability in refining the precision of lung cancer identification within the realm of medical imaging. This exhaustive investigation not only highlights the relentless pursuit of advanced deep learning models but also underscores the urgent need for robust methodologies aimed at enhancing the efficacy of medical imaging analysis, thereby making significant strides toward the timely and accurate diagnosis of lung cancer.

TITLE: Image Segmentation Techniques for Lung Cancer Detection

AUTHOR: Tripathi P, Tyagi S, et al.

YEAR: 2019

DESCRIPTION: Tripathi P, Tyagi S, et al.'s seminal 2019 inquiry embarked boldly into the intricate domain of image segmentation techniques tailored explicitly for lung cancer detection. Through a meticulous process of scrutiny and comparative analysis of various segmentation methodologies, the study advocated fervently for marker-controlled watershed segmentation as an unparalleled approach in accurately delineating lung cancer pathology from medical images. By emphasizing the paramount importance of meticulously segmenting lung structures, the authors not only provided invaluable insights into the nuanced development of robust image analysis algorithms but also laid the groundwork for achieving precise lung cancer diagnosis.

TITLE: 3D CNN for Lung Cancer Nodule Detection

AUTHOR: Nasrullah Nasrullah et al.

YEAR: 2019

DESCRIPTION: In their groundbreaking research, Nasrullah Nasrullah et al. pioneered the application of 3D Convolutional Neural Networks (CNNs) for the detection of lung cancer nodules, a crucial aspect of early diagnosis and treatment. Their innovative approach achieved an impressive Free-Response Receiver

Operating Characteristic (FROC) score of 94.21%, demonstrating the efficacy of advanced machine learning techniques in medical imaging analysis. By integrating Gradient Boosting Machine methods, they optimized the predictive capabilities of their 3D CNN architecture, offering a promising avenue for improving the accuracy and efficiency of lung cancer diagnosis. This study represents a significant leap forward in leveraging deep learning for medical image analysis and holds immense potential for enhancing patient outcomes in clinical settings.

TITLE: Deep Residual Learning for Lung Cancer Detection

AUTHOR: Siddharth Bhatia et al.

YEAR: 2019

DESCRIPTION: Siddharth Bhatia et al. introduced a groundbreaking methodology for lung cancer detection based on deep residual learning principles, marking a significant advancement in the field of medical image analysis. Through the fusion of Random Forest and XGBoost classifiers, their approach achieved an impressive accuracy rate of 84%, showcasing the potential of integrating deep learning techniques with traditional machine learning algorithms. By effectively capturing complex features inherent in lung images, their model demonstrated robust performance in identifying cancerous patterns, thereby facilitating early detection and intervention.

CHAPTER 3

SYSTEM ANALYSIS

System analysis involves a comprehensive examination of the lung cancer detection system to understand its existing structure and identify areas for improvement. This chapter delves into the analysis of both the existing system and the proposed system for lung cancer detection.

3.1 EXISTING SYSTEM ANALYSIS

The current methods for lung cancer detection primarily rely on traditional diagnostic techniques such as MRI, isotopes, X-rays, and CT scans. While these methods are effective to some extent, they have limitations, especially in detecting small nodules that may indicate early-stage cancer.

3.1.1 Limitations of the Existing System

a) **Limited Accuracy:** Traditional diagnostic methods may fail to detect small nodules or early-stage lung cancer accurately.

b) **Subjectivity:** Manual interpretation of diagnostic results may vary depending on the expertise of the healthcare professional, leading to inconsistencies in diagnosis.

c) **Time-Consuming:** The process of analyzing diagnostic images and patient data manually can be time-consuming, delaying the diagnosis and treatment initiation.

3.2 PROPOSED SYSTEM ANALYSIS

The proposed lung cancer detection system aims to address the limitations of the existing system by leveraging machine learning and deep learning algorithms to improve accuracy, efficiency, and accessibility.

3.2.1 Key Features of the Proposed System:

a) **Machine Learning-Based Prediction:** Utilizing historical patient data and symptoms, machine learning algorithms can predict the likelihood of lung cancer development in individuals.

b) **Deep Learning-Based Image Analysis:** Deep learning models analyze CT scan images to detect and classify lung nodules, enabling early-stage cancer detection with high accuracy.

c) **Automated Risk Assessment:** The system automatically calculates the patient's estimated medical insurance costs based on their cancer risk status and other relevant factors.

d) **Web Application Interface:** All functionalities are integrated into a user-friendly web application, allowing healthcare professionals and patients to access the system easily.

3.2.2 Benefits of the Proposed System

a) **Early Detection:** By detecting lung cancer at an early stage, the system improves patients' chances of survival and facilitates timely treatment initiation.

b) **Accuracy and Consistency:** Machine learning and deep learning algorithms offer more accurate and consistent results compared to manual interpretation, reducing diagnostic errors.

c) **Cost-Efficiency:** The proposed system optimizes healthcare resources by streamlining diagnostic procedures and reducing unnecessary costs associated with traditional methods.

d) **Accessibility:** The web-based interface makes the system accessible to healthcare professionals and patients, regardless of their geographical location.

3.2.3 Challenges and Considerations

a) **Data Quality:** Ensuring the quality and reliability of the input data is crucial for the performance of machine learning and deep learning models.

b) **Regulatory Compliance**: Adhering to regulatory requirements and standards for medical software development and data privacy is essential to ensure patient safety and data security.

c) **Integration with Existing Systems:** Seamless integration with existing healthcare IT infrastructure and electronic health record systems may be required to maximize the system's utility and interoperability.

3.3 SYSTEM REQUIREMENTS

The system requirements outline the functional and non-functional specifications necessary for the development and implementation of the proposed lung cancer detection system. These requirements serve as the foundation for system design and development efforts.

3.3.1 Functional Requirements

a) **Patient Data Input:** Allow healthcare professionals to input patient demographic data, medical history, and symptoms into the system.

b) **Diagnostic Image Upload:** Enable the uploading and storage of CT scan images for analysis by deep learning algorithms.

c) **Machine Learning Prediction:** Implement machine learning algorithms to predict the likelihood of lung cancer based on patient data and historical records.

d) **Deep Learning Image Analysis:** Develop deep learning models capable of analyzing CT scan images to detect and classify lung nodules.

e) **Risk Assessment:** Automatically calculate and display the patient's estimated medical insurance costs based on their cancer risk status.

f) **Reporting and Visualization:** Generate comprehensive reports and visualizations of diagnostic results for healthcare professionals and patients.

3.3.2 Non-Functional Requirements

a) **Accuracy:** Ensure the system achieves high accuracy in predicting lung cancer risk and analyzing CT scan images.

b) **Scalability:** Design the system to handle large volumes of patient data and diagnostic images efficiently as the user base grows.

c) **Usability:** Develop a user-friendly interface that is intuitive and easy to navigate for healthcare professionals and patients.

d) **Performance:** Optimize system performance to minimize processing time for machine learning and deep learning algorithms.

e) **Security:** Implement robust security measures to protect patient data confidentiality and prevent unauthorized access or data breaches.

CHAPTER 4

SYSTEM REQUIREMENTS

4.1 SYSTEM REQUIREMENTS.

This is what is required to run this software. The system requirements are separated into three phase which are hardware, software and personal requirement of the system.

4.2 HARDWARE REQUIREMENTS & SOFTWARE REQUIREMENT

4.2.1 HARDWARE REQUIREMENTS

Pentium IV computer with following configurations;

- I3 Processor GHz
- 4 GB RAM or more
- 450 GB Hard Disk
- Logitech Mouse
- Logitech Keyboard

4.2.2 SOFTWARE REQUIREMENT

- PHP, HTML, CSS, PYTHON Programming language
- TensorFlow, PyTorch, Keras, NumPy, Pandas, Matplotlib and Seaborn, Scikit-learn are Libraries & Frameworks needed
- Development Environment: Visual Studio Code.

4.3 FUNCTIONAL REQUIREMENTS

The Functional requirements for the proposed system are:

- The user can upload his age, height, weight, and other details for lung cancer prediction based on symptoms.

- A user's CT scan can be uploaded for lung cancer classification to know if the tumor is benign, malignant or no tumor.
- A CT scan of the user can be uploaded for lung nodule detection.
- The user can acquire a diagnosis and learn how advanced his/her lung cancer.
- A cost estimate for the user's medical insurance is available.

4.4 NON - FUNCTIONAL REQUIREMENTS

4.4.1 Usability:

Ensuring a user-friendly experience is essential for the website's effectiveness among patients and medical professionals. It should feature intuitive design and easy navigation, minimizing clutter. User feedback is crucial for addressing any usability issues, while compatibility across devices enhances accessibility.

4.4.2 Reliability:

Reliability is paramount in lung cancer detection and classification, where precision is crucial. Even the slightest error could have life-threatening consequences for the patient. Therefore, the system must be meticulously designed and rigorously tested to minimize the risk of misdiagnosis or false results. Accuracy and consistency are non-negotiable when it comes to medical applications, underscoring the importance of robust algorithms and quality assurance measures to ensure reliable performance at all times.

4.4.3 Performance:

Performance is a critical aspect of the lung cancer detection system, requiring both efficiency and accuracy. When dealing with new datasets, it's imperative that the training time is relatively short to facilitate timely analysis and decision-making. Additionally, the classification process must yield correct results consistently, with a minimal margin of error. This necessitates the use of efficient algorithms and optimization techniques to ensure swift and accurate predictions, thereby enhancing the system's overall performance and utility in clinical settings.

4.4.4 Supportability:

Supportability of the website is essential for ensuring its accessibility and usability across different browsers. It should be designed to function properly in the latest versions of popular browsers such as Google Chrome and Mozilla Firefox. This ensures that users can access the website seamlessly regardless of their browser preference, thereby maximizing its reach and effectiveness. Regular updates and testing should be conducted to maintain compatibility with evolving browser technologies and standards, ensuring a consistent user experience for all visitors.

4.4.5 Data set requirements:

The better the training and testing phases may be, the more CT scans and medical data there are available. The Non-Functional requirements for the proposed system are:

- The system uses the patient's information to anticipate lung cancer.
- The system classifies the CT scan images into normal, benign and malignant.

CHAPTER 5

SYSTEM DESIGN

5.1 INSTALLATION

To establish the project environment, adhere to the following steps:

- Export the project folder to the designated directory:
- For local development: Select a directory conducive to your workflow.
- For deployment on a web server: Adhere to the server's directory structure guidelines.
- Ensure the requisite software and libraries are installed, encompassing PHP, HTML, CSS, and Python.
- Employ Visual Studio Code as the primary development environment for streamlined workflow and enhanced productivity.

5.2 USER GUIDE

For seamless local usage:

- Deploy the project folder to your preferred directory.
- Verify the presence of essential software and libraries.
- Access the application through the browser by initiating the HTML files.
- Follow any provided installation scripts or dependency management tools.
- Seek additional assistance from community forums or support channels if needed.

5.3 INITIATING THE APPLICATION ONLINE

For online accessibility:

- Host the application on a web server to enable remote access.
- Facilitate user interaction via the application's designated URL.

5.4 SYSTEM TESTING

Conduct comprehensive testing to ensure optimal performance:

- Verify functionality across diverse datasets to uphold standards.
- Evaluate model accuracy and reliability.
- Conduct security testing to identify vulnerabilities.
- Validate responsiveness across multiple devices.
- Implement regression testing to ensure updates don't impact existing functionality.
- Gather user feedback for continuous improvement post-launch.

5.5 USER TRAINING

Facilitate user proficiency through comprehensive training sessions:

- Equip medical professionals with skills to interpret predictions and integrate them into patient care effectively.
- Empower data scientists with knowledge of machine learning techniques and model evaluation methodologies.
- Provide personalized training sessions based on the specific needs and expertise levels of users, ensuring maximum comprehension and proficiency.
- Offer certification programs or assessments to validate user skills and knowledge, incentivizing continued engagement and mastery of the prediction system.

5.6 SYSTEM DOCUMENTATION

Ensure thorough documentation encapsulating system design, functionality, and implementation details:

- Provide a comprehensive overview of the system's architecture, facilitating a nuanced understanding of its operation.

5.7 PROGRAM DOCUMENTATION

Document the program's intricacies to ensure clarity and transparency:

- Leverage the Python programming language for the lung cancer prediction model.
- Harness prominent libraries and frameworks such as TensorFlow, PyTorch, Keras, NumPy, Pandas, Matplotlib, Seaborn, and Scikit-learn for enhanced machine learning capabilities.
- Utilize Visual Studio Code as the preferred development environment for streamlined workflow management.

5.8 TECHNICAL DETAILS

Furnish technical specifications catering to proficient users in programming and machine learning:

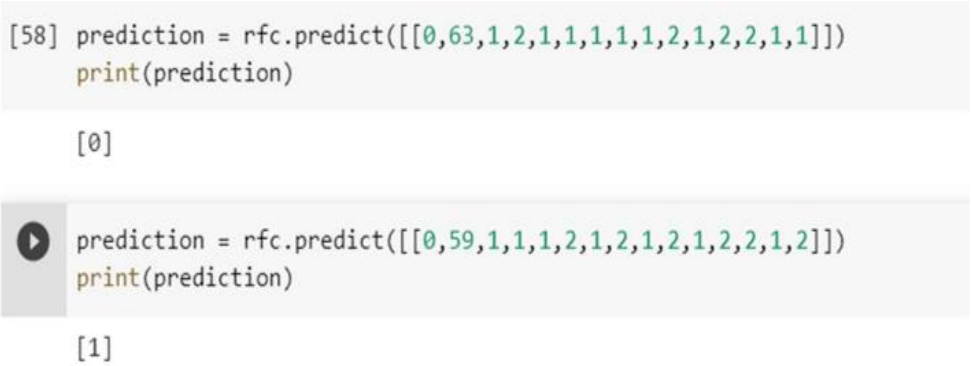
- Provide comprehensive documentation outlining the system architecture and data processing pipelines.
- Offer guidance on data preprocessing techniques and model optimization strategies.
- Include code snippets and sample datasets for implementation.
- Conduct tutorials covering advanced topics like hyperparameter tuning and model interpretation.
- Establish a dedicated support channel for user assistance and collaboration.

CHAPTER 6

TESTING AND MAINTENANCE

6.1 LUNG CANCER PREDICTION ON THE BASIS OF SYMPTOMS

A prediction with values from dataset to test if prediction is right or wrong provided 100% correct results. Fig 6.1 shows the result of the predictions.



```
[58] prediction = rfc.predict([[0,63,1,2,1,1,1,1,1,2,1,2,2,1,1]])
     print(prediction)

     [0]

     prediction = rfc.predict([[0,59,1,1,1,2,1,2,1,2,1,2,2,1,2]])
     print(prediction)

     [1]
```

Figure 6.1 Results for Symptom based predictions

The models used provided variable scores but RFC and SVC proved to be the best as shown in Figure 6.2.

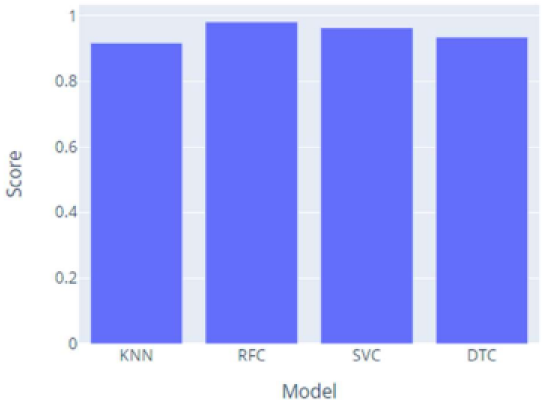
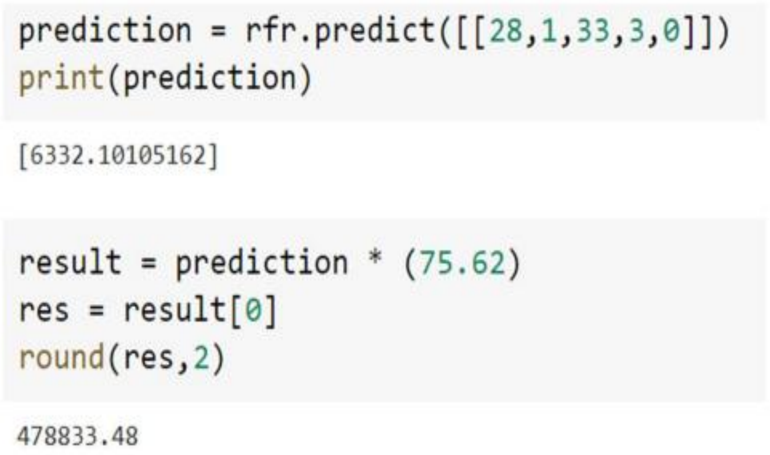


Figure 6.2 Scores of all models for Lung Cancer Prediction

6.2 INSURANCE PREDICTION

A prediction with values from dataset to test if prediction is right or wrong provided 100% correct results.



```
prediction = rfr.predict([[28,1,33,3,0]])
print(prediction)
```

```
[6332.10105162]
```

```
result = prediction * (75.62)
res = result[0]
round(res,2)
```

```
478833.48
```

Figure 6.3 Results for Insurance cost prediction

6.3 LUNG CANCER CLASSIFICATION

If the accuracy is high and the loss is low, then the model makes small errors on just some of the data, which would be the ideal case. The plots show the loss at and accuracy of the CNN model on both training and testing sets.

A Confusion matrix is an N x N matrix used for evaluating the performance of a classification model. The matrix compares the actual target values with those predicted by the machine learning model. This gives us a holistic view of how well our classification model is performing and what kinds of errors it is making. Figure 6.6 shows the confusion matrix of the CNN Model.

A classification report is a performance evaluation metric in machine learning. It is used to show the precision, recall, and F1 Score of your trained classification model. Figure 6.5 shows the trained CNN model’s classification report.

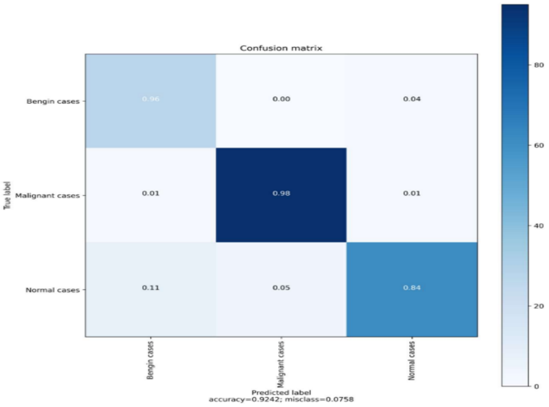


Figure 6.4 CNN Model Confusion Matrix

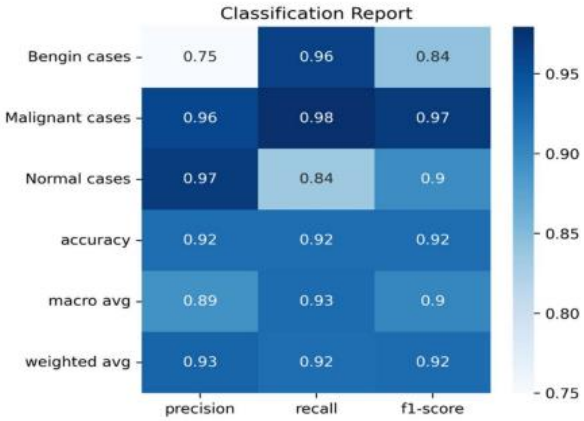


Figure 6.5 CNN Model Classification Report

6.4 LUNG CANCER NODULE DETECTION USING DEEP LEARNING

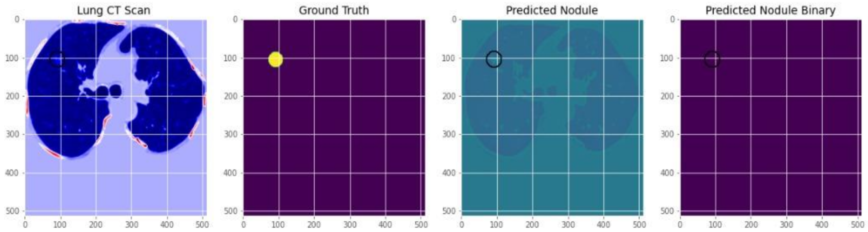


Figure 6.6 Output image of detected nodule

6.5 COMPARISON OF PROPOSED MODELS

The accuracy of the proposed model is compared with different models from previously done experiments. The proposed models outperformed the others in terms of performance. The proposed UNet with an accuracy of 98% outperformed CNN and ResNet models for nodule detection. The classification of CT Scans had accuracy upto about 90% in previous done research which was also outperformed with our proposed CNN model by achieveing an accuracy of 92.42%.

CHAPTER 7

SOFTWARE DESCRIPTION

Implementation involves the realization of a design or idea using specific tools and technologies. The new system has been implemented using a combination of Dreamweaver, PHP, MySQL database, HTML, CSS, and JavaScript. These technologies were chosen for their ease of development, flexibility, online capabilities, and provision of graphical user interfaces.

7.1 OPERATING SYSTEM

Windows

Windows operating system offers robust features suitable for businesses of all sizes. It provides standards-based security, manageability, and reliability, coupled with user-friendly features like Plug and Play and simplified interfaces. Windows enhances computing power while reducing ownership costs, making it an ideal choice for desktop computing in various business environments.

Linux

Linux is a family of open-source and free software operating systems renowned for their stability and flexibility. Based on the Linux kernel, Linux distributions (distros) include system software and libraries, often provided by the GNU Project. Popular distros such as Debian, Fedora, and Ubuntu offer desktop environments and support a wide range of hardware configurations. Linux emphasizes collaboration and community-driven development, providing users with diverse options for customization and deployment.

7.2 CHARACTERISTICS

Multiuser Capability

Both Linux and Windows support multiuser environments, allowing multiple users to access computer resources simultaneously from different terminals.

Multitasking

Linux and Windows are capable of multitasking, enabling the execution of multiple tasks concurrently for enhanced productivity.

Security

Linux and Windows prioritize security, offering features such as authentication and authorization to safeguard user data and prevent unauthorized access.

Communication

Operating systems facilitate communication between users within a network or across multiple networks, enabling the exchange of data, emails, and programs for seamless collaboration.

7.3 SOFTWARE FEATURES

MySQL

MySQL is a robust relational database management system (RDBMS) that provides efficient storage and retrieval of structured data. It offers multi-user access and supports SQL queries for data manipulation. MySQL is widely used for web applications, offering reliability, scalability, and performance.

PHP

PHP is a versatile server-side scripting language used for developing dynamic web pages and web applications. It integrates seamlessly with HTML and supports a wide range of databases. PHP is renowned for its flexibility, ease of integration, and extensive community support.

Python

Python is a high-level programming language known for its simplicity and readability. It offers powerful libraries and frameworks such as TensorFlow, PyTorch, Keras, NumPy, Pandas, Matplotlib, Seaborn, and Scikit-learn for machine learning tasks.

Development Environment

Visual Studio Code is the preferred development environment for coding and debugging PHP, HTML, CSS, and Python projects. It provides a lightweight yet powerful editor with built-in support for various languages and extensions, facilitating efficient software development.

CHAPTER 8

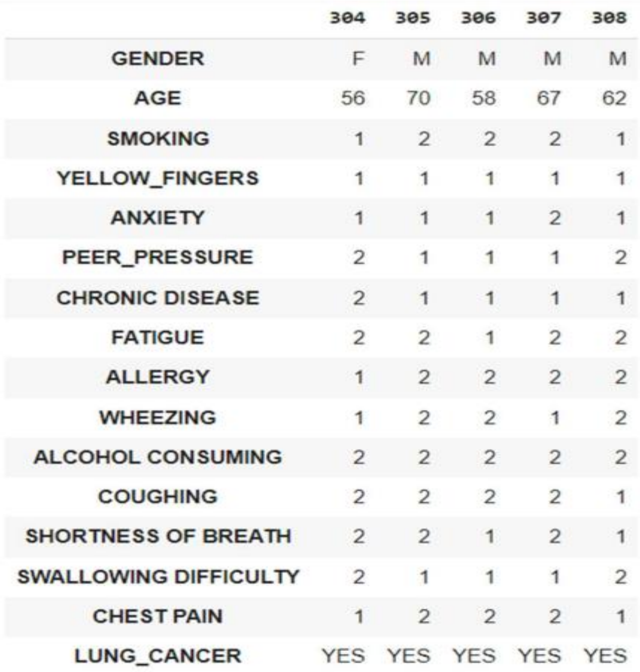
SYSTEM IMPLEMENTATION

8.1 LUNG CANCER DETECTION BASED ON SYMPTOMS

Various symptoms and habitual practices can lead to lung cancer. Using such data from users, we build models to predict if or not a patient has lung cancer.

Data Analysis

A thorough data analysis to be able to see the kind of data we are dealing with. The dataset used was the Survey Lung Cancer dataset collected from data.world website. The data is collected from the website online lung cancer prediction system and gets feedback from the user. This site was implemented during the period of August 2013 by the people who visited this site. The data has 309 records with 16 variable columns as shown in Figure 8.1.



	304	305	306	307	308
GENDER	F	M	M	M	M
AGE	56	70	58	67	62
SMOKING	1	2	2	2	1
YELLOW_FINGERS	1	1	1	1	1
ANXIETY	1	1	1	2	1
PEER_PRESSURE	2	1	1	1	2
CHRONIC DISEASE	2	1	1	1	1
FATIGUE	2	2	1	2	2
ALLERGY	1	2	2	2	2
WHEEZING	1	2	2	1	2
ALCOHOL CONSUMING	2	2	2	2	2
COUGHING	2	2	2	2	1
SHORTNESS OF BREATH	2	2	1	2	1
SWALLOWING DIFFICULTY	2	1	1	1	2
CHEST PAIN	1	2	2	2	1
LUNG_CANCER	YES	YES	YES	YES	YES

Figure 8.1 Preview of the Dataset

Data processing

Certain modifications which are to be done on the data are done in this stage. Renaming columns and encoding categorical data were some changes. Since the dataset was imbalanced, resampling was used to balance it. Synthetic Minority Oversampling Technique (SMOTE) was implemented as an oversampling technique to increase the number of cases in the dataset in a balanced way.

Data exploration

Count of males are more than females

- Mean age of males is 0.4 more than females
- Males smoke more than females
- Yellow fingers are more common among females
- Anxiety is also commonly found symptoms in females
- Peer pressure is also more for females
- Chronic disease is also more for females.

Further exploration was done with respect to all variables. The obtained results were used to create a dashboard page, shown in the following chapters.

Model Building

The creation of different samples for training and testing helps us evaluate model performance. Hence the split of our modelling dataset into training and testing samples is performed using the train_test_split() function of the scikit-learn library.

Following the data split, the train data is fed to various models in order to train them. The models used are:

- KNN: K-Nearest Neighbors
- RFC: Random Forest Classifier

- SVC: Support Vector Classifier
- DTC: Decision Tree Classifier

After training the models, they are made to predict. The predicted values are matched against the validation data to obtain accuracies of each of the models. The model with the highest accuracy was the RFC, which was chosen as the best model to predict the presence of cancer based on user's symptoms.

8.2 MEDICAL INSURANCE COST PREDICTION

The insurance dataset is used to build a model to be able to predict an approximate cost of insurance that a cancer patient will be in need for. The prediction will be based on the given user's details.

A thorough data analysis to be able to see the kind of data we are dealing with. The dataset used was the Medical Insurance Dataset collected from Kaggle. The data has 1338 records with 7 variable columns.

Model Building

The creation of different samples for training and testing helps us evaluate model performance. Hence the split of our modelling dataset into training and testing samples is performed using the train_test_split() function of the scikit-learn library. With about 1070 samples for training and 268 for testing, models were implemented.

Following the data split, the train data is fed to various models in order to train them. The models used are:

- Linear Regression (LR)
- Random Forest Regression (RFR)
- Decision Tree Regression
- Lasso Regression

The models are built to predict after they have been trained. The predicted values are compared to the validation data to determine the accuracy of each model. The RFR model had the highest accuracy and was chosen as the best model to forecast insurance rates based on user information.

8.3 LUNG CANCER CLASSIFICATION USING CT SCANS

Data Analysis

The Iraq-Oncology Teaching Hospital/National Center for Cancer Diseases (IQOTH/NCCD) lung cancer dataset was collected in the above-mentioned specialist hospitals over a period of three months in fall 2019. It includes CT scans of patients diagnosed with lung cancer in different stages, as well as healthy subjects. IQ-OTH/NCCD slides were marked by oncologists and radiologists in these two centers. The dataset contains a total of 1190 images representing CT scan slices of 110 cases. These cases are grouped into three classes: normal, benign, and malignant. of these, 40 cases are diagnosed as malignant; 15 cases diagnosed with benign; and 55 cases classified as normal cases. The 110 cases vary in gender, age, educational attainment, area of residence and living status.

Data processing and exploration

The image data is shuffled then viewed using colormaps like Figure 8.2 that uses hot cmap. The data is normalized, reshaped and encoded using the one hot encoding.

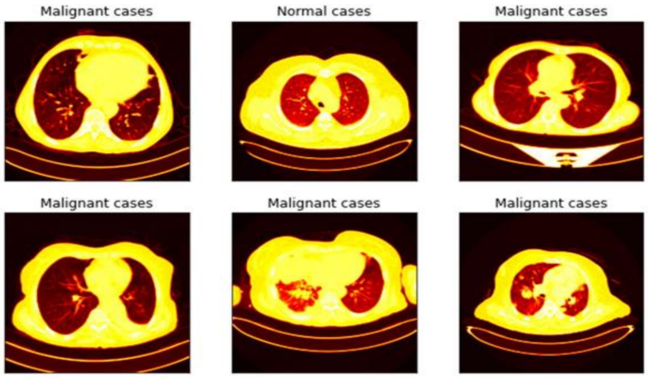


Figure 8.2 Image data with hot colormap

Model Building

A Convolutional Neural Network to predict the correct classes of cells from the images was created. We have used 3 Conv2D layers with MaxPool2D layers after each for the feature extraction from the images. The activation function used is ReLU. The output layer has only three neurons corresponding to the three classes of tumors (Benign, Malignant, Normal), with SoftMax activation function. Fig 8.10 shows the model summary.

Then, the model is compiled with Adam as the optimizer and Categorical Crossentropy as the loss function. We will train the model for 7 epochs with the class weights. The training was stopped at the 7th epoch as a good accuracy was obtained of about 92%

8.4 LUNG CANCER NODULE DETECTION USING DEEP LEARNING

Data Analysis

The dataset excluded scans with a slice thickness greater than 2.5 mm. In total, 888 CT scans are included. The LIDC/IDRI database also contains annotations which were collected during a two-phase annotation process using 4 experienced radiologists. Each radiologist marked lesions they identified as non-nodule, nodule

< 3 mm, and nodules >= 3 mm. See this publication for the details of the annotation process. The reference standard of challenge consists of all nodules >= 3 mm accepted by at least 3 out of 4 radiologists. Annotations that are not included in the reference standard (non-nodules, nodules < 3 mm, and nodules annotated by only 1 or 2 radiologists) are referred as irrelevant findings. The subset0 is a zip file which contains all CT images and annotations.csv file contains the annotations used as reference standard for the 'nodule detection' track.

In subset0, CT images are stored in MetaImage (mhd/raw) format. Each mhd file is stored with a separate raw binary file for the pixel data. The annotation file is a csv file that contains one finding per line. Each line holds the SeriesInstanceUID of the scan, the x, y, and z position of each finding in world coordinates and the corresponding diameter in mm. The annotation file contains 1186 nodules.

Data processing and exploration

Before inputting the CT images into the U-net architecture, it is important to reduce the domain size for more accurate results. A variety of preprocessing steps are performed to segment out the ROI (the lungs) from the surrounding regions of bones and fatty tissues. These include

- Binary Thresholding
- Selecting the two largest connected regions
- Erosion to separate nodules attached to blood vessels
- Dilation to keep nodules attached to the lung walls
- Filling holes by dilation
- Converting the mhd files to png

An index of the image file is captured and the directory of that index image is stored. After that, the image is opened, resized and converted into a grayscale image and its index is stored. The mask of the image corresponding to the index is also stored along with the grayscale image. After that, masks can be read by giving the directory of the mask. Finally, the mask image is preprocessed by resizing it and normalizing the pixel value then stored at the pre-processed mask image at the output array at the same index position. X[n] stores the image and y[n] stores the corresponding mask.

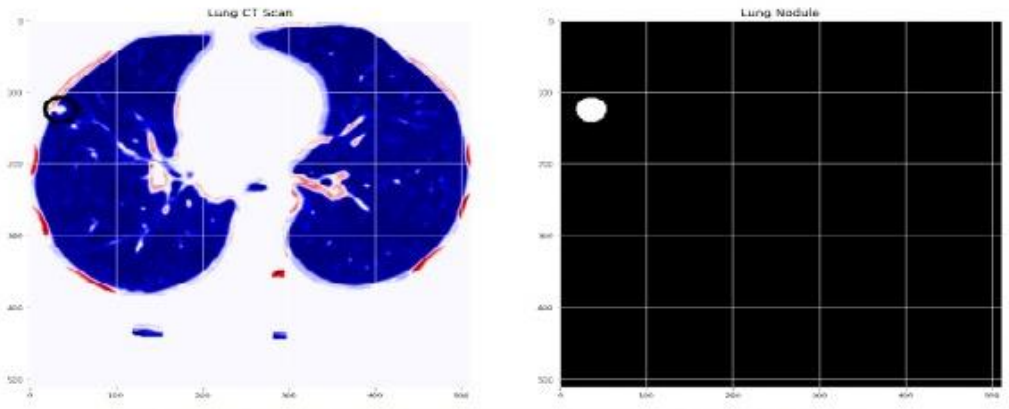


Figure 8.3 CT Scan and its corresponding lung nodule mask visualization

The plot of the pre-processed sample image and mask of the image displaying thenodule is shown above in Fig 8.3 The above image shows the pre-processed image as well as its mask.

CHAPTER 9

CONCLUSION, FUTURE ENHANCEMENT

9.1 CONCLUSION

The goal of this project was to advance the fight against lung cancer by providing a comprehensive solution that integrates diagnosis, prediction, and financial planning. Through the implementation of advanced techniques such as the Random Forest Classifier, CT scan prediction, and nodule staging, we aimed to empower healthcare providers with tools for early detection and accurate prognosis, thereby improving patient outcomes and enabling personalized treatment strategies. Additionally, our system's capability to estimate medical insurance costs aimed to provide essential financial support, easing the burden for patients and ensuring better preparedness for treatment expenses. In essence, the goal of this project was to make a meaningful impact in the lives of those affected by lung cancer, combining innovation with compassion to contribute to the larger goal of eradicating lung cancer.

9.2 FUTURE ENHANCEMENT

Further research and studies are to be conducted and validation of the proposed models of convolutional neural networks has to be performed. Verification of the presented models is necessary before they may be used in lung cancer screening procedures, enhancing the detection rate at an earlier stage. Models can also be enhanced by training with additional data from a wider range of scenarios. Also, consider employing multi-segmentation models with a high number of processors to train models to detect additional nodules. This can help train and predict output in less time.

CHAPTER 10

APPENDICES

10.1 SOURCE CODE

```
#homepage
<!doctype html>
<html>
<head>
  <style>
    body {
      background-color: white;
      font-family: arial, 'sans serif';
    }
    h1 {
      background: #009ffd;
      padding: 10px;
      text-align: center;
    }
    h3 {
      text-align: center;
      color: black;
      margin-top: 10px;
      font-size: large; }
    .grid-container {
      display: grid;
      grid-template-columns: auto auto auto auto;
      gap: 10px;
      background-color: #009ffd;
      padding: 10px; }
    .grid-container>div {
      background-color: white;
      text-align: center;
      padding: 20px}
    img {
      height: 30%;
      display: block;
      margin-left: auto;
      margin-right: auto;
      width: 30%;
      font-weight: bold;
      color: black;
      font-size: 1rem;
   a:hover {
      color: #009ffd;
      text-decoration: underline;
    }
  </style>
</head>
<body>
  <h1>Lung Cancer Analysis and Prediction</h1>
  <h3>What would you like to have a look at?</h3>
  <img src="https://dailynorthwestern.com/wp-content/uploads/2019/05/CANCER_Courtesy_WEB-900x600.jpg"
    alt="lung cancer" />
  <div class="grid-container">
    <div> <a href="symptoms">Lung Cancer Detection based on Symptoms</a></div>
    <div> <a href="stage">Lung Cancer Type Detection</a></div>
    <div> <a href="mainmedicalinsurance">Medical Insurance Calculation</a></div>
  </div>
</body>
</html>
```

#healthy.html

```
<!DOCTYPE html>
<html>
<head>
            vertical-align: unset;
            line-height: 18px;
            font-weight: 400;
            word-break: break-all;
        }
        .first-col {
            width: 25%;
            text-align: left;
        }
        .btn-block {
            margin-top: 20px;
            text-align: center;
        }
        button {
            width: 150px;
            padding: 10px;
            border: none;
            -webkit-border-radius: 5px;
            -moz-border-radius: 5px;
            border-radius: 5px;
            background-color: #009ffd;
            font-size: 16px;
            color: #fff;
            cursor: pointer;
        }
        button:hover {
            background-color: #009ffd;
        }
        @media (min-width: 568px) {
            .name {
                display: flex;
                justify-content: space-between;
            }
            .name input {
                width: 47%;
                margin-bottom: 0;
            }
            th,
            td {
                word-break: keep-all;
            }
                <span>
                    <p>
                    <h2 style="text-align: center;">{{prediction_text}}</h2><br><br>
                    {% with messages = get_flashed_messages() %}
                    {% if messages %}
                    <ul>
                        {% for message in messages %}
                        <li style="text-align: justify;position: relative;left: 5%; ">{{ message }}</li>
                        {% endfor %}
                    </ul>
                    {% endif %}
                    {% endwith %}
                  <div><br><href="{{url_for('unet')}}" style="position:relative;"><button>Insurance
                                  Calculation</button></a><br><br><br></div>
                    {% endif %}
                </span>
            </div>
        </div>
</body>
</html>
```

#predict.html

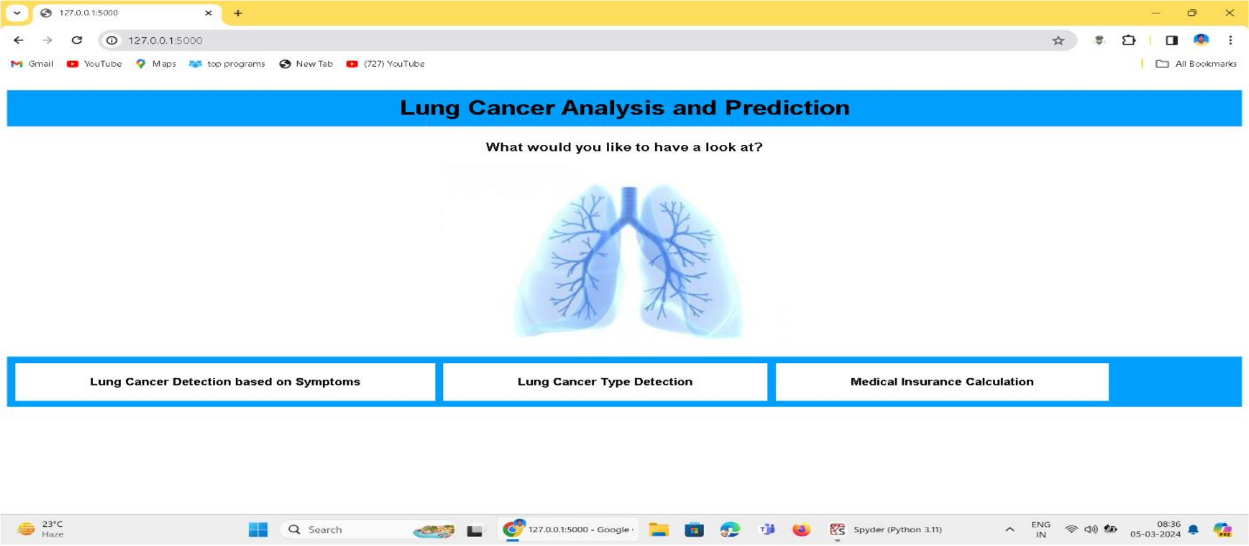
```
<!DOCTYPE html>
<html>
<head>
  <!-- Required meta tags -->
    <meta charset="utf-8">
    <meta name="viewport" content="width=device-width, initial-scale=1, shrink-to-fit=no">
<link rel="stylesheet"
href="https://maxcdn.bootstrapcdn.com/bootstrap/4.0.0/css/bootstrap.min.css"
integrity="sha384-
Gn5384xqQ1aoWXA+058RXPxPg6fy4IWvTNh0E263XmFcJlSAwiGgFAW/dAiS6JXm"
crossorigin="anonymous">
<style>
        body {
            margin: 4rem;
            background-color: #009ffd;
            box-shadow: -6px 13px 20px 0px rgba(0, 0, 0, 0.75);
        }
OUTPUT_FOLDER = os.path.join('static', 'output')
app.config['UPLOAD_FOLDER'] = OUTPUT_FOLDER
@app.route('/')
def index():
    return render_template('index.html')
@app.route('/symptoms')
def symptoms():
    return render_template('symptoms.html')
@app.route('/mainmedicalinsurance')
def mainmedicalinsurance():
    return render_template('mainmedicalinsurance.html')
@app.route('/lungcancer')
def lungcancer():
    return render_template('lungcancerdetection.html')
@app.route('/chart1//')
def chart1():
    data = pd.read_csv(r'C:\Users\lenovo\Desktop\Enhancing-Early-Detection-of-Lung-Cancer-Integrative-Analysis-and-Prediction-System,manoj ,praveen\Code\Web Application\output\dataset1.csv')
    trace0 = go.Histogram(x=data['GENDER'],name="Gender")
    trace1 = go.Histogram(x=data['AGE'],name="Age")
    trace2 = go.Histogram(x=data['SMOKING'],name="Smoking")
    trace3 = go.Histogram(x=data['YELLOW_FINGERS'],name="Yellow Fingers")
    trace5 = go.Histogram(x=data['ANXIETY'],name="Anxiety")
    trace6 = go.Histogram(x=data['PEER_PRESSURE'],name="Peer Pressure")
    trace7 = go.Histogram(x=data['CHRONIC_DISEASE'],name="Chronic Disease")
    trace8 = go.Histogram(x=data['FATIGUE'],name="Fatigue")
    trace9 = go.Histogram(x=data['ALLERGY'],name="Allergy")
    trace10 = go.Histogram(x=data['WHEEZING'],name="Wheezing")
    trace11= go.Histogram(x=data['ALCOHOL_CONSUMING'],name="Alcohol Consuming")
    trace12= go.Histogram(x=data['COUGHING'],name="Coughing")
    trace13= go.Histogram(x=data['SHORTNESS_OF_BREATH'],name="Shortness Of Breath")
    trace14= go.Histogram(x=data['SWALLOWING_DIFFICULTY'],name="Swallowing Difficulty")
    trace15= go.Histogram(x=data['CHEST_PAIN'],name="Chest Pain")
    trace16 = go.Histogram(x=data['LUNG_CANCER'],name="Lung Cancer")
    data1 = pd.read_csv(r'C:\Users\lenovo\Desktop\Enhancing-Early-Detection-of-Lung-Cancer-Integrative-Analysis-and-Prediction-System,manoj ,praveen\Code\Web Application\output\dataset2.csv')
    corrmat = data1.corr()
    trace17 = go.Heatmap( z = corrmat.values, x = list(corrmat.columns),y = list(corrmat.index),colorscale = 'Viridis',showscale=False)
    models = pd.read_csv(r'C:\Users\lenovo\Desktop\Enhancing-Early-Detection-of-Lung-Cancer-
Integrative-Analysis-and-Prediction-System,manoj ,praveen\Code\Web
Application\output\models.csv')
    trace18 = go.Bar(x=models['Model'], y= models['Score'],name="Model Comparison")
    fig = plotly.tools.make_subplots(
        rows=9,
        cols=3,
        specs=[[{}, {}, {}],[{},{},{}],[{}, {'colspan': 2, 'rowspan': 3}, None], [{} , None, None],[{}
, None, None],[{}, {}, {}],[{},{},{}],[{'colspan': 3, 'rowspan': 1},None, None],[{'colspan': 3,
'rowspan': 1},None, None]],
        subplot_titles=('Gender','Allergy', 'Smoking',"Yellow Fingers", "Anxiety","Peer
Pressure","Chronic Disease","Age","Fatigue","Lung Cancer","Wheezing","Alcohol
Consuming","Coughing","Shortness Of Breath","Swallowing Difficulty","Chest
Pain","Correlation Matrix"," Model Comparison")
    )
    fig.update_layout(width=1100,height=3000,title="Data Distribution, X-axis indicates 'Features'
and Y-axis indicates 'Count'")
    fig.append_trace(trace0, 1, 1)
    fig.append_trace(trace9, 1, 2)
    fig.append_trace(trace16, 5, 1)
    fig.append_trace(trace10, 6, 1)
    fig.append_trace(trace11, 6, 2)
    fig.append_trace(trace12, 6, 3)
 graphJSON = json.dumps(fig, cls=plotly.utils.PlotlyJSONEncoder)
    header="Lung Cancer Analysis"
    description="The Random Forest Classifer is the most accurate model for predicting Lung
Cancer"
    return render_template('notdash.html',
graphJSON=graphJSON,header=header,description=description)
@app.route('/chart2')
def chart2():
    data = pd.read_csv(r'C:\Users\lenovo\Desktop\Enhancing-Early-Detection-of-Lung-Cancer-
Integrative-Analysis-and-Prediction-System,manoj ,praveen\Code\Web
Application\output\insurance.csv')
    trace0 = go.Histogram(x=data['age'],name="AGE")
    trace1 = go.Histogram(x=data['sex'],name="SEX")
    trace2 = go.Histogram(x=data['bmi'],name="BMI")
    trace3 = go.Histogram(x=data['children'],name="CHILDREN"
    trace11 = go.Scatter(x=data['age'], y=data['bmi'], name="AGExBMI",mode='markers')
    trace12 = go.Scatter(x=data['age'], y=data['children'], name="AGExCHILDREN",
mode='markers')
    trace13= go.Scatter(x=data['age'], y=data['charges'], name="AGExCHARGES",
mode='markers')
    corrmat = data.corr()
    trace21= go.Heatmap( z = corrmat.values, x = list(corrmat.columns),y =
list(corrmat.index),colorscale = 'Viridis',showscale=False
```

```
    models = pd.read_csv(r'C:\Users\lenovo\Desktop\Enhancing-Early-Detection-of-Lung-Cancer-
Integrative-Analysis-and-Prediction-System,manoj ,praveen\Code\Web
Application\output\models2.csv')
    trace22 = go.Bar(x=models['Model'], y= models['Score'],name="Model Comparison",marker =
{'color' : 'teal'})

    fig = plotly.tools.make_subplots(
        rows=11,
        cols=3
specs=[[{'rowspan':2},{},{'rowspan':2}],[None,{},None],[{},{},{}],[{'colspan':3,'rowspan':3},N
one,None],[None,None,None],[None,None,None],[{},{},{}],[{},{},{}],[{'colspan':3,'rowspan':3
},None,None],[None,None,None],[None,None,None]],
subplot_titles=('CHARGES','SEX','BMI','CHILDREN','SMOKER','REGION','AGE','CORRELA
TION
MATRIX','AGExBMI','AGExCHILDREN','AGExCHARGES','BMIxCHARGES','CHILDRENx
CHARGES','SMOKERxCHARGES','MODELS COMPARISON'
    )
    fig.update_layout(width=1300,height=1300)
    fig.append_trace(trace0, 3, 3)
    fig.append_trace(trace1, 1, 2)
    fig.update_layout(bargap=0.2,title="Data Distribution, X-axis indicates 'Features' and Y-axis
indicates 'Count'")
    header="Medical Insurance for Lung Cancer"
    description="The Random Forest Regression is the best for obtaining medical insurance"
    graphJSON = json.dumps(fig, cls=plotly.utils.PlotlyJSONEncoder)
    return render_template('notdash.html',
graphJSON=graphJSON,header=header,description=description)

@app.route('/predicts',methods=['POST','GET'])
def predicts():
    int_features = [int(x) for x in request.form.values()]
```

10.2 SCREENSHOTS

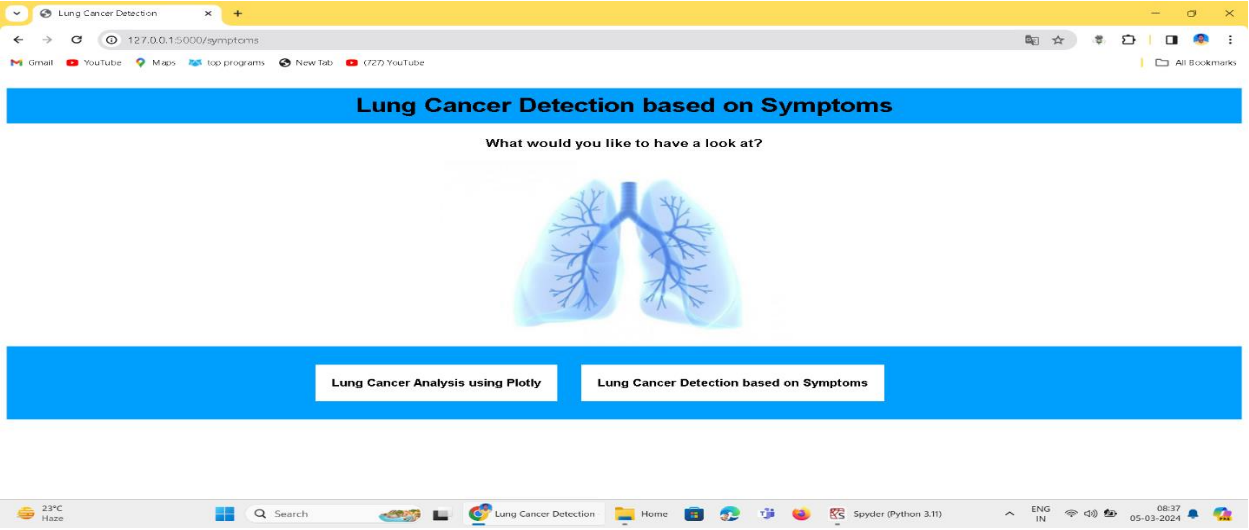


Lung Cancer Analysis and Prediction

What would you like to have a look at?

127.0.0.1:5000
Lung Cancer Detection based on Symptoms
Lung Cancer Type Detection
Medical Insurance Calculation

Figure 10.2.1 Home Page

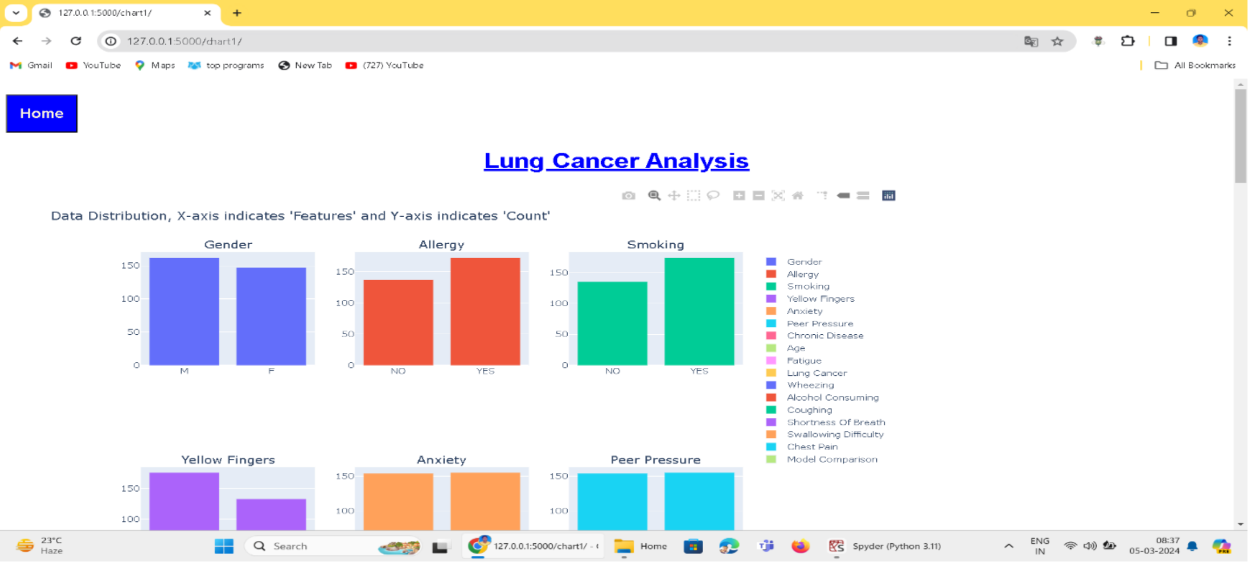


Lung Cancer Detection based on Symptoms

What would you like to have a look at?

127.0.0.1:5000/symptoms
Lung Cancer Analysis using Plotly
Lung Cancer Detection based on Symptoms

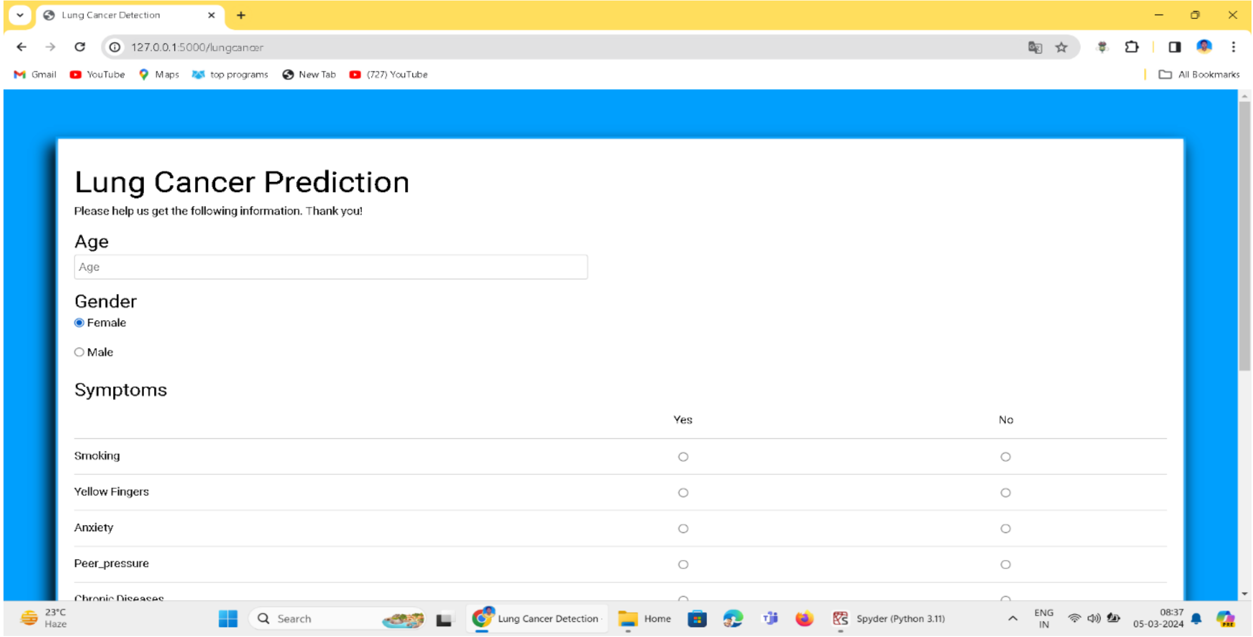
Figure 10.2.2 Lung Cancer Detection based on Symptoms



Lung Cancer Analysis

Data Distribution, X-axis indicates 'Features' and Y-axis indicates 'Count'

Figure 10.2.3 Lung Cancer Analysis using Plotly



Lung Cancer Prediction

Please help us get the following information. Thank you!

Age

Age

Gender

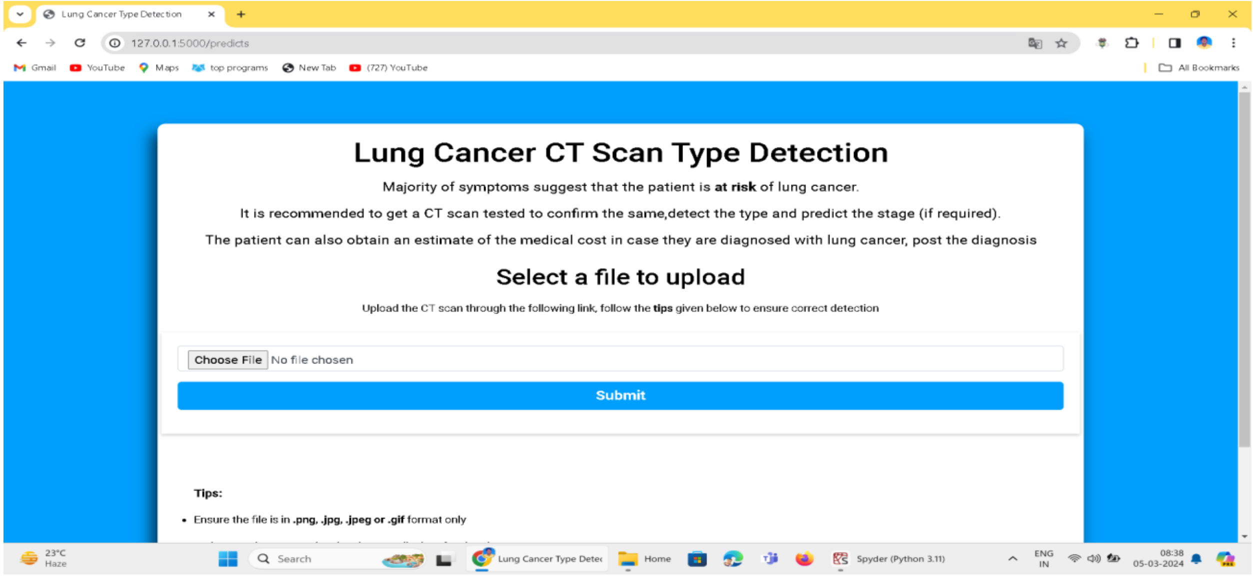
● Female

○ Male

Symptoms

	Yes	No
Smoking	○	○
Yellow Fingers	○	○
Anxiety	○	○
Peer_pressure	○	○
Chronic Diseases	○	○

Figure 10.2.4 Lung Cancer Based on Symptoms



Lung Cancer CT Scan Type Detection

Majority of symptoms suggest that the patient is **at risk** of lung cancer.

It is recommended to get a CT scan tested to confirm the same,detect the type and predict the stage (if required).

The patient can also obtain an estimate of the medical cost in case they are diagnosed with lung cancer, post the diagnosis

Select a file to upload

Upload the CT scan through the following link, follow the **tips** given below to ensure correct detection

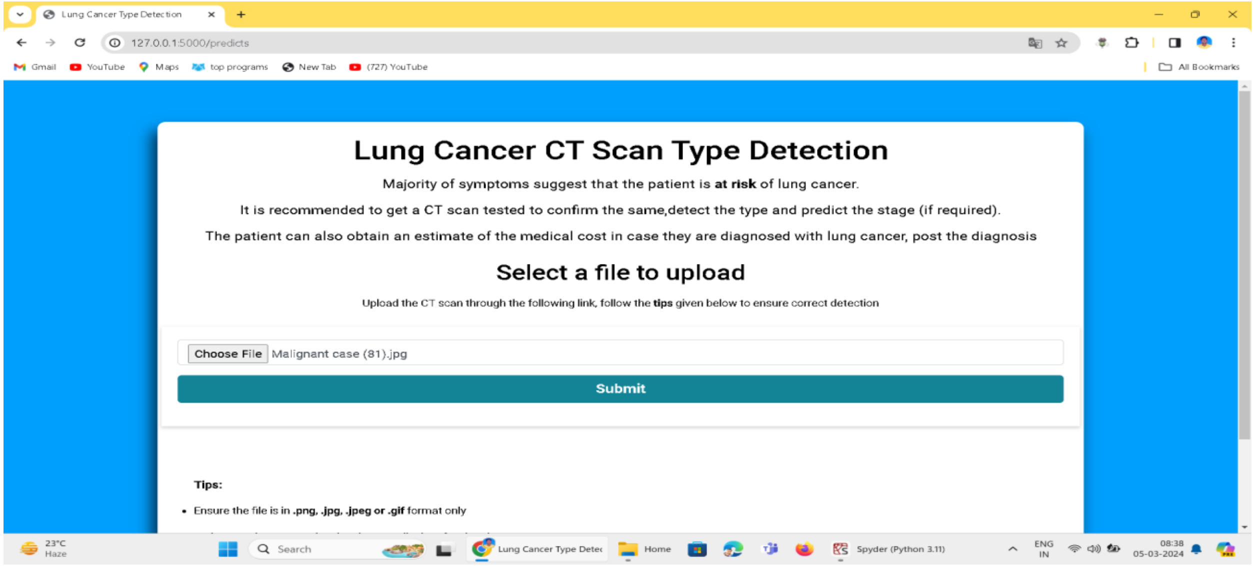
Choose File No file chosen

Submit

Tips:

- Ensure the file is in **.png**, **.jpg**, **.jpeg** or **.gif** format only

Figure 10.2.5 Type Detection Page



Lung Cancer CT Scan Type Detection

Majority of symptoms suggest that the patient is **at risk** of lung cancer.

It is recommended to get a CT scan tested to confirm the same,detect the type and predict the stage (if required).

The patient can also obtain an estimate of the medical cost in case they are diagnosed with lung cancer, post the diagnosis

Select a file to upload

Upload the CT scan through the following link, follow the **tips** given below to ensure correct detection

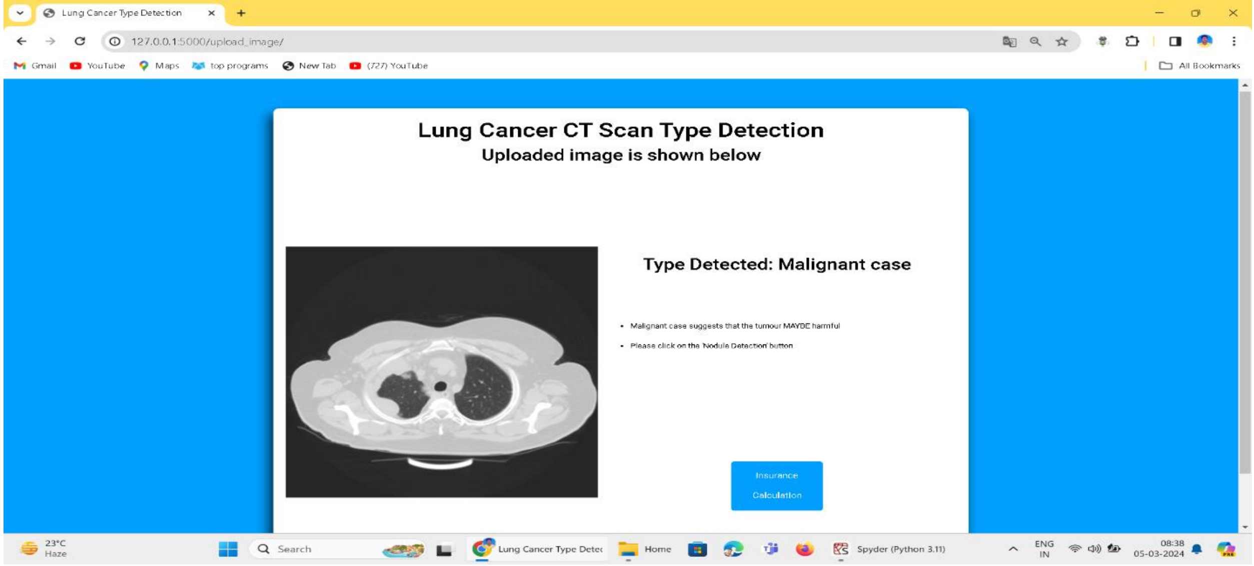
Choose File Malignant case (81).jpg

Submit

Tips:

- Ensure the file is in **.png**, **.jpg**, **.jpeg** or **.gif** format only

Figure 10.2.6 Type Detection: Uploading (Benign Case)



Lung Cancer CT Scan Type Detection

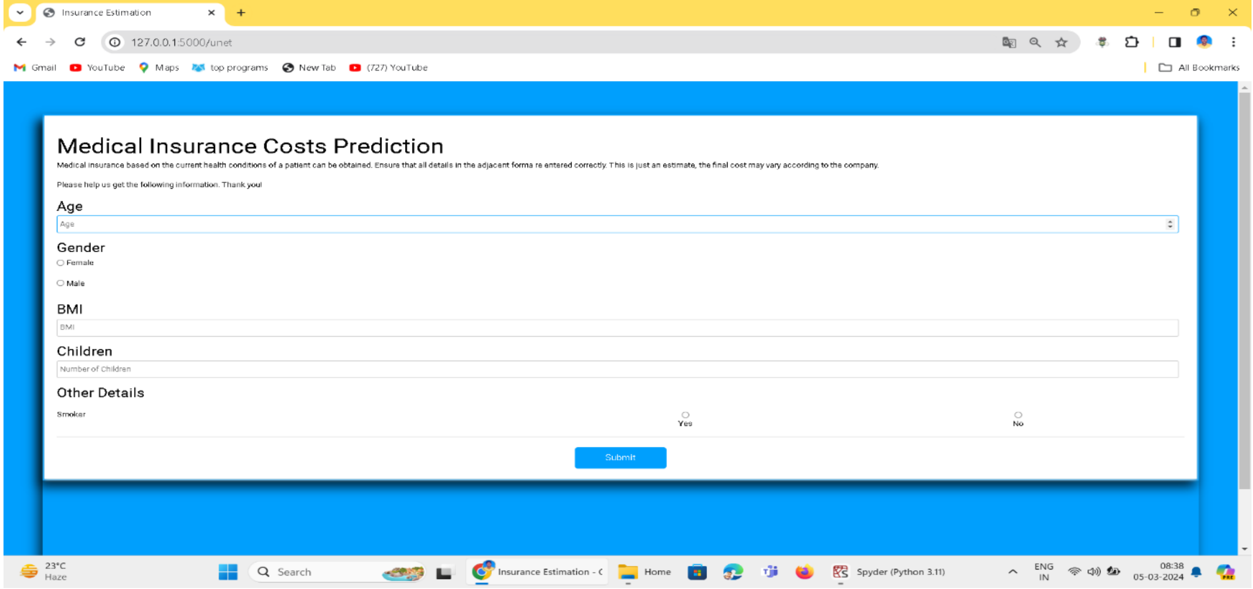
Uploaded image is shown below

Type Detected: Malignant case

- Malignant case suggests that the tumour MAYBE harmful
- Please click on the 'Nodule Detection' button

Insurance Calculation

Figure 10.2.7 Type Detection Output



Medical Insurance Costs Prediction

Medical insurance based on the current health conditions of a patient can be obtained. Ensure that all details in the adjacent forma re entered correctly. This is just an estimate, the final cost may vary according to the company.

Please help us get the following information. Thank you!

Age

Age

Gender

○ Female

○ Male

BMI

BMI

Children

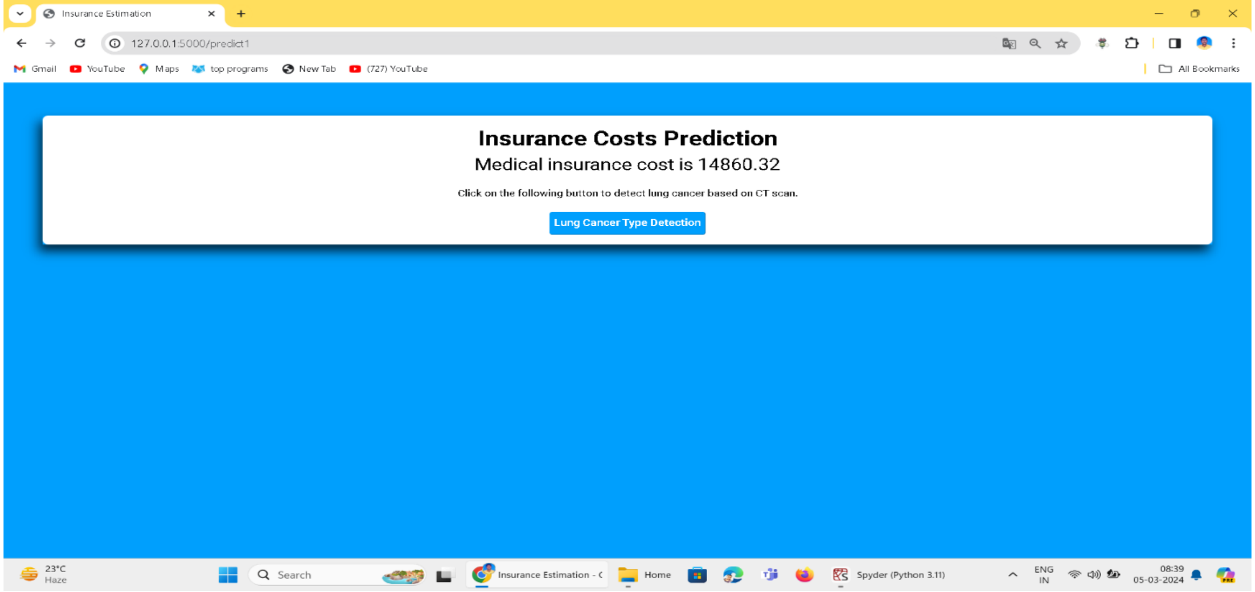
Number of Children

Other Details

Smoker ○ Yes ○ No

Submit

Figure 10.2.8 Medical Insurance Prediction



Insurance Costs Prediction

Medical insurance cost is 14860.32

Click on the following button to detect lung cancer based on CT scan.

Figure 10.2.9 Medical Insurance Output

REFERENCE

1. Ali, I., Hart, G. R., 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.
2. Bhatia, S., Sinha, Y., & Goel, L.: Lung cancer detection: a deep learning approach. In Soft Computing for Problem Solving. Springer, Singapore, pp. 699-705, 2019.
3. 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.
4. Hanafy, Mohamed: "Predict Health Insurance Cost by using Machine Learning and DNN Regression Models", International Journal of Innovative Technology and Exploring Engineering, Volume-10, p. 137, 2021.
5. Hosny, Ahmed, et al.: "Deep learning for lung cancer prognostication: a retrospective multi-cohort radiomics study." PLoS medicine, 15.11: e1002711, 2018.
6. 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.
7. Kadir, Timor, and Fergus Gleeson: "Lung cancer prediction using machine learning and advanced imaging techniques." Translational lung cancer research, 7.3: 304, 2018.
8. Lakshmanaprabu, S. K., et al.: "Optimal deep learning model for classification of lung cancer on CT images." Future Generation Computer Systems, 92: 374-382, 2019.
9. 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.

10. 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 classification framework." Sensors, 21(3):748, 2021.
11. Makaju, S., Prasad, PW., Alsadoon, A., Singh, AK., Elchouemi, A.: "Lung cancer detection using CT scan images." Procedia Computer Science, 1;125:107-14, 2018.
12. 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, 2:1-9, 2020.
13. 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, 19(17):3722, 2019.
14. Nasser, IM., & Abu-Naser, SS.: "Lung cancer detection using artificial neural network." International Journal of Engineering and Information Systems (IJEAIS), Mar;3(3):17-23, 2019.
15. Raoof, Syed Saba, M. A. Jabbar, and Syed Aley Fathima: "Lung Cancer prediction using machine learning: A comprehensive approach." 2nd International conference on innovative mechanisms for industry applications (ICIMIA). IEEE, 2020.
16. Sang, J., Alam, MS., 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.

17. 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, 201
18. Shin, Hyunku, et al.: "Early-stage lung cancer diagnosis by deep learning-based spectroscopic analysis of circulating exosomes." ACS nano, 14.5: 5435-5444, 2020.
19. Singh, Gur Amrit Pal, and P. K. Gupta: "Performance analysis of various machine learning-based approaches for detection and classification of lung cancer in humans." Neural Computing and Applications, 31.10: 6863-6877, 2019.
20. 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.
21. Xie, Ying, et al.: "Early lung cancer diagnostic biomarker discovery by machine learning methods." Translational oncology, 14.1: 100907, 2021.