MES COLLEGE OF ENGINEERING, KUTTIPPURAM

DEPARTMENT OF COMPUTER APPLICATIONS 20MCA245 – MINI PROJECT

PRO FORMA FOR THE APPROVAL OF THE THIRD SEMESTER MINI PROJECT

*(Note: All entries of the pro forma for approval should be filled up with appropriate and complete information. Incomplete Pro forma of approval in any respect will be rejected.)*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| |  | | --- | | Mini Project Proposal No :  *(Filled by the Department)* | | |  |  | | --- | --- | | Academic Year | : 2021-2022 | | Year of Admission | : 2020 | |

1. Title of the Project : Analysis of CT Scan Images to Predict Lung

Cancer Stages Using Image Processing Techniques

2.Name of the guide: Mr. Mohammad Jabir C

3. Number of the Student: MES20MCA-2037

4. Details (in BLOCK LETTERS)

Name: NASRIN BP Roll Number : 37

Signature

Date:15/12/2021

**Approval Status :** Approved / Not Approved

Signature of

Committee Members

# Comments of The Mini Project Guide Dated Signature

Initial Submission :

First Review :

Second Review :

# Comments of The Project Coordinator Dated Signature

Initial Submission:

First Review

Second Review

Final Comments :

Page - 1

Dated Signature of HOD

Analysis of CT Scan Images to Predict Lung Cancer Stages Using Image Processing Techniques

NASRIN BP

**Introduction:**

Lung cancer is one of the most dangerous and common cancer diseases in the world. Early detection of lung cancer can increase survival time of a patient. It is difficult for doctors to identify the cancer stages from Computed Tomography (CT) scan images. In this era of technology computer-aided system can help us to predict lung cancer stages more accurately. Inspired by the recent success of image processing and machine learning techniques in medical field we have developed models using Gray level co-occurrence matrix (GLCM) based texture image analysis and Statistical parametric approach for helping doctors to detect lung cancer stages. Our approach involves image acquisition, preprocessing, feature extraction and finally classification. For feature extraction purpose two approaches are used: Gray level co-occurrence matrix (GLCM) based texture image analysis and Statistical parametric approach. For detecting lung cancer stages four different classifiers are used and obtained the highest accuracy 78.95% with 0.77 precision and 0.83 recall using Support Vector Machine(SVM) in the Statistical parametric approach of feature selection.

Lung cancer also familiar as lung carcinoma caused by malignant lung tumor which have uncontrolled cell growth. This cell growth can spread to the other part of the body by the metastasis process. Mainly there are two types of lung cancer, one is small-cell lung cancer (SCLC) and the other one is nonsmall-cell lung cancer (NSCLC). The primary symptoms of lung cancer are coughing, losing weight, breath shortness and chest pain . One of the main reasons of lung cancer is smoking besides being a passive smoker, air pollution and genetic factors are also responsible for lung cancer. Avoiding smoking with other risking factors can prevent lung cancer primarily. Lung cancer stage can be divided into limited stage and excessive stage. In limited stage, cancer is confined in one lung, and in excessive stage cancer has spread to the other parts of the body .CT images have chosen as it is more efficient compared to X-ray for detecting lung cancer stages. The main reasons to choose digital image processing techniques (DIP) for detecting lung cancer stage is image gives better visualization and information compared to other forms. Image processing techniques are easier way to analyze image cells and extract data from them. According to our knowledge a few research work has been done on Lung Cancer stage detection. In this study, statistical parametric approach has least amount of features compared to GLCM approach. As the number of features increases, dimension increases on the ten to the power of number of features. Using statistical parametric approach computation is not getting complex compared to GLCM approach. Statistical parametric approach gives best result on our dataset. Support Vector Machine (SVM), K Nearest Neighbor(KNN), Random Forest and Naive Bayes classifiers used for learning purposes. SVM with highest accuracy 78.95% in Statistical parametric approach performs best for our selected dataset.

**Objectives:**

•To idententify the positive and negative comments in a comment section

• It can be applied in any section like feedbacks,reviews,post…etc

**Problem Definition:**

Disease prediction using patient treatment history and health data by applying data mining and machine learning techniques is ongoing struggle for the past decades. Many works have been applied data mining techniques to pathological data or medical profiles for prediction of specific diseases. These approaches tried to predict the reoccurrence of disease. Also, some approaches try to do prediction on control and progression of disease. The recent success of deep learning in disparate areas of machine learning has driven a shift towards machine learning models that can learn rich, hierarchical representations of raw data with little preprocessing and produce more accurate results. Numbers of papers have been published on several data mining techniques for diagnosis of heart disease such as Decision Tree, Naive Bayes, neural network, kernel density, automatically defined groups, bagging algorithm and support vector machine showing different levels of accuracies in diseases prediction. In this type of research generally used tool is Waikato Environment for Knowledge Analysis (WEKA).

**Basic functionalities:**

**Functional module**

**SVM Algorithm**

SVM is a supervised machine learning algorithm which can be used forclassification or regression problems. It uses a technique called the kernel trick to transform your data and then based on these transformations it finds an optimal boundary between the possible outputs.

**Module Discription**

Admin

Doctor

User

**ADMIN**

* Login
* View user
* Doctor management
* View feedback
* View complaint and send reply
* View comments
* View questions

**Doctor**

* login
* Upload patient result (san report,cancer stages etc)
* View questions & send reply
* View user
* View patient(lung cancer stage wise)
* View questions & send reply

**USERS**

* Registration
* login
* Update profile
* View result(san report,cancer stages etc)
* send doubts
* View answer
* Add comments
* view comments
* Send feedback
* Send complaint and view reply

**Tools / Platform, Hardware and Software Requirements:**

**Hardware Requirements:**

The selection of hardware is very important in the existence and proper working of any software. Then selection hardware, the size and capacity requirements are also important.

* Processor : Intel Pentium Core i3 and above, 64 bits

* RAM : Min3GB RAM

* HARD DISK: 10 GB

**Software Requirements:**

One of the most difficult task is selecting software for the system, once the system requirements is found out then we have to determine whether a particular software package fits for those system requirements. The application requirement:

* OPERATING SYSTEM: WINDOWS 10
* FRONT END: HTML, CSS, JAVASCRIPT
* BACK END: Mysql
* IDE USED: Jetbrains Pycharm, Android studio
* TECHNOLOGY USED: PYTHON JAVA
* FRAME WORK USED: Flask