**Lung Cancer Detection Using ML**

**Table of Contents**

* **Introduction**
* **Objective**
* **Background**
* **Hardware & Software Requirements**
* **Future Scope**
* **Conclusion**
* **Bibliography & References**

**Introduction-**

Lung cancer is one of the main cause of the death and health issue in many countries with a 5-year survival rate of only 10-16%. In this project we use machine learning algorithms to

diagnose a cancer and start treatment in early stages.

We use KNN & Decision Tree algorithms to predict the accuracy of the cancer. In this

project we use scikit-learn libraries like s-klearn and pandas to predict and classify the dataset

of the lung cancer patients. Slicing the dataset and feature scaling options are used to train the

dataset. After that we use confusion matrix, f1 score and accuracy score to predict the accuracy

of the result.

**Objectives-**

Lung cancer is considered as the deadliest cancer worldwide. For this reason, many countries

are developing strategies for the early diagnosis of lung cancer. In this project the objective is

to give best result accuracy of lung cancer patients. To achieve this objective, we use

Neighbors Classifiers & Decision Tree algorithm to classify the data set and give the best

accuracy of the result.

Background-

To perform KNN algorithm and Decision Tree algorithm we use scikit-learn library. In

background of this project, used libraries are given below-

* Numpy - NumPy is a python library used for working with arrays. It also has functions for working in the domain of linear algebra, fourier transform, and matrices.
* Pandas - Pandas is a high-level data manipulation tool developed by Wes McKinney. It is built on the Numpy package and its key data structure is called the DataFrame. DataFrames allow you to store and manipulate tabular data in rows of observations and columns of variables
* Sklearn - Scikit-learn is a free machine learning library for Python. It features various

algorithms like support vector machine, random forests, and k-neighbors, and it also supports Python numerical and scientific libraries like NumPy and SciPy.

* Neighbors Classifier - KNN is a non-parametric and lazy learning algorithm. Non-

parametric means there is no assumption for underlying data distribution. All training data used in the testing phase. This makes training faster and testing phase slower and costlier. Costly testing phase means time and memory.

* Decision Tree - DTs are a non-parametric supervised learning method used for classification and regression. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features.

| **Hardware Tools** | **Minimum Requirements** |
| --- | --- |
| Processor | i5 or above |
| Ram | 4GB |
| Monitor | 15.6’’ colored |
| Mouse | optical |
| Keyboard | 122 keys |
| Hard Disk | 50Gb |

| **Software Tools** | **Minimum Requirements** |
| --- | --- |
| **Platform** | **Windows** |
| **Operating System** | **Windows 8 or above** |
| **Technology** | **Machine Learning - Python** |
| **Scripting Language** | **Python** |
| **IDE** | **Pycharm** |

**Future Scope-**

The lung cancer detection system using the machine learning technique is much efficient and

gives the betterment result to the radiologist and assist them. This enhances with the additional

features for upgrading in the future. On this processing system to support the radiologist to

detect the affected patients as accurate as the result.

**Conclusion-**

We processed the dataset to differentiate the affected patient and its level of the growth of the

cancer by the machine learning system. Here it presented an approach to find best accuracy of

the cancer result to assist the radiologist and for the future enhancement. Further loads ought

to be directed at improving the classifying accuracy levels of result through experiments with

various alternatives

**Bibliography & References-**

Dataset - Lung Cancer Dataset ([**https://www.kaggle.com/datasets/yusufdede/lung-cancer-dataset?resource=download**](https://www.kaggle.com/datasets/yusufdede/lung-cancer-dataset?resource=download))

Machine Learning Journal

ML Algorithms (machinelearningmastery.com)

Scikit-Learn Library (scikit-learn.org)

Lung Cancer Detection Using Machine Learning Techniques by IJAREEIE

(www.ijareeie.com)