

“A Novel Framework for Lung Cancer Prediction using Machine Learning Algorithms”

Naveen Kumar¹, Bhavya B M²

PG Student, Dept. of MCA, P.E.S College of Engineering Mandya, Karnataka, India¹

Assistant professor, Dept. of MCA, P.E.S College of Engineering Mandya, Karnataka, India²

Abstract: Researchers have broadly used statistical and device getting to know strategies to assemble prediction fashions in numerous domains, consisting of software program fault prediction, junk mail detection, disorder diagnosis, and monetary fraud detection. Predicting sufferers who're susceptible to lung most cancers facilitates medical doctors make choices approximately treatment. In this regard, this observe seeks to evaluate the strength of numerous predictors within side the observe to growth the performance of detecting lung most cancers primarily based totally on symptoms. A range of classifiers together with Support Vector Machine (SVM), C4.5 Decision tree, Multi-Layer Perceptron, Neural Network, Naïve Bayes (NB) are to be had and we use one of the green set of rules called “KNN set of rules” and evaluated on a benchmark datasets received from UCI repository. The overall performance is likewise analyzed with famous ensembles confusion matrix. We construct automation for lung most cancers prediction the usage of Microsoft technology consisting of VISUAL STUDIO AND SQL SERVER.

Keyword: Machine Learning, Lung Cancer, KNN Algorithm, UCI Repository.

I. INTRODUCTION

One of the important fields of medical study is disease detection. The prediction of lung cancer is not automated. The current method is laborious, costly, dependent on medical equipment, unsatisfactory to users, inefficient, and inaccurate. The idea may be used at a hospital or clinic to analyses patient data related to lung cancer. It may be used as a platform for an online health community where people can learn more about lung cancer.

II. LITERATURE SURVEY

This paper proposes an efficient algorithm for the detection and prediction of lung cancer using a multiclass SVM (Support Vector Machine) classifier. Multi-level classification is used to detect cancer. Image enhancement and segmentation were performed separately at each stage of classification. Threshold and marker-driven basin-based segmentation was used for segmentation. For classification, a binary SVM classifier was utilized in [1]. Radiologists may identify, forecast, and diagnose lung cancer with the use of a clever computer-aided diagnostic system. An efficient multi-class SVM classifier-based system for lung cancer detection and prediction was created in this work. The classification of cancer was done in multiple stages. At each classification stage, image enhancement and segmentation have been done independently[2]. Using lung X-ray images, present an autonomous system for the early diagnosis of Plano cellular lung cancer. The approach is to try early diagnosis at the level of general practitioners using the least expensive diagnostic technology, chest radiography. Lung cancer is one of the most deadly malignancies and is typically identified too late. The suggested technique locates and separates the suspicious region in the lung image. It entails two steps: computing similarity coefficients to compare the retrieved Plano lung cancer structure to the x-ray picture. And then identifying the largest similarity coefficient to identify the likely cancer-affected region[3].

III. PROBLEM STATEMENT

Lung cancer prediction has become a worldwide health issue and could be a part of concern. Current medical system requires longer for carcinoma prediction because it requires longer for diagnosis, consultation, doctor's experience etc. The current technique could be manual, time-consuming, costly, and need extra training for doctors. Our study primarily focuses on applying machine learning algorithms to identify life-threatening illnesses like lung cancer.

IV. PROPOSED SYSTEM

System developed with the use of effective tools, including C# as the programming language, Visual Studio as the front end technology, and SQL Server as the back end technology. Compared to other technologies micro-soft technologies

are more compatible and more supportive with real time application. We use efficient algorithm for prediction, we use "KNN algorithm" for prediction. System collects data from [www.Kaggle.com](https://www.kaggle.com) and model is built to predict Lung Cancer disease.

V. METHODOLOGY

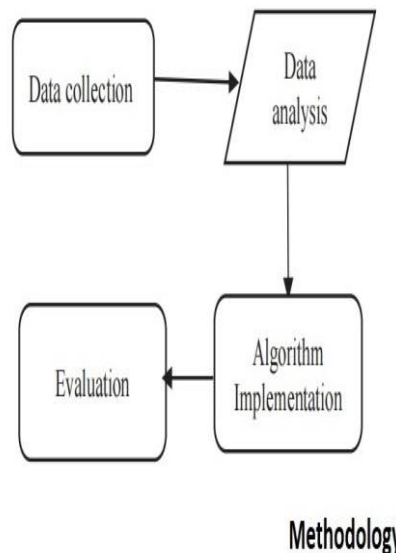
Machine Learning

The process of investigating a system using data is called machine learning. Data science includes machine learning that processes data using machine learning algorithms.

Supervised Learning Technique

It is a predictive model used in jobs that require predicting one value using the data set's other values. There will be predetermined labels for supervised learning. Based on the characteristics of one of the established sets of labels, it categorizes an item.

In supervised learning, there are several algorithms we may use to develop models, including KNN, Naive Bayes, Decision Trees, ID3, Random Forest, SVM, Regression techniques. Choose the best algorithm for your predictions based on your requirements, labels, parameters, and datasets. An algorithm is used to create a model that, in the face of uncertainty, generates predictions based on data.



Step 1: Raw data Collected

This is the first step in the prediction process where we collect medical data. Previous year's patient's data collected for processing. More than 5000 + data-sets used for processing. Training data-sets will contain patient details and also parameters that are required for prediction.

Step 2: Extract and Segment Data

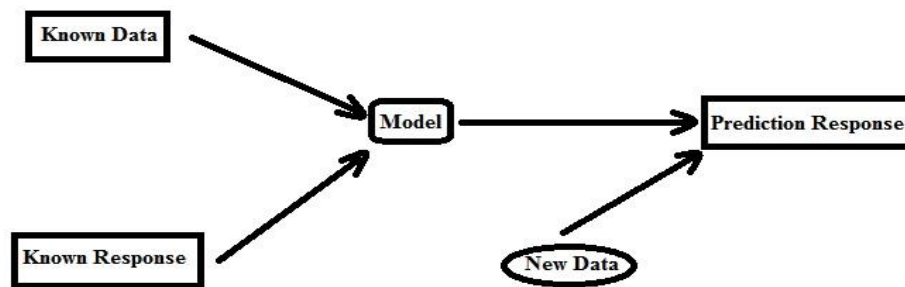
Here medical data analyzed and only relevant data extracted. The data required for processing extracted and segmented according to the requirement. Required data extraction is done because entire training data not required for processing and if we input all data, it requires too much of time for processing, so data processing is done.

Step 3: Train Data

Once required data extracted and segmented, we need to train the data, train means converting the data into the required format such as numerical values or binary or string etc... Conversion depends on the algorithm type.

Step 4: ML Technique for Disease Prediction

In this project, we employ the "Bayesian Classifier or KNN method" for prediction, which is effective and suitable for all possible parameter settings. Additionally, it produces reliable findings.



Proposed Model

Step 5: Lung Cancer Prediction

Doctor can access to the core module where system predicts the Lung Cancer and types for the new patients based on the inputted parameters. Here system uses “KNN” algorithm or “Naive Bayes” for disease prediction.

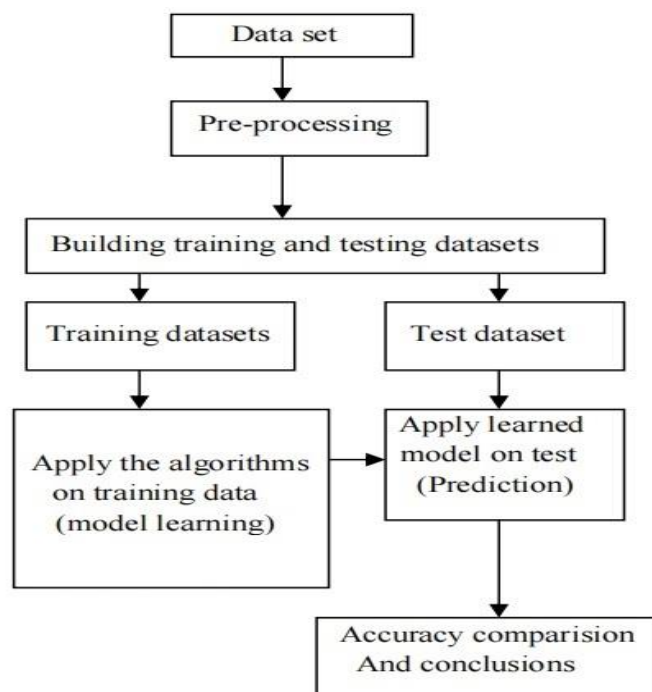
Step 6: Results

Results generated by the algorithm are checked with the accuracy using confusion matrix method. Here we validate the results generated by the algorithm “Bayesian classifier” and “KNN algorithm”.

Step 7: Visual Representation

Final outputs represented on GUI. When user’s gets login to the application system predicts the disease and displays on a GUI.

FLOW DIAGRAM



Flow chart for prediction

VI CONCLUSION

For hospitals and clinicians, developing a lung cancer prediction system is important. Early lung cancer detection made possible by the method enables doctors to treat patients more successfully. The suggested solution is a real-time application designed for many hospitals that can more quickly predict lung cancer. With the application of machine



learning algorithms, we can forecast lung cancer with more efficiency and accuracy. By using the classification methods, it is completed effectively. This method of categorization falls under data science technology.

VII REFERENCES

- [1] Moffy Vas , Amita Dessai “Lung cancer detection system using lung CT image processing”. International Conference on computing ,Communication, Control and Automation (ICCUBEA) 2017:8463851
- [2] Janee Alam ,Sabrina Alam , Alamgir Hossan “Multi-Stage Lung Cancer Detection and Prediction Using Multi-class SVM Classifier”. International Conference on computer ,communication ,chemical, material and Electronic Engineering (IC4ME2)2018:8465593
- [3] Vesna Zeljkovic , Milena Bojic ”Automatic detection of abnormalities in lung radiographs caused by Plano cellular lung cancer”. 1st Middle East Conference on Biomedical Engineering (MECBME)2011:5752067