IMPACT OF DIMENSIONALITY REDUCTION USING KNN AND SVM FOR THE CLASSIFICATION OF LUNG CANCER

Y.Vineela Sravya[1], P Mrudula[2], P V Vaidehi [3], V Anusha [4] , V Lakshmi Sahithi [5] and R Nandini [6]
[1,2,3,4,5,6] Department of Computer Science and Engineering, Vignan's Institute of Engineering for Women, Visakhapatnam, Andhra Pradesh, India.

ABSTRACT
Lung Cancer is a type of cancer that form in the tissues of the lung and most often occurs in people who smoke. The detection of lung cancer in early stage provides the best possibility of its cure .We use Lung Cancer dataset (age, smokes, alcohol etc.).The stage of Cancer is determined by the size of the tumor. Air pollution and chemicals at work are also the risk factors .And we follow the classifiers such as KNN (k-Nearest Neighbor), SVM(Support Vector Machine), and we use dimensionality Reduction concept on symptoms and CT scan images to provide low error rate and high accuracy.

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%. People who smoke have the greatest risk of lung cancer, though lung cancer can also occur in people who have never smoked. The risk of lung cancer increases with the length of time and number of cigarettes you've smoked. If you quit smoking even after smoking for many years, you can significantly reduce your chances of developing lung cancer. In this project we use machine learning algorithms to diagnose a cancer and start treatment in early stages. With the expected increase in the number of preventive/early detection measures, scientists are working in computerized solutions that help alleviate the work of doctors, improve diagnostics' precision by reducing the subjectivity factor, speedup the analysis and reduce medical costs. In order to detect malignant nodules, specific features need to be recognized and measured. Based on the detected features and their combination cancer probability can be assessed. However, this task is very difficult, even for an experienced medical doctor, since nodule presence and positive cancer diagnosis are not easily related. Early diagnosis is very important in lung cancer. Because lung cancer often has the ability to rapidly spread to the bones, liver, brain and adrenal glands. However, with the newly developed lung cancer treatment methods, the average life expectancy and quality have increased. Today, with advances in imaging techniques such as low-dose spiral computed tomography ,lung cancer can be detected at the early stage.

Lung cancer can start in the trachea, main airway or Lung cancer can start in the trachea, main airway or lung. It is caused by the uncontrolled growth and spread of certain cells in the lungs. People with lung diseases such as emphysema and previous chest problems are more likely to be diagnosed with lung cancer. Excessive use of tobacco, cigarettes and beedis is the major risk factor for lung cancer in Indian men; however, smoking is rare among Indian women, suggesting that other factors contribute to lung cancer. Other risk factors include radon exposure in the workplace, air pollution, and chemicals. Cancers that start in the lungs are primary lung cancers, and cancers that start in the lungs and spread to other parts of the body are secondary lung cancers. The size of the tumour and its extent determine the stage of the cancer. Early-stage cancers are small cancers diagnosed in lung cancer, and advanced-stage cancers are cancers that have spread to surrounding tissues or other parts of the body. A better understanding of risk factors can help prevent lung cancer disease. The key to improving survival is early detection using machine learning techniques, and if we can make the diagnostic process more efficient and effective for the radiologists who use it, that will be a critical step towards aim to improve early detection.

LITERATURE SURVEY

[1]Özge GÜNAYDIN, Melike GÜNAY, Öznur ŞENGEL,"Comparison of Lung Cancer Detection Algorithms ". numerous ways have been employed to find cancer in its early stages. In this study, machine literacy ways were examined for excrescence discovery in lung cancer. To find anomalies, we used machine literacy ways similar as star element Analysis, K Nearest Neighbors, Support Vector Machines, Naive Bayes, Decision Trees, and Artificial Neural Networks. Both with and without pre processing, we compared every system. The experimental findings indicate that Decision Tree provides the stylish result with,24 delicacy without image processing while Artificial Neural Networks provides the stylish result with,43 delicacy after image processing. In this study, we used colorful machine literacy ways to dissect casket radiographs and identify lung cancer. also, we used PCA to gauge back casket radiographs by a factor of1/8. Loss of features would affect from reducing dimension. In our script, there was not a significant information loss. also, it bettered SVM and KNN delicacy(where k = 2, and k = 3). delicacy results only affected nanosecond quantities, which can be disregarded to simplify calculation and conserve storehouse. Due to the large quantum of data, we were unfit to apply Naive Bayes and a 10- subcaste Feed Forward Neural Network to our prints. Despite the fact that neural networks are more accurate than other machine literacy ways. The top scores across all performance criteria in the original data belong to the decision tree.

[2]Swati Mukherjee, Prof.S.U. Bohra,"Lung Cancer Disease opinion Using Machine Learning Approach". From the morning to the present, the analysis and study of lung diseases has been the most fascinating exploration area for medical professionals .The performance of a neural network model is estimated in the following gospel to address the problem of relating nasty cells in image data, a typical problem in remedial imaging applications .A frame for relating lung cancer grounded on AI and deep neural networks has been established in an trouble to complete this thing. The methodology relies on supervised literacy, for which a advanced position of perfection has been attained, particularly when employing the deep literacy medium. A strategy for classifying lung excrescences is CNN bracket. The frame uses a number of ways, similar as image accession ,pre-processing, addition, segmentation, point birth, and neural frame identification. In a nutshell, a machine learning fashion can offer a preliminarily unheard- of implicit to enhance decision backing in the low- cost treatment birth, picture pre-processing, and deep literacy ways to produce prognostications that were more accurate. In the end, a significant neural system fashion is used to descry lung cancer with lesser delicacy and prognosticate its stages with lesser perfection. We could also show how using AI might potentially unnaturally separate and group low crowd. The use of further images, similar as X-rays, CT reviews, MRIs, and PET reviews, can ameliorate delicacy and enable croakers to give timely and affordable treatment. farther disquisition can be done to determine the knowledge gaps in illness control and opinion styles, which can prop the development of a vaccine or other control measures, given the complaint's negative profitable goods.

[3]Radhika P RRakhi.A.S.Nair,"A relative Study of Lung Cancer Discovery using Machine Learning Algorithms". Lung cancer is a condition in which lung cells gain unbridled. Although lung cancer can not be prevented, the threat can be dropped. thus, early identification of lung cancer is essential for perfecting patient survival. Lung cancer prevalence is directly equally identified with the number of chain smokers. colorful bracket ways, including Naive Bayes, SVM, Decision trees, and Logistic Retrogression, were used to dissect the lung cancer vaticination. The primary thing of this study is to probe the effectiveness of bracket algorithms in the early identification of lung cancer. In the history, a croakerwould need to do a number of tests to determine whether a case had lung cancer or not. still, this was a lengthy procedure. A case may sometimes be needed to suffer meaningless examinations or farther tests in order to diagnose lung cancer. These days, machine literacy algorithms are pivotal for the bracket and vaticination of medical data. The machine literacy styles employed for this comparison study are Nave Bayes, Logistic Regression, SVM, decision trees, and Logistic Retrogression. The delicacy rates of each classifier are compared and bandied. Quantitative comparisons are made between classifiers' vaticination capacities. For each classifier on the

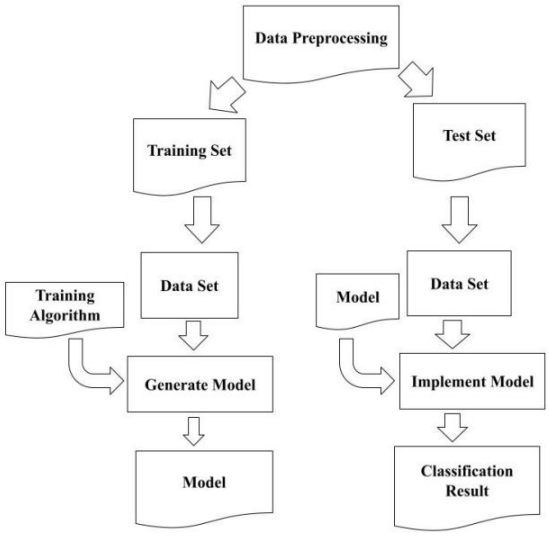
lung cancer dataset, colorful issues are shown in the performance map. The support vector machine fashion yields the stylish result when considering the correct bracket(CA) and other criteria. For the topmost performance, the SVM algorithm used a high dimension into classify the observation. This system can be used to diagnose lung cancer more precisely. There are hence smaller crimes. Last but not least, more pre-processing can ameliorate delicacy

PROPOSED SYSTEM

- The proposed work focuses on finding tumour, early symptoms of the diseases, appearing in patients‟ lungs. In this, available lung cancer images are passed through the system: pre-processing stage, feature Extraction stage and classification
- The dimensionality reduction is achieved by the feature selection algorithm. It also provide less memory utilization and latency for critical applications. Most of the existing methods have proven appreciable results when the classification is carried out after feature selection.
- The algorithms used are KNN, SVM .The comparison of the various classifiers based on the high detection accuracy and the reduced error yields the best classifier and the model is then trained and tested with the Lung cancer dataset.
- The performance of the system is evaluated using various metrics such as accuracy, true positive, false positive, Precision, Recall and F-measure.

Precision=TP/TP+FP
Recall=TP/TP+FN
F-measure=(2*Precision*Recall)/(Precision+Recall)
Accuracy=TP+TN/TP+FP+FN+TN

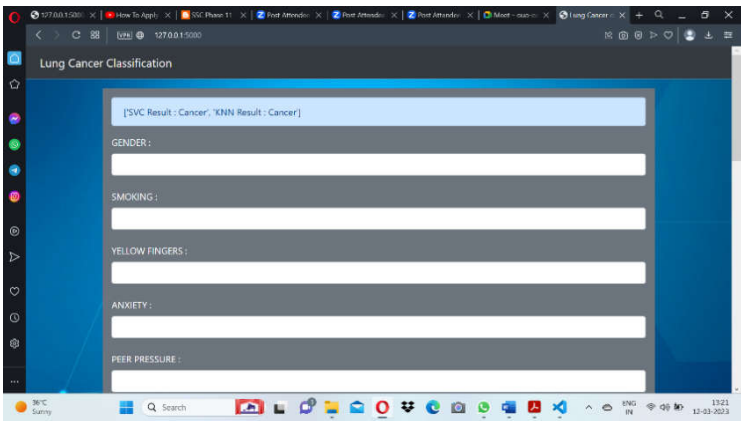
PROPOSED SYSTEM ARCHITECTURE



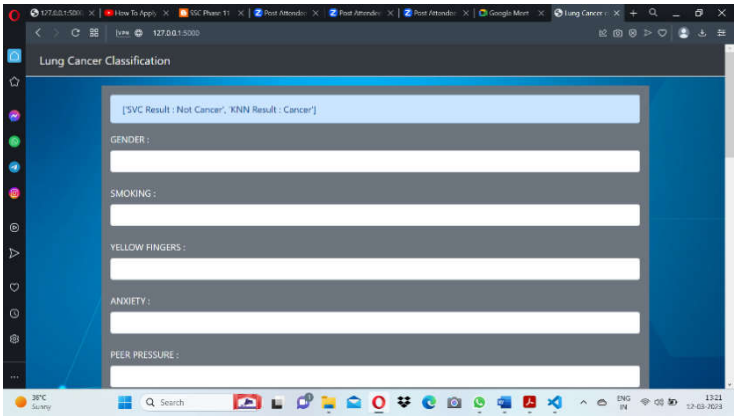
RESULTS AND DISCUSSIONS

DETECTING LUNG CANCER THROUGH SYMPTOMS

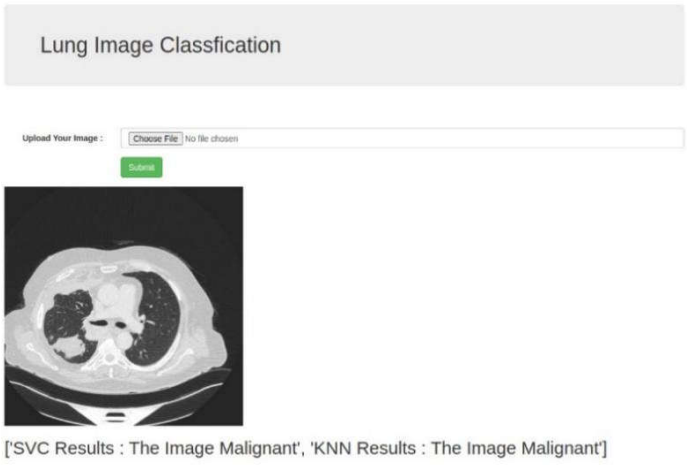
1.PATIENT OF HAVING CANCER



2. PATIENT HAVING NO CANCER



DETECTING LUNG CANCER THROUGH LUNG CT IMAGES



CONCLUSION

This project is entitled “Impact Of Dimensionality Reduction using KNN and SVM for the classification of Lung cancer.” is useful for the detection of lung cancer using symptoms and CT scan images. The project gives the results whether a person is having the lung cancer or not, so that the person can cure the cancer at an early stage. This project finally leads to the improvement of accuracy in detecting the lung cancer.

FUTURE SCOPE

As of now, the accuracy of svm is 86%,The feasibility of this project is increasing the accuracy of svm to 100%

REFERENCES

1. Ada, RajneetKaur, ” A Study of Detection of Lung Cancer Using Data Mining Classification Techniques ” International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 3, March 2013
2. Lung Cancer detection and Classification by using Machine Learning & Multinomial Bayesian-IOSR Journal of Electronics and Communication Engineering (IOSR-JECE) e-ISSN: 2278-2834,p-ISSN: 2278-8735.Volume 9, Issue 1, Ver. III (Jan. 2014), PP 69-75
3. KwetisheJoroDanjuma, ” Performance Evaluation of Machine Learning Algorithms in Post-operative Life Expectancy in the Lung Cancer Patients” Department of Computer Science, ModibboAdama University of Technology, Yola, Adamawa State, Nigeria
4. Sathyan H, Panicker,J.V,,”Lung Nodule Classification Using Deep ConvNets on CT Images”,9th International Conference on Computing, Communication and Networking Technologies, ICCCNT 2018
5. Anita chaudhary, SonitSukhraj Singh“Lung Cancer Detection on CT Images by Using Image Processing”2012 International Conference on Computing Sciences Nihad Mesanovic , Mislav Grgic, Haris Huseinagic,Matija Males, Emir Skejic, Muamer Smajlovic”Automatic CT Image Segmentation of the Lungs with Region Growing Algorithm”
6. Nikita Pandey, Sayani Nandy *“*A Novel Approach of Cancerous Cells Detection from Lungs CT Scan Images*’’*International Journal of Advanced Research in Computer Science and Software Engineering Volume 2, Issue 8,August 2012
7. Prof. Samir Kumar Bandyopadhyay “Edge Detection From Ct Images Of Lung““ International Journal Of Engineering Science & Advanced Technology Volume - 2,Issue - 1, 34 – 37
8. FatmTaher, NaoufelWerghi and Hussain Al-Ahmad “Extraction of Sputum Cells using Thresholding Techniques for Lung Cancer Detection” 2012 International Conference on Innovations in Information Technology
9. QinghuaJi,Ronggang Shi “A Noval Method of Image Segmentation Using Watershed Transformation”2011 International Conference on Computer Science and Network Technology