DETECTION OF CANCER IN LUNGS USING AN IMPROVED MACHINE LEARNING ALGORITHM

A.Vinitha[1], G.Anitha[2]*

ABSTRACT

In the current scenario lung cancer has become most threatful disease in our world. Due to use of smoking cigarette tends to be a major reason. This type of cancer could be found in the tissues of lungs. It could be found when there is a huge growth in the cells that are present in the lungs. Initial step of lung cancer can start from anywhere in the lungs later it begins to affect the entire part of the respiratory system. Hence, for this reason the cancer needs to be detected during the early stages and this could be done by using Machine learning algorithms. Here In this paper, for predicting lung cancer we have used KNN algorithm. For predicting the cancer, we have imported the dataset to get the accurate results whether the patient has been affected with cancer or not.

KEYWORDS: Lung cancer, KNN, Prediction, Distance Metrices.

I. INTRODUCTION

In today's world day to day death is increasing worldwide for this one of the main reasons is lung cancer. In most of the countries the health issue is of the range 10- 16% with the last 5 year of survival rate. The deadliest cancer in this world is considered to be as lung cancer. This could damage the cells in the respiratory system and also damages the whole tissues which provides problem to breathing and becomes difficult to intake hygienic air and as the final it leads to death. Here prediction of lung cancer has been analyzed by using the classification algorithm known to be KNN (K Nearest Neighbor). The major key objective of this paper is to detect the lung cancer at the early stage by using this algorithm. This also helps in predicting the stage of cancer. As the final this reveals the accuracy of 95 % in this proposed approach on lung cancer database.

ADVANTAGES OF THE SYSTEM

1. Early-stage diagnosis of cancer in individuals which could help doctors to provide better results for patients.
2. This could save millions of people life.

II. LITERATURE SURVEY

The lung cancer diagnosis should be done in a systematic process. Such as blood test, Imaging test and collection of samples. Blood tests should be taken to find out whether the tumor is present in the lung or not. Blood test shows the AFP(Alpha-Fetoprotein) level [1].

Next will be an MRI scan or ultrasound which is basically an imaging test. Through imaging tests, we can know the spread of cancer, location of cancer. So, the imaging test plays a predominant role. A third step sample will be collected such as a piece of tissue from the liver to know the malignancy level and to make the diagnosis process according to malignancy rate [2].

The goal of this paper is to get the useful information from a dataset. Likewise, to know the occurrence or spread of the malignancy in lung is a most needed one. To know through that image, it should be processed in step by step. The first thing is to elevation of image. Elevation or enhancement of image leads to a maximum error-less image [3].

Picture enhancement is the first stage of image processing. Improving the image quality for the human eye is the goal of the image enhancement procedure. Additionally,

[1,2]Department of Computer Applications,
Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India
* Corresponding Author

this procedure is necessary to provide greater feedback. Image for additional processing, ensuring that the final image contains less mistakes after going through all the various processing stages [4].

The two components of the picture enhancement technique are the spatial domain technique and the frequency domain technique. While the frequency domain approach deals with the rate of change of pixels that are changing as a result of the spatial technique changes the value of the pixel according to the demand. What kind of technique is domain, the spatial domain best for improving images cannot be decided? There are numerous methods [5].

III. METHODOLOGY

3.1 KNN Algorithm

This algorithm known is a supervised machine learning and also non- parametric. This works with the help of proximity in order to predict the individual data point from the group. Therefore, this is also used for neither classification problem or regression problem most widely this is used for classification problem because this type of algorithm works with the assumption of similar data points which is placed nearly.[6] In this problem a label is used as of majority vote basis which represents the data point around the used data. This is known to be plurality voting this word is derived from majority vote most widely used in literature.

3.2 DISTANCE METRICS

P=2 this is known to be Euclidean distance which is most widely used to measure distance in the real valued data point vectors. This helps to measure the value by a straight line distinct between two points which need to be measured. [7]

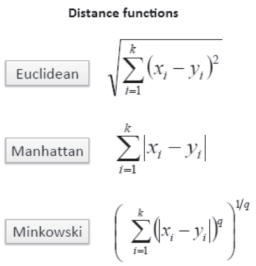

Figure 1: Distance metrices (Euclidean distance)

ADVANTAGES

1. Implementation of this algorithm is simple and easy.
2. This also adjusts the new training data set even though the data is stored to memory.

3.3 DATASET AND ATTRIBUTES

The dataset that has been used to predict the lung cancer includes the attributes such as age, Gender, Air pollution, Dust Allergy, Obesity, Wheezing, Alcohol Consumption etc. Thus, this attribute provides the results in the form of 2, 6 and 9 for lung cancer prediction.[8]

IV. RESULT

Therefore, the result provides the accuracy of correct and incorrect prediction that has been done by using classifier. This also measures the entire measure of the performance of this model. This keeps on evaluated for the better performance via a calculation of metrices such as accuracy, prediction.[9]

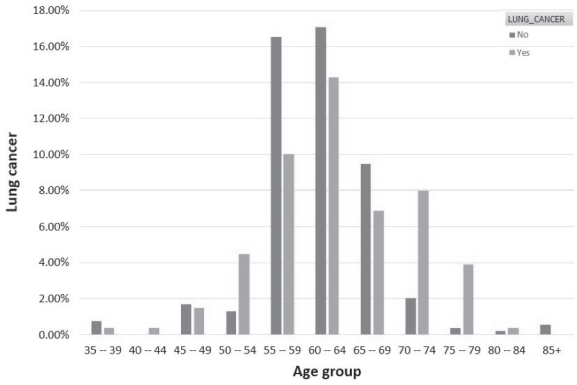

Figure 2: Prediction of Lung cancer among age group

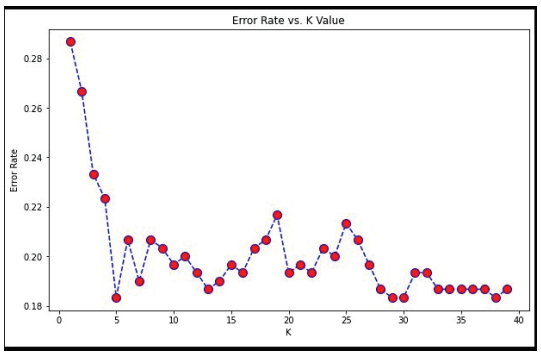

Figure 3: Error rate with K-Value data point

V. CONCLUSION

Machine learning algorithm that is used to predict the lung cancer in this paper is KNN even though there are many machine learning algorithms. The accurate prediction tends to help the doctors to provide the better treatment for the patients in the specified duration by using the stages of lung cancer. [10] In case of providing the incorrect prediction which may lead to harmful consequences for the individual. Hence, we need to use the better algorithm to provide the better result. Therefore, we have used KNN algorithm which provided the good accuracy that could able to predict the lung cancer.

REFERENCES

[1] A. Sharma and P. Kaur, “Optimized Lung cancer Detection and Segmentation Using Neural Network,” Systems Science & Control Engineering ,2022.

[2] R. Aarthi, S. Nivetha, and P. Vikashini, “lung cancer detection using image processing,” Pattern Analysis and Applications, 2020.

[3] A. Raj and M. Jayasree, “Automated Lung cancer Detection Using Markov Random Field Segmentation,” Journal of Ambient Intelligence and Humanized Computing , 2016.

[4] A. Krishan, “Detection and Classification of Lung Cancer using CT Images,” Emerging Science Journal , 2021.

[5] N. Dhanachandra and Y. J. Chanu, “Image Segmentation Method using K-means Clustering Algorithm for Color Image lung cancer detection,” European Journal of Cancer ,2021.

[6] G. Rajesh and A. Muthukumaravel, “Role of Artificial Neural Networks (ANN),” Computer Methods and Programs in Biomedicine, 2021.

[7] A.H. Ali and E. M. Hadi, “Diagnosis of Lung cancer from CT Images using Digital Image Processing,” IEEE Transactions on Medical Imaging , 2022.

[8] Y. Zhang, L. Wu and S. Wang, "Magnetic resonance brain image classification by an improved artificial bee colony algorithm", Journal of the American College of Radiology , 2020.

[9] S. A. Mohsin, N. M. Sheikh and U. Saeed, "MRI induced heating of deep brain stimulation leads: Effect of the air-tissue interface", Computer Methods and Programs in Biomedicine , 2018.

[10] N. P. Asimakis, I. S. Karanasiou, P. K. Gkonis and N. K. Uzunoglu, "Theoretical analysis of a passive acoustic brain monitoring system", IEEE Journal of Biomedical and Health Informatics, 2021.