

# Lung Cancer Prediction and Classification Using Recurrent Neural Network

V. Raaga Varsini<sup>1\*</sup>, M. Mohanasundari<sup>2</sup>

<sup>1</sup>Student, Department of Computer Science and Engineering, Velalar College of Engineering and Technology, Erode, India

<sup>2</sup>Assistant Professor, Department of Computer Science and Engineering, Velalar College of Engineering and Technology, Erode, India

**Abstract:** Lung cancer is the leading causes which affect both men and women in all over countries. Lung cancer which gives the low prognosis, and result in a high death rate. This world is computing sector and it is fully automated and now-a-days the medical industry is also fast moving into the digital world so they automating itself with the uses of image recognition and data analytics. This paper endeavors to inspect the prediction and the classification by the uses of three classifiers which is Support Vector Machine (SVM), K-Nearest Neighbor (KNN) and Recurrent Neural Network (RNN), this all will classify and give the accuracy value for the lung cancer and this accuracy value is more helpful to find whether the cancer is in early stage are critical stage. So that we can save many patients before the advanced stage. Basically, all the information of patients will be taken from UCI datasets to known how much of patients are affected by lung cancer. The principle of this paper is to the execution of the prediction, classification and accuracy value of the algorithms. The experimental results of prediction and classification shows that SVM gives the result with 78.56%, and KNN with 65.40%. and accuracy value result is from RNN with 92.75%.

**Keywords:** Lung cancer, Machine Learning, SVM, KNN, RNN.

## 1. Introduction

Cancer is a disease in which some of the body's cells of a human will grow uncontrollably and it will spread to all other parts of the human body. cancer will be come anywhere in human body. Cancer is made of trillions of cells. We were unnoticed the symptoms of cancer it will take the lives of human. Basically human cells will grow and it became old the old cells will be dead and new cells will be formed when the new cell is damaged and old cells were grown then it leads for the formation of tumors. lung cancer is one of the leading causes for cancer death in the world. Lung cancer is causing because of smoking. We can only prevent the lung cancer in the early stages, it will be more critical when we find it late lung cancer can be predicted and classified by using the machine learning techniques.

In this paper, they were clearly explained about the effective methods and prediction for the lung cancer treatments. Lung cancer is not so easy to cure this we should detect the disease early only we have chance for a patient to long lives. By using

deep learning techniques identification, classification and accuracy values have been done. The dataset is taken from the UCI dataset. The three classifiers are Support Vector Machine (SVM), K-Nearest Neighbor (KNN) and Recurrent Neural Network (RNN). These three algorithms gives the prediction, classifications accuracy value. This current paper is used for prediction and classification.

This paper is organized as follows, 2. Introduce to lung cancer, 3. Machine learning, 4. Related works, 5. Conclusion.

## 2. Lung Cancer Prediction

Carcinogenesis is used to multiply one or more cell types. Good tissues will not be support the growth of normal cells, and they will be separated quickly and become tumors. Primary lung cancer originates in the body and it will spreads to the lungs, while secondary lung cancer will be starts e in the body and it will spreads from there. This is the one of the most aggressive types of cancer and it will gives life-threatening to the human body. Whether the unchecked development is identified correctly at the early stages.it will help us for the diagnose and we can avoid unnecessary surgery and it will improve the chance of recovery. Chronic Obstructive Pulmonary (COPD) illness will attacks the areas of the lungs and it will cause many diseases like measles, influenza, pneumonia, and other respiratory issues and gives a breathing trouble. There are two types of lung cancer they are Small Cell Lung Cancer (SCLC) and Non-Small Cell Lung Cancer (NSCLC) this are the two main forms for lung cancer this will be develop and expand in their own ways. This non-small cell lung cancer has three subtypes they are (adenocarcinomas, squamous cell carcinomas, large cell carcinomas). The small/large cell cancer is a disease that occurs a patient and shows the symptoms for both types of cancer. (NSCLC) Adenocarcinoma is affect more common and it will be progressed more slowly than small cell lung cancer. Small cell lung cancer is linked to smoking which progresses more quickly and it will become a large tumor which spread all over the body.

\*Corresponding author: raagaradha1999@gmail.com

### 3. Related Works

[1] This paper proposed and a research in gene expression data to help with cancer classification in the field of oncogenomics. A small sample size is used for an adversely affect the accuracy of classification, as the performance of a classifier depends largely on the data. There is to propose an improved variant of K-nearest neighborhood (KNN) rules. To use Counting Quotient Filter, Euclidean distance and the best value from the k-neighbors for each target sample to a synthetic samples. A comparison is done among the raw data from public domain and that data are generated using the standard of K-nearest neighbor. The data generated through these approaches is then further classified using state-of-art classifiers like SVM, J48 and DNN.

[2] In this paper worked on very crucial and sensitive data for cancer classification. there are so many privacy considerations. Public domain datasets is a big help for researchers who can experiment with genes of individuals whose identity is undisclosed. The gene expression data at NCBI is a pre-processed, but the datasets are imbalanced; this needs to be handled for classifiers to work at their best. And have proposed a unique enhancement on the standard k-nearest neighborhood algorithm which helps to generate samples that have a greater likelihood of being behaviourally very close to the originals.

[3] In this paper they have said that the techniques of machine learning is used for the classifying breast lesions. They used for the mammogram accuracy in detecting malignant masses. One of the top causes of death for women remains breast cancer. In the early times diagnosis can facilitate adequate treatment to reduce the morbidity and mortality. mammogram will be an efficient method to detect breast lumps at an earlier stage. The main difficulty occurs during cancer detection and the distinction between a diagnosis to check whether a patient has a malignant and the type of disease. Machine learning algorithms such as K-Nearest Neighbors algorithm is help to solve the problem by providing high accuracy performance. K-Nearest Neighbors is one of the machine learning the diagnostic accuracy of the mammogram. This paper says about some recent studies that highlight algorithms used to enhance K-Nearest Neighbors algorithm, as a machine learning algorithm, in diagnosing cancer of breast.

[4] They are proposed the fusion scheme which is based on the Uniform-Local Binary Pattern (LBP) and filtered noise reduction. this new descriptor says that the LBP features is based on the new threshold. This paper proposes a multi-level fusion scheme for the auto-classification for the static of ultrasound images of breast cancer, in which we can attained two stages. In that several images will be generate from a single image using the pre-processing method. The median and Wiener filters will be utilized to lessen the speckle noise and the ultrasound image texture. This will allow the extraction reducing the overlap that benign and malignant image classes. the fusion mechanism will allow the production of diverse features from different filtered images. The feasibility of using the LBP-based texture feature to categorize the ultrasound images was demonstrated. As a result, the fusion process that can help achieve a powerful decision based on different features

produced from different filtered images improved the results of the new descriptor of LBP features in terms of accuracy, sensitivity, and specificity

[5] They were proposed that the disease have been existed as the mankind lung cancer has emerged as one of the most fata one time and again. it was one of the most common disease among all the cancers. There are more than 70,000 cases per year in India. This disease has more tendency and to be asymptomatic mostly in its earlier stages thus making it nearly impossible to detect. Only in early stages and we can detect and we can cure and recover the disease. In this technology plays a major role in detecting cancer. Many researchers have proposed different methods based on their studies. By using several computer-aided diagnosis (CAD) techniques to develop the detection. They will use the various Machine learning techniques and we can also deep learning techniques, there also have been several methods based off of image processing-based techniques to predict the malignancy level of cancer. In this paper, the aim will be focused onto list, discuss, compare and analyse several methods in image segmentation and in feature extractions we can use various techniques for prediction and classification. This disease is more dangerous and extremely tough to treat after it spreads upto an extent or reaching a serious stage. Computer-Aided Detection (CAD) is one of the constantly growing technologies that help detect cancer by feeding in certain inputs containing patient-related information such as scans like CT-Scan, X-Ray, MRI Scan. Unusual symptoms in patients or biomarkers, etc. SVM, CNN, ANN, Watershed Segmentation, Image enhancement, Image processing are a few methods used to improve the accuracy and aid the process. For training, the most popular datasets used are LUNA16, Super Bowl Dataset 2016, and LIDC-IDRI.

### 4. Conclusion

Lung cancer is the most dangerous diseases and which cause of death, the severity of the disease lies in the difficulty of diagnosing it in the early stages. This paper tries to endeavor to inspect the prediction and classification. We should find the best classifier to find the easy way for the detection so that we can give the treatment for lung cancer in early stage. The informational indices included in this study were derived from UCI databases for lung cancer patients. The focus of this paper is on using WEKA Tool to investigate the prediction and classification algorithms. This results show that the Support Vector Machine (SVM) give the prediction of 75.56%, and K-Nearest Neighbor (KNN)with 65.40%.

### References

- [1] S. F. Khorshid and A. M. Abdulazeez, "Breast Cancer Diagnosis Based On K-Nearest Neighbors: A Review," *Pal Arch's J. Archaeol. Egypt/Egyptology*, 2021.
- [2] P. Chaudhari, H. Agarwal, and V. Bhateja, "Data augmentation for cancer classification in oncogenomics: an improved KNN based approach," 2019.
- [3] H. R. Abdulqadir and A. M. Abdulazeez, "Reinforcement Learning and Modeling Techniques: A Review," 2021.
- [4] D. M. Abdullah and N. S. Ahmed, "A Review of most Recent Lung Cancer Detection Techniques using Machine Learning," *Int. J. Sci. Bus.*, 2021.

- [5] H. Tang, J. Zhao, and X. Yang, "Explore machine learning for analysis and prediction of lung cancer related risk factors," in Proceedings of the 2018 2nd International Conference on Computer Science and Artificial Intelligence, 2018.
- [6] P. R. Radhika, R. A. S. Nair, and G. Veena, "A Comparative Study of Lung Cancer Detection using Machine Learning Algorithms," in 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), 2019.
- [7] S. Uddin, A. Khan, M. E. Hossain, and M. A. Moni, "Comparing different supervised machine learning algorithms for disease prediction," *BMC Med. Inform. Decis. Mak.*, 2019.
- [8] A. Sreekumar, K. R. Nair, S. Sudheer, H. G. Nayar, and J. J. Nair, "Malignant Lung Nodule Detection using Deep Learning," in 2020 International Conference on Communication and Signal Processing (ICCSP), 2020.
- [9] N. Maleki, Y. Zeinali, and S. T. A. Niaki, "A k-NN method for lung cancer prognosis with the use of a genetic algorithm for feature selection," *Expert Syst. Appl.*, 2021.
- [10] D. Reddy, E. N. H. Kumar, D. Reddy, and P. Monika, "Integrated Machine Learning Model for Prediction of Lung Cancer Stages from Textual data using Ensemble Method," in 2019 1st International Conference on Advances in Information Technology (ICAIT), 2019.
- [11] A. Asuntha and A. Srinivasan, "Deep learning for lung Cancer detection and classification," *Multimed. Tools Appl.*, 2020.