



e-ISSN: 2319-8753 | p-ISSN: 2320-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Issue 7, July 2021

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.569

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com



Lung Cancer Detection using KNN

Priyanka Agawane, Manisha Khebale, Yogendra Patil

UG Student, Dept. of Computer, CCIS, JSPM's, Bhivarabai Sawant Institute of Technology & Research, Wagholi,
Pune, India

UG Student, Dept. of Computer, CCIS, JSPM's, Bhivarabai Sawant Institute of Technology & Research, Wagholi,
Pune, India

Assistant Professor, Dept. of Computer, CCIS, JSPM's, Bhivarabai Sawant Institute of Technology & Research,
Wagholi, Pune, India

ABSTRACT: Image inpainting is the process of restoring the lost or damaged regions or modifying the image contents imperceptibly. It refers to the process of filling-in missing data in a designated region of the visual input. In this paper, the technique presented is for detection and removal of text from images. The system detects text using morphological operations, connected component labelling and a set of selection criteria which helps to filter out non text regions. So, the resultant image is the image with only texts. Text Inpainting is done in two steps. The first step detects the text region automatically, without user interaction and in the second step; the text is removed from the image using exemplar based Inpainting algorithm.

KEYWORDS: Text detection, Inpainting, Morphological operations, Connected component labelling.

I. INTRODUCTION

Machine learning (ML) is already playing a vital role in healthcare with use into various fields like medical imaging and surgery to robots making the crucial and time taking tasks perform such actions. It is now widely used in detecting the diseases and discovering the new drugs. However, the question right here is posted what is the scope of ML in healthcare industry.

In this project I have use KNN & Decision Tree algorithms to predict the accuracy of the cancer. In this project we use scikit-learn libraries like sklearn 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.

II. THEORY

1) Small cell lung cancer

Small cell lung cancer occurs almost exclusively in heavy smokers and is less common than non-small cell lung cancer.

2) Non-small cell lung cancer.

Non-small cell lung cancer is an umbrella term for several types of lung cancers. Non-small cell lung cancers include squamous cell carcinoma, adenocarcinoma and large cell carcinoma.

Objectives:

- 1) 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 KNeighbors Classifiers & Decision Tree algorithm to classify the data set and give the best accuracy of the result.
- 2) Using machine learning approaches for classifying whether a person suffering from heart disease or not, using the given dataset.



- 3) In the dataset, the result attributes which have 0 value which represent person have no lung cancer & 1 value represents person have lung cancer

III. BACKGROUND

Cancer that starts in the lung is called primary lung cancer. There are several different types of lung cancer, and these are divided into two main groups:

A. Small cell lung cancer

B. Non-small cell lung cancer which has three subtypes:

- 1) Carcinoma
- 2) Adenocarcinoma
- 3) Squamous cell carcinomas.

The rank order of cancers for both males and females among Jordanians in 2008 indicated that there were 356 cases of lung cancer accounting for (7.7 %) of all newly diagnosed cancer cases in 2008. Lung cancer affected 297 (13.1 %) males and 59 (2.5%) females with a male to female ratio of 5:1 which Lung cancer ranked second among males and 10th among females. Lung cancer is the most dangerous and widespread cancer in the world according to stage of discovery of the cancer cells in the lungs, so the process early detection of the disease plays a very important and essential role to avoid serious advanced stages to reduce its percentage of distribution.

IV. METHODOLOGY

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 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-neighbours, and it also supports Python numerical and scientific libraries like NumPy and SciPy.

KNeighbors Classifier - KNN is a non-parametric and lazy learning algorithm. Nonparametric 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.



V. EXPERIMENTAL RESULTS

1	Name	Surname	Age	Smokes	AreaQ	Alkhol	Result
2	John	Wick	35	3	5	4	1
3	John	Constantin	27	20	2	5	1
4	Camela	Anderson	30	0	5	2	0
5	Alex	Telles	28	0	8	1	0
6	Diego	Maradona	68	4	5	6	1
7	Cristiano	Ronaldo	34	0	10	0	0
8	Mihail	Tal	58	15	10	0	0
9	Kathy	Bates	22	12	5	2	0
10	Nicole	Kidman	45	2	6	0	0
11	Ray	Milland	52	18	4	5	1
12	Fredric	March	33	4	8	0	0
13	Yul	Bryner	18	10	6	3	0
14	Joan	Crawford	25	2	5	1	0
15	Jane	Wyman	28	20	2	8	1
16	Anna	Magnani	34	25	4	8	1
17	Katharine	Hepburn	39	18	8	1	0
18	Katharine	Hepburn	42	22	3	5	1
19	Barbra	Streisand	19	12	8	0	0
20	Maggie	Smith	62	5	4	3	1

Fig 1 First 20 Instance Database

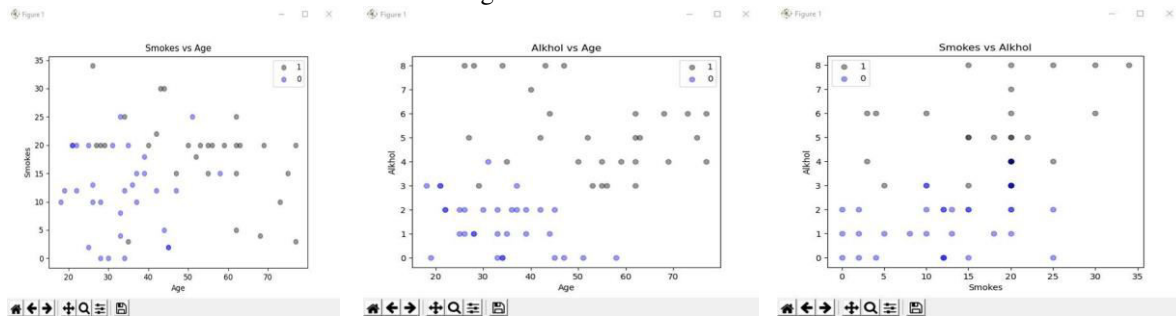


Fig 2 Graph Of Output Results

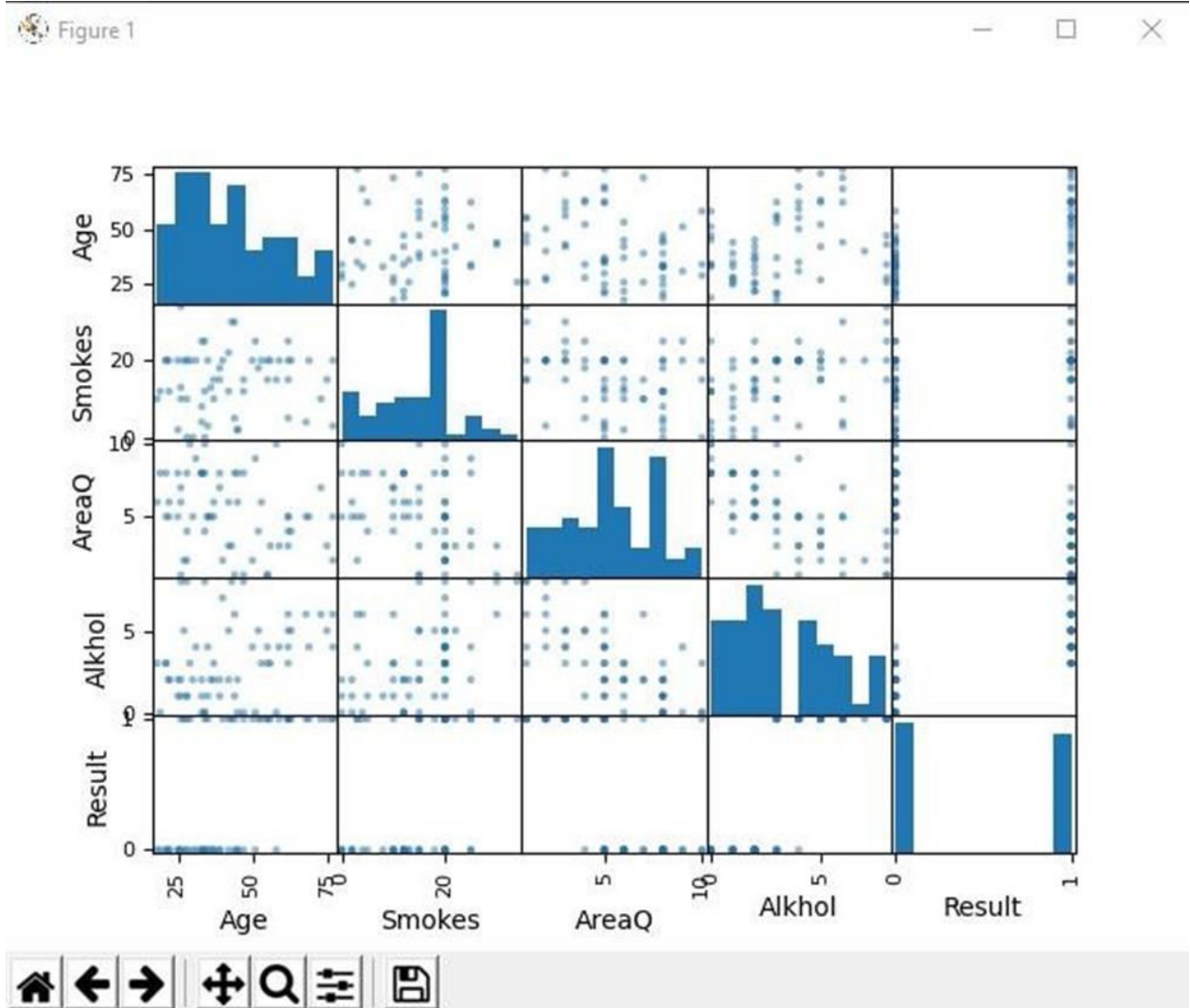


Fig 3 Output

VI. 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.

A benchmarking of the most performing architectures on available datasets using similar metrics can help in their comparative analysis. Finally, one of the current limitations is the data and their imbalanced nature. The use of new loss functions designed to tackle the problem of unbalanced classes such as focal loss, could improve the existing results, and help achieve more efficient training. With more datasets and more balanced data, I think that better results can be achieved.

VII. 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. Machine learning is the key to enabling Artificial Intelligence and the future of healthcare is data-driven. Big data and machine learning



have a tremendous potential in the healthcare field. All these technologies are not only improving treatment and diagnosis options, they also have the potential to take control of their own health by empowering individuals. With the help of advanced analytics, artificial intelligence and machine learning some of the most exciting advances are coming about in healthcare. Advances in AI interfaces, personalized medicine, predictive healthcare and advances in diagnostics all come down to the application of machine learning to help patients have access to smarter healthcare.

REFERENCES

1. <https://archive.ics.uci.edu/ml/dataset/Lung+cancer>. Accessed 12 Feb 2020
2. WHO Deaths by cause, sex and mortality stratum, World Health Organization. <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>. Accessed 25 Jan 2020
3. Ada, R.K.: Early detection and prediction of lung cancer survival using neural network classifier (2013)
4. Alcantud, J.C.R., Varela, G., Santos-Buitrago, B., Santos-Garcia, G., Jimenez, M.F.: Analysis of survival for lung cancer resections cases with fuzzy and soft set theory in surgical decision making. PLoS ONE **14**(6), e0218283 (2019)
5. Asuntha, A., Srinivasan, A.: Deep learning for lung cancer detection and classification. Multimedia Tools Appl. **79**, 1–32 (2020)
6. Bhatia, S., Sinha, Y., Goel, L.: Lung cancer detection: a deep learning approach. In: Bansal, J.C., Das, K.N., Nagar, A., Deep, K., Ojha, A.K. (eds.) Soft Computing for Problem Solving. AISC, vol. 817, pp. 699–705. Springer, Singapore (2019).
7. Bhuvanewari, P., Therese, A.B.: Detection of cancer in lung with k- m classification using genetic algorithm. Procedia Mater. Sci. **10**, 433–440 (2015)



**Impact Factor:
7.569**



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details