International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 5 Volume No.6 Aug-Sept – 2022
DOI:10.46647/ijetms.2022.v06i05.029 ISSN: 2581-4621

An High Accuracy Lung Cancer Detection System Using MRMR Method

Samarjeet kar[1], Dr S.Venkatesan[2], Dr vasantha swaminathan[3]
[1]PhD Scholar,Dept of ECE, GIET University Gunupur,Odisha-765022
[2]Professor, Dept of ECE, GIET University Gunupur,Odisha-765022
[3]Associate Professor, Dept of ECE, GIET University Gunupur,Odisha-765022

ABSTRACT
Lung cancer is one of the main cause of the death and health issue in many countries with 5-year survival rate of only10–16%. In this project we use machine learning algorithms to diagnose a cancer and start treatment in early stages.
MRMR method and decision tree algorithms to predict the high 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.Success obtained was 99.51% with 200 features provided by MRMR.
In the dataset lung cancer.csv the result attributes which have 0 value which represent person have no lung cancer and 1 value represents person have lung cancer. These researchers then performed the 10-fold cross-validation for model evaluation.IOT arduino UNO is interfaced with WIFI module to collect the data. wifi sensor is connected with various IOT sensor to obtain the information of the patient like temp, BP, Pulse rate..Using wifi module patient data is transfer to doctor to protect the patient in early stages.
Keywords –MRMR,decision tree,IOT,confusion matrix

Objective
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. Using machine learning approaches for classifying whether a person suffering from heart disease or not, using the given dataset.In the dataset, the result attributes which have 0 value which represent person have no lung cancer & 1 value represents person have lung cancer.
To achieve this objective, we use MRMR method and Decision Tree algorithm to classify the data set and give the best accuracy of the result.

1.Introduction

In this project we use machine learning algorithms to diagnose a cancer and start treatment in early stages. We use MRMR and 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.1000 efficient features were selected by the RMR method, obtained from the las layer of AlexNet without using the PCA method .Success obtained was 99.51% with 200 features provided by MRMR. These researchers then performed the 10-fold cross-validation for model evaluation.IOT Arduino UNO is interfaced with WIFI module ESP 8266 to collect the data. Wifi module is connected with temperature sensor DHT11,BP sensor,

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 5 Volume No.6 Aug-Sept – 2022
DOI:10.46647/ijetms.2022.v06i05.029 ISSN: 2581-4621

ECG sensor and tilt sensor to obtain the information of the patient like temp, BP, Pulse rate.Tilt sensor is involved to obtain shievering level of patient. Using wifi module patient data is transfer to doctor to protect the patient in early stages.
These values are stored in dataset using pyserial python libraray.pyserial library will connect IOT with python. Lung cancer.csv dataset is involved in our paper to store attributes of patients.Small cell lung cancer occurs almost exclusively in heavy smokers and is less common than 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. Decision Tree are a non-parametric supervised learning method used for classification and regression. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features.
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.

MRMR method comparision
model flow

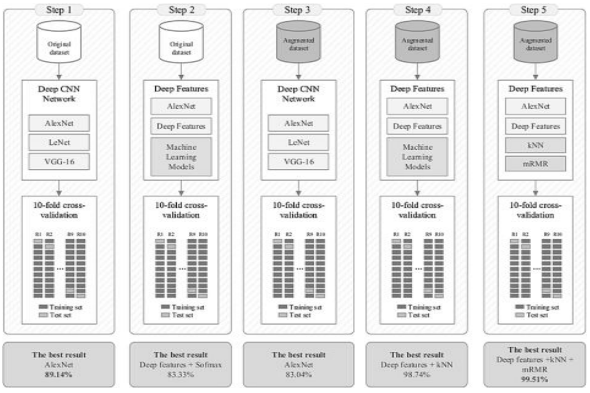


Fig 1

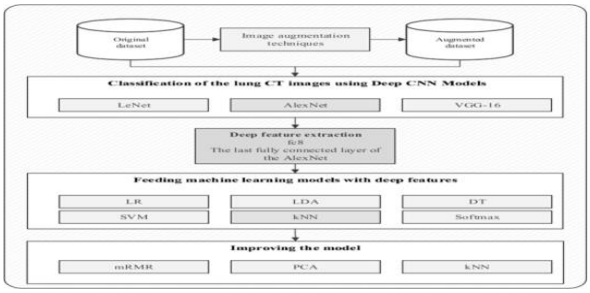


Fig 2

2.Methodology

2.1 Types of ml algorithm:

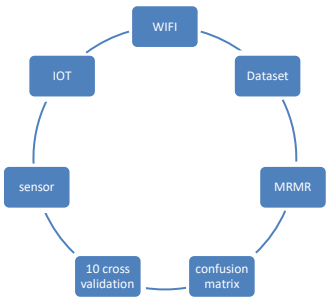


Fig 3

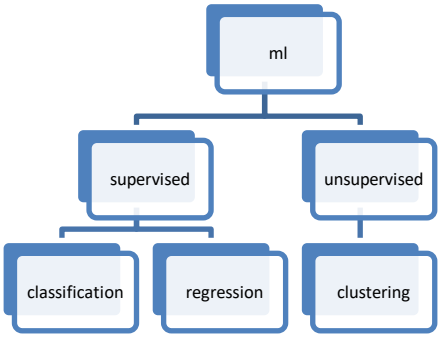


Fig 4

2.2.Pre processing
The working structure is simple, it consists of three major steps, they are
1. Data Pre-Processing

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 5 Volume No.6 Aug-Sept – 2022
DOI:10.46647/ijetms.2022.v06i05.029 ISSN: 2581-4621

2. Computation of data by MRMR method
3. Computation of confusion matrix and cancer prediction

The working structure of the project demanded learning concepts of MRMR and going through research papers to get a comprehensive understanding of the activities involved in processing, training and testing a neural network.

n=165	Predicted: NO	Predicted: YES	
Actual: NO	TN = 50	FP = 10	60
Actual: YES	FN = 5	TP = 100	105
	55	110	

Table 1 confusion matrix

Literature survey

Sno	title	technology	year	isbn
1.	Automatic lung cancer detection using artificial intelligence bardhrushiti	cnn	nov '19	UBT Knowledge Center
2.	Lung cancer detection using 3d convolution neural networks	3d cnn	jun' 20	Journal of Healthcare Engineering
3..	Lung Cancer Diagnosis Using Deep Attention Based Multiple Instance Learning and Radiomics Junhua	computer-aided diagno	JUN' 19	Computer Methods and Programs in Biomedicine

2.3 PROBLEM STATEMENT

The project deals with the identification of lung cancer, which affects roughly about 225,000 people every year and accounts to about 12 billion in health care costs in the final stages.

Therefore, early detection of cancer is critical to give the patients the best chance of recovery and survival.

The goal of this project is to evaluate the data (Slices of CT scans) provided with various pre-processing techniques and analyze the data using machine learning algorithms, in this case MRMR method to train and validate the model, to create an accurate model which can be used to determine whether a person has cancer or not. This will greatly help in the identification and elimination of cancer cells in the early stages.Therefore, an automated method capable of determining whether the patient will be diagnosed with lung cancer is the aim of this project.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 5 Volume No.6 Aug-Sept – 2022
DOI:10.46647/ijetms.2022.v06i05.029 ISSN: 2581-4621

2.4.PROPOSED METHODOLOGY

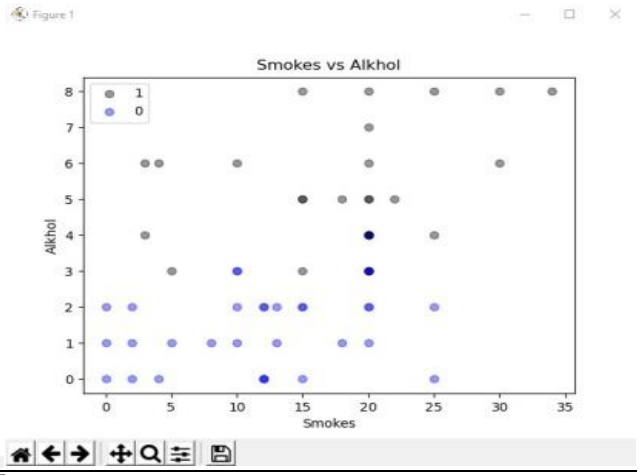
The lung cancer detection system using the machine learning technique is much efficient and gives the betterment result to the radiologist and assist them. This enhances with the additional features for upgrading in the future. On this processing system to support the radiologist to detect the affected patients as accurate as the result.

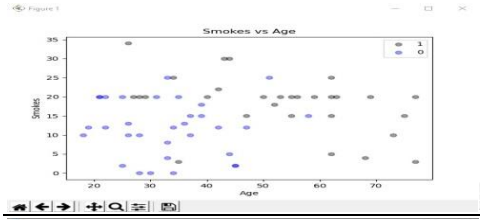
Machine learning is the key to enabling Artificial Intelligence and the future of healthcare is data-driven. Big data and machine learning have a tremendous potential in the healthcarefield. All these technologies are not only improving treatment and diagnosis options, they also have the potential to take control of their own health by empowering individuals.

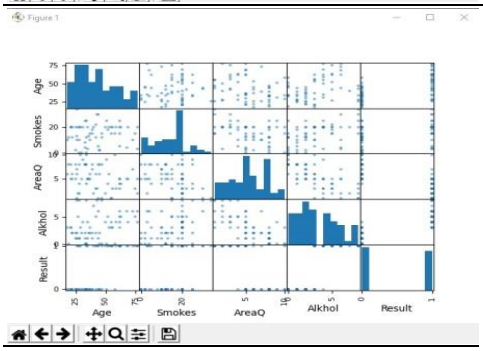
With the help of advanced analytics, **artificial intelligence** and machine learning some of the most exciting advances are coming about in healthcare. Advances in AI interfaces, personalized medicine, predictive healthcare and advances in diagnostics all come down to the application of **machine learning** to help patients have access to smarter healthcare.

The idea that I have presented is at the very beginning of its deployment phase and thus I wish to extend my project by implementing a model that can make valuable predictions based on the reports specified in the datasets and achieve a higher value of accuracy by using Microsoft's ResNet approach to build a network, to provide much higher learning rate of the nodes increasing the performance of the model and minimizing the losses effectively.The dataset contains 61 instances and 5 attributes which are attributes InformationAge,Smokes,AreaQ,Alcohol and Result(0,1).IDE used is Pycharm based on Python 3 interpreter.

3.Results and Discussion:



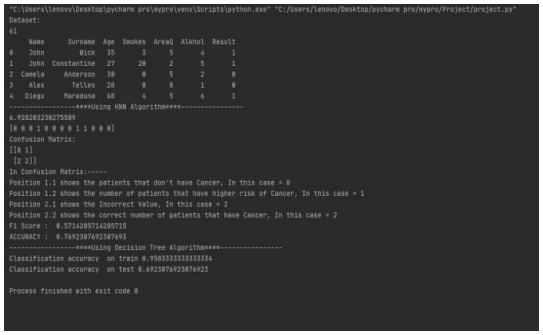




International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 5 Volume No.6 Aug-Sept – 2022
DOI:10.46647/ijetms.2022.v06i05.029 ISSN: 2581-4621

Dataset :

	Name	Surname	Age	Smokes	AreaQ	Alkhol	Result
2	John	Wick	35	3	5	4	1
3	John	Constantii	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	Brynner	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



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.

References

1.Automatic lung cancer detection using artificial intelligence mentor: dr. sc. bertankarahoda ,november 2019

2. Lung Cancer Diagnosis Using Deep Attention Based Multiple Instance Learning and RadiomicsJunhua Chen a,1, MS; HaiyanZeng 1, MD; Chong Zhang 1, MS; Zhenwei Shi 1, PhD Submitted to Computer Methods and Programs in Biomedicine

3.Cancer(2018)[WWWdocument].URL https://www.who.int/newsroom/factsheets/detail/cancer. [access on 20 June 2020]

4. Siegel, R. L., Miller, K. D., &Jemal, A. (2020). Cancer statistics, 2020.*CA: a cancer journal for clinicians, 70(1), 7–30.*

5.3D CNN with Visual Insights for Early Detection of Lung CancerUsing Gradient-Weighted Class ActivationEali Stephen Neal Joshua ,1Debnath Bhattacharyya ,2 MidhunChakkravarthy Hindawi,Journal of Healthcare ID 6695518, 11 pages

6. H. Xie, D. Yang, N. Sun, Z. Chen, and Y. Zhang, "Automated pulmonary nodule detection in CT images using deep convolutionalneural networks," *Pattern Recognition, vol. 85,* pp. 109–119, 2019.

7. W. W. Labaki, T. Gu, S. Murray et al., "Voxel-wise longitudinalparametric response mapping analysis of chest computed tomography in smokers," *Academic Radiology,*vol. 26, no. 2, pp. 217–223, 2019.

8. D. Ardila, A. P. Kiraly, S. Bharadwaj et al., "End-to-end lungcancer screening with three-dimensional deep learning onlow-dose chest computed tomography," *Nature Medicine,*vol. 25, no. 6, pp. 954–961, 2019.

9. M. Kroenke, K. Hirata, A. Gafita et al., "Voxel-based comparison and texture analysis of 18 F-FDG and 18 F-FMISO PET of patients with head-and-neck cancer," *PLoS One,*vol. 14, no. 2, 2019.

DOI:10.46647/ijetms.2022.v06i05.029 ISSN: 2581-4621

10. F. Liao, M. Liang, Z. Li, X. Hu, and S. Song, "Evaluate the malignancy of pulmonary nodules using the 3-D deep leaky noisy-OR network," *IEEE Transactions on Neural Networks and Learning Systems, vol. 30, no. 11, pp. 3484–3495, 2019.*