Lung Cancer Detection Using Machine Learning

Shivam Singh Vats
School of Computer Science Engineering
Galgotias University

Dr. Balamurugan Km
School of Computer Science Engineering
Galgotias University

Ayush Chauhan
School of Computer Science Engineering
Galgotias University

Abstract:- **Cellular breakdown in the lungs is one of the darkest deadly disease in the world . Earlier treatment recognition identification are hardly more crucial for patient. Medical expert use some biological picture just like histological image of Adencarcinoma of biopsy tissue from possible lungs images for analysis. Now onwards than not the analyzation of cellular breakdown In the lungs liver are hardly And Deadly .CNN means convolutional neural organization could be distinct different characterize in lungs type with more than prominent pression. Cellular breakdown in a limited time which is really hardly and crucial for deciding patient good treatment method and their success rate very lesser in harmless tissue in biological term Adenocarcinoma, and squamous. It will automatically detect without any human supervision. Decision Tree, Random Forest, Support Vector machine, Logistic Regression Machine Learning, a analysis process to predict the probability of a discrete outcome for any input variable will be used for classifying whether a person is suffering from heart disease or not, for the given used dataset by predicting the output in the form of Yes or NO.**

Keywords— CNN, Machine learning ,Artificial Intelligence, Segment, PET-CT

I. INTRODUCTION

There is a critical change the malignant growth types are not readable and understand .some coordinates which are part of this In cellular breakdown in the lungs is deadly disease among most of the people. Particular 28% of all death of lungs. They was not the patients of lung cancer. Cellular breakdown can be caused due to exposure to container or smoke and various element or harmful element present in the atmosphere. this is ideally in this research different or many kinds of image , X-ray. In lung cancer detection many types of test are perform to detect presence of harmful elements. Now onward we are performing biological images just as an input we give . In phase of biopsy when lab technician study these reports is a right way or simply way to analysis. After recognizing it forward image, X-ray ,CT-PET to next process and once access the result of cellular breakdown in the lungs. For pathologists and other clinical experts diagnosing cellular breakdown in the lungs and the sorts is a tedious interaction process, these don't works direct to incorrect fully treatment because for this special work probably it may take money to patients (men, women). We can find the accuracy of given image which is discussed already.

In this paper, we are using administered learning methods for foreseeing the beginning phase of lungs illness. Troupe calculations and a few calculations, for example, a Logistic Regressiin, k-nearest neighbor (KNN), support vector machine (SVM), Decision tree (DT), Random forest(RF) algorithms are utilized to group whether individuals tried have a place with the class of heart disease or sound individuals.

II. EASE OF USE

To develop an AI based model using various Machine learning algorithms and techniques which has the highest accuracy. Here we only need to train the models by some input data of individuals having heart disease and normal hearts both and validate it with some data. Later check the predicti

on of the model by comparing to actual data whether prediction is correct or not. This is very easy process inwhich we only need to compare the predictive accuracy of different models and choose the model with highest accuracy.

III. LITERATURE REVIEW

The author S. Shashikala, B. R. Sowmya proposed a modal using CNN on CT scan images to check weather the cancer is present in the lungs or not. Using they work and find two different phases in training volumetric features from input data as the first phase and classification as the second phase. The system proposed by them could classify the cancerous and non-cancerous cells with accuracy of 98%.

T. Atsushi, T. Tetsuya, K. Yuka, and F. Hiroshi applied Convolutional Neural Network on cytological images to lung cancer type classification. They considered little cell carcinoma, Squamous cell , Adenocarcinoma images in maintained dataset. The DCNN architecture of 3 convolution and pooling layers

Ritu-et-al introduced a feature determine technique for recognizing mortality occasions in patients withlung illness during treatment to track down the most basic elements Various AI techniques are used, including LDA, KNN, RF, SVM, DT, and GBC. The test discoveries showed that the consecutive component determination method accomplishes a precision of 86.67% for the irregular Random forest classifier.

IV. PROPOSED WORKING

This paper presents cellular breakdown in the lungs location dependent on CT scan of chest utilizing CNN. In the initial step , lung locales are been removed from CT picture and in that Each cuts are segmented to ready to get. Higher growth of segmented higher growth vertex area is used to prepare CNN (Convolutional Neural Network) and all these things. Then at the point of convolutional neural network is taken to test to modify of patient pictures. The principle of this theory of this research paper is to test there is high amount of cancer cells in a patient's lungs is dangerous or crucial. The summary of this proposed work firstly works on image, X-Ray segmented feature extraction, lungs cancer cell identification diagnosis result. In the cellular breakdown in the lungs.

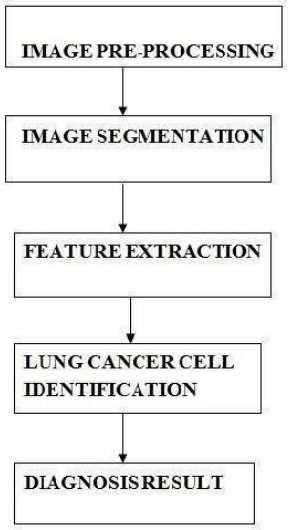

Figure 1 General model Proposed system

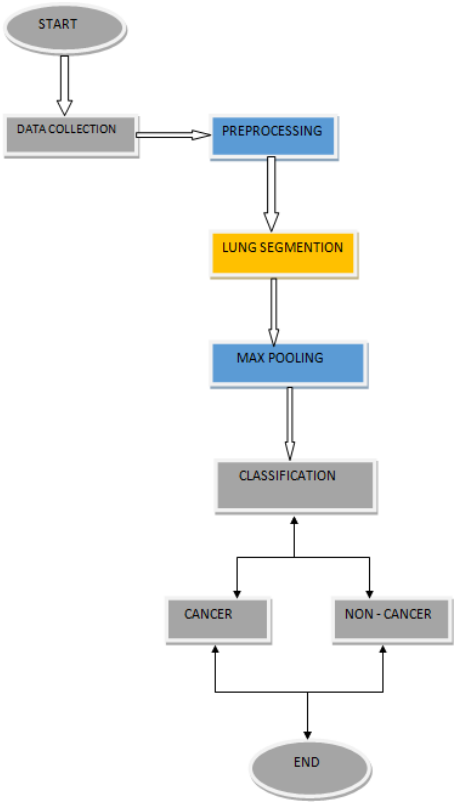

Figure 2 Flowchart of proposed mode

DATA PRE-PROCESSING

Preprocessing the dataset is essential for the viable portrayal of the of data quality. The dataset has been preprocessed utilizing strategies, for example, removing missing qualities from highlights, StandardScaler (SS), and MinMaxScaler. Missng value hndling is an information preprocessing method used to make a smooth dataset. In this way, it was begun by deciding whether the dataset contains any missing qualities.

The lung illness dataset utilized in this article comprises of six missing qualities. The missing qualities are supplanted with the mean or mode upsides of the elements.

Figure3. Data Preprocessing

FEATURE EXTRACTION

The feature extraction stage is one of the important while we train our model machine. Feature selection methods are expected to decrease the quantity of input dimensions to those that are accepted to be generally helpful to a model to predict the objective variable.

As there is big amount of data input, Some prescient displaying issues have an enormous number of factors that can slow the working and efficiency and preparing of models and require a lot of framework memory. Furthermore, the presentation of certain models can debase when including input factors that are not applicable to the objective variable.

In this stage, tests were led with and without feature selection to evaluate the impact of feature selection. The component choice plans to r ecognize the main elements of Infections.

V. IMPLEMENTATION

STAGE 1: Start

Convolution activity is the most important structure component in our assault plan in initial step detect segment images . Include locators, filter based with CNN which effectively fill in as the filter neurological organization's channels being a first step, will be addressed in this phase.

- Step 1(b): DATA COLLECTION

The Rectified Linear Unit, or ReLU, will be the second step in this development. We'll look at ReLU layers and how they affect Convolutional Neural Networks' linearity powers.

- Step 1(c): input Image
- Step 2(a):Detect X-Ray

STAGE 2: IMAGE SEGMENTATION

This will be a celluler breakdown of the straightening system and in another stage how we move from pooled to smoothed layers filter bades with CNN when working with Convolutional Neural Networks.

In this part it is used to bitwise addition to convert thebinary images

Ya= CNN(xa)

Where , xa = input of CNN network Ya = output of CNN network

Further to the next LSTM networks and RNN

STAGE 3: CNN

For gaining knowledge from 2D model of any input scanned input image we mostly used CNN , it’s structure used to perform such acts. The model was built from 22 layer deep CNN. More than 8 inception module as a group are stored in Google net. Every module has it’s convolutions at various measures, some pooling and integration operations.

Stage 3.2: MAX POOLING LAYER

In this layer taking the maximum value from a small window layer map which gives down sampling of thefeature maps

$F = \max\{F(a,b)\}$

Where, F(a,b) = Set of feature maps produced by conventionallayer maps

Several machine learning algorithms are used by the machine. Models are designed using those algorithms

Algorithms Used:-

- Logistic Regression
- K-Nearest Neighbor (KNN)
- Decision Tree (DT)
- Random Forest
- Support Vector Machine

Stage 4: Output Layer

The final result is obtained in the form of Yes or No, wether a person is having lung cancer or not.

VI. RESULT AND DISCUSSION

Efficiency of all the machine learning algorithms is calculated. Different models based on those efficiency are designed to predict the cardiac disease in a person.

Different machine learning models tried:

1. Logisic Regression
2. K-NN Classifier
3. Decision Tree Classifier
4. Random Forest Classifier
5. Support Vector Machine

K-NN ALGORITHM TRAIN RESULT

```
================================================
Accuracy Score: 86.79%

________________________________________________
Classification Report:  Precision Score: 87.18%
                        Recall Score: 88.70%
                        F1 score: 87.93%

________________________________________________
Confusion Matrix:
 [[ 82  15]
 [ 13 102]]

Test Result:
================================================
Accuracy Score: 86.81%

________________________________________________
Classification Report:  Precision Score: 88.00%
                        Recall Score: 88.00%
                        F1 score: 88.00%

________________________________________________
Confusion Matrix:
 [[35  6]
 [ 6 44]]
```

DECISION TREE TRAIN RESULT

```
================================================
Accuracy Score: 95.75%

________________________________________________
Classification Report:  Precision Score: 93.44%
                        Recall Score: 99.13%
                        F1 score: 96.20%

________________________________________________
Confusion Matrix:
 [[ 89   8]
 [  1 114]]

Test Result:
================================================
Accuracy Score: 80.22%

________________________________________________
Classification Report:  Precision Score: 84.78%
                        Recall Score: 78.00%
                        F1 score: 81.25%

________________________________________________
Confusion Matrix:
 [[34  7]
 [11 39]]
```

LOGISTIC REGRESSION TRAIN RESULT

```
================================================
Accuracy Score: 86.79%

________________________________________________
Classification Report:  Precision Score: 87.18%
                        Recall Score: 88.70%
                        F1 score: 87.93%

________________________________________________
Confusion Matrix:
 [[ 82  15]
 [ 13 102]]

Test Result:
================================================
Accuracy Score: 86.81%

________________________________________________
Classification Report:  Precision Score: 88.00%
                        Recall Score: 88.00%
                        F1 score: 88.00%

________________________________________________
Confusion Matrix:
 [[35  6]
 [ 6 44]]
```

SUPPORT VECTOR MACHINE TRAIN RESULT

```
================================================
Accuracy Score: 93.40%

________________________________________________
Classification Report:  Precision Score: 93.16%
                        Recall Score: 94.78%
                        F1 score: 93.97%

________________________________________________
Confusion Matrix:
 [[ 89   8]
 [  6 109]]

Test Result:
================================================
Accuracy Score: 87.91%

________________________________________________
Classification Report:  Precision Score: 89.80%
                        Recall Score: 88.00%
                        F1 score: 88.89%

________________________________________________
Confusion Matrix:
 [[36  5]
 [ 6 44]]
```

RANDOM FOREST CLASSIFIER TRAIN RESULT

```
Train Result:
================================================
Accuracy Score: 94.81%

________________________________________________
Classification Report:  Precision Score: 94.07%
                        Recall Score: 96.52%
                        F1 score: 95.28%
________________________________________________
Confusion Matrix:
 [[ 90   7]
 [  4 111]]

Test Result:
================================================
Accuracy Score: 84.62%

________________________________________________
Classification Report:  Precision Score: 86.00%
                        Recall Score: 86.00%
                        F1 score: 86.00%
________________________________________________
Confusion Matrix:
 [[34  7]
 [ 7 43]]
```

Accuracy table of Models

The prediction accuracy of different models is given:

Model	Training Ac	Testing Accu
K-nearest ne	86.79%	86.81%
Decision Tre	95.75%	80.22%
Random For	94.81	84.62%
Logistic Reg	86.79%	86.81
Support Vect	93.4	87.91

Table1. Accuracy table of models

VII. CONCLUSION

The proposed model using CNN convolutional neural network was performed to detect the pressence of cancer causing celles using CT-PET ,X-RAY various kinds of cancer causing tissueof different dimention in the organ have helped for performing the projet that will be implementd will have the effeciancy to detect weather cancer causing celles are there or not with accuracy of approximate 95%.the fact of this research paper the accuracy of finding lungs cancer with the help of this CNN was technical based on which we are performing in cellulerbreakdown in the lungs.

IX. REFERENCES

1. http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx.

2. 1iQO/Ind ia-recordedabout39-million-cancer-cases-in2016data.html.

3. Using Deep Learning for classification of Lung Tumorsof Computed Tomography Images.

4. M.S. Al-Tarawneh, “Lung cancer detection using image processing techniques,” Leonardo Electronic
Journal of Practices and Technologies, vol. 20, pp. 147– 58,May 2012.

5. K.Malarvizhi, R.Kiruba, “A Novel Method OfSupervision And Control Of First Order Level Process Using Internet Of Things “,Journal Of Advanced Research In Dynamical And Control Systems, Vol 9 No.6 2017, Pp.1876-1894.

6. S Sasikala, M Ezhilarasi, Combination of MammographicTexture Feature Descriptors for Improved Breast Cancer Diagnosis. Asian Journal ofInformation Technology, 2016.

7. A.Kavitha, Anusiyasaral and P.Senthil,” Design Model of Retiming Multiplier For FIR Filter &its Verification”, International Journal of Pure and Applied Mathematics, Vol116

No12, 2017, pp. 239-247 11. S Sasikala, M Ezhilarasi, Combination of Mammographic Texture8.LUNA16, "Lung tumor analysis 2016." https://luna16.grandchallenge.org

8. Jin, Taolin, et al. "Learning Deep Spatial Lung Features by 3D Convolutional Neural Network for Early Cancer Detection." (DICTA), 2017 International Conference on. IEEE, 2017.

9. da Nóbrega, Raul Victor Medeiros, et al. "Lung Nodule Classification via Deep Transfer Learning in CT Lung Images."31st International Symposium on. IEEE, 2018.

10. Fan, Lei, et al. "Lung nodule detection based on 3D convolutional neural networks." (FADS), 2017 International Conference.
Tang, Hao, Daniel R. Kim, and Xiaohui Xie. "Automated pulmonary nodule detection using 3D deep convolutional neural networks."15th International Symposium on IEEE, 2018.