Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 8, August : 2024

Detection of Lung Cancer Using a Conventional Neural Network

1MRS.CH.DEEPTI, 2 Mounika Vadlamudi

1Assistant Professor, Department of Master of Computer Applications,

QIS College of Engineering & Technology, Ongole, Andhra Pradesh, India

2PG Scholar, Department of Master of Computer Applications,

QIS College of Engineering & Technology, Ongole, Andhra Pradesh, India

Abstract: Lung most cancers is the main purpose for cancer-related death. Lung most cancers can provoke in the windpipe, major airway or lungs. It is brought about by way of unchecked increase and unfold of some cells from the lungs. People with lung disorder such as emphysema and preceding chest troubles have greater danger to be identified with lung cancer. Over utilization of tobacco, cigarettes and beedis, are the fundamental chance thing that leads to lung most cancers in Indian men; however, amongst Indian women, smoking is now not so common, which point out that there are different elements which lead to lung cancer. Other danger elements consist of publicity to radon gas, air-pollutions and chemical compounds in the workplace.

Lung most cancers detection at early stage has come to be very vital and additionally very handy with photograph processing and deep getting to know techniques. In this find out about lung affected person Computer Tomography (CT) scan pix are used to discover and classify the lung nodules and to notice the malignancy stage of that nodules. In this challenge we are the usage of CNN algorithm to become aware of Lung most cancers from CT-SCAN pics and to educate CNN we have CT-SCAN photographs dataset

1.INTRODUCTION

It is most common in smokers accounting 85% of cases among all. So many Computer Aided Diagnosis (CAD) Systems are developed in recent years. Detection of lung cancer at early stage is necessary to prevent deaths and to increase survival rate. Lung nodules are the small masses of tissues which can be cancerous or noncancerous also called as malignant or benign. Benign tissues are most commonly non-cancerous and does not have much growth where malignant tissues grows very fast and can affect to

the other body parts and are dangerous to health.

1.1 SCOPE:

For medical imaging so many different types of images are used but computer Tomography (CT) scans are generally preferred because of less noise. Deep learning is proven to be the best method for medical imaging, feature extraction and classification of objects. Several types of deep learning architectures are introduced by so many researchers to classify the lung cancer.

2.LITERATURE SUREVY

2.1 An Automatic Detection System of Lung Nodule Based on Multi-Group Patch-Based Deep Learning Network

High-efficiency lung nodule detection dramatically contributes to the risk assessment of lung cancer. It is a significant and challenging task to quickly locate the exact positions of lung nodules. Extensive work has been done by researchers around this domain for approximately two decades. However, previous computer-aided detection (CADe) schemes are mostly intricate and time- consuming since they may require more image processing modules, such as the computed tomography image transformation, the lung nodule segmentation, and the feature extraction, to construct a whole CADe system. It is difficult for these schemes to process and analyze enormous data when the medical images continue to increase. Besides, some state of the art deep learning schemes may be strict in the standard of database. This study proposes an effective lung nodule detection scheme based on multigroup patches cut out from the lung images, which are enhanced by the Frangi filter. Through combining two groups of images, a four-channel convolution neural networks model is designed to learn the knowledge of radiologists for detecting nodules of four levels. This CADe scheme can acquire the sensitivity of 80.06% with
4.7 false positives per scan and the sensitivity of 94% with 15.1 false positives per scan. The results demonstrate that the multigroup patch-based learning system is efficient to improve the performance of lung nodule detection and greatly reduce the false positives under a huge amount of image data.

2.2 Deep residual learning for image recognition

Deeper neural networks are more difficult to train. We present a residual learning framework to ease the training of networks that are substantially deeper than those used previously. We explicitly reformulate the layers as learning residual functions with reference to the layer inputs, instead of learning unreferenced functions. We provide comprehensive empirical evidence showing that these residual networks are easier to optimize, and can gain accuracy from considerably increased depth. On the ImageNet dataset we evaluate residual nets with a depth of up to 152 layers - 8× deeper than VGG nets [40] but still having lower complexity. An ensemble of these residual nets achieves 3.57% error on the ImageNet test set. This result won the 1st place on the ILSVRC 2015 classification task.

We also present analysis on CIFAR- 10 with 100 and 1000 layers. The depth of representations is of central importance for many visual recognition tasks. Solely due to our extremely deep representations, we obtain a 28% relative improvement on the COCO object detection dataset. Deep residual nets are foundations of our submissions to ILSVRC & COCO 2015 competitions1, where we also won the 1st places on the tasks of ImageNet detection, ImageNet localization, COCO detection, and COCO segmentation.

2.3 Accurate Pulmonary Nodule Detection in computed Tomography Images Using Deep Convolutional Neural Networks

Early detection of pulmonary cancer is the most promising way to enhance a patient's chance for survival. Accurate pulmonary nodule detection in computed tomography (CT) images is a crucial step in diagnosing pulmonary cancer. In this paper, inspired by the successful use of deep convolutional neural networks (DCNNs) in natural image recognition, we propose a novel pulmonary nodule detection approach based on DCNNs. We first introduce a deconvolutional structure to Faster Region-based Convolutional Neural Network (Faster R-CNN) for candidate detection on axial slices. Then, a three-dimensional DCNN is presented for the subsequent false positive reduction. Experimental results of the LUng Nodule Analysis 2016 (LUNA16) Challenge demonstrate the superior detection performance of the proposed approach on nodule detection (average FROC-score of 0.893, ranking the 1st place over all submitted results), which outperforms the best result on the leaderboard of the LUNA16 Challenge (average FROC-score of 0.864).

3.Praposed Work

Lung cancer identification at an early stage has become extremely crucial, as well as quite simple, thanks to image processing and deep learning techniques. Lung patient Computer Tomography (CT) scan images are used in this study to locate and classify lung nodules, as well as to determine the malignancy level of those nodules. In this research, we are utilising the CNN algorithm to detect lung cancer from CT-SCAN images, and we have a dataset of CT-SCAN images to train the CNN. The primary goal of this research is to investigate the performance of a classification algorithm in order to make an early diagnosis of lung cancer.

Fig 1:System Architecture

4. IMPLIMENTATION

- Upload Lung Cancer Dataset: In this module use upload dataset.
- Preprocess Dataset: In this module data undergoes preprocessing.
- Model Generation: In this module model generation is take place to predict disease.
- Build CNN Model: In this module CNN model is build.
- Accuracy & Loss Graph: In this module comparison graph is shown.
- Upload Test Image & Predict Cancer: In this module, user uploads test image to predict diseases.

5. ALGORITHM:

CNN :

```
def executeCNN():
    global cnn_acc
    X = np.load('features/X.txt.npy')
    Y = np.load('features/Y.txt.npy')
    Y = to_categorical(Y)
    classifier = Sequential()
    classifier.add(Convolution2D(32, 3, 3, input_shape = (64, 64, 3), activation = 'rel
    classifier.add(MaxPooling2D(pool_size = (2, 2)))
    classifier.add(Convolution2D(32, 3, 3, activation = 'relu'))
    classifier.add(MaxPooling2D(pool_size = (2, 2)))
    classifier.add(Flatten())
    classifier.add(Dense(output_dim = 256, activation = 'relu'))
    classifier.add(Dense(output_dim = 2, activation = 'softmax'))
    print(classifier.summary())
    classifier.compile(optimizer = 'adam', loss = 'categorical_crossentropy', metrics =
['accuracy'])
    hist = classifier.fit(X, Y, batch_size=16, epochs=12, shuffle=True, verbose=2)
    hist = hist.history
    acc = hist['accuracy']
    cnn_acc = acc[9] * 100
    text.insert(END,"CNN Accuracy : "+str(cnn_acc)+"\n")
```

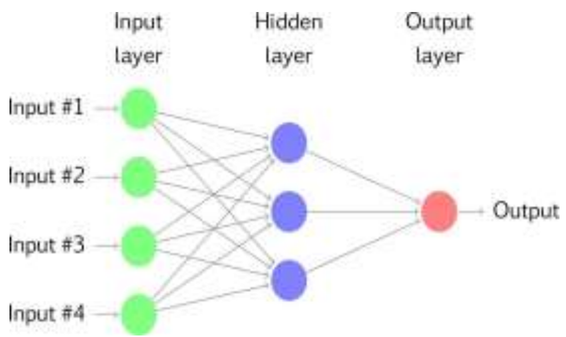
To demonstrate how to build a convolutional neural network based image classifier, we shall build a 6 layer neural network that will identify and separate one image from other. This network that we shall build is a very small network that we can run on a CPU as well. Traditional neural networks that are very good at doing image classification have many more parameters and take a lot of time if trained on normal CPU. However, our objective is to show how to build a real- world convolutional neural network using

TENSORFLOW.

Neural Networks ar essentially mathematical models

to solve an optimization problem. They are made of neurons, the basic computation unit of neural networks. A neuron takes an input (say x), do some computation on it (say: multiply it with a variable w and adds another variable b) to produce a value (say; z= wx+b). This value is passed to a non- linear function called activation function (f) to produce the final output(activation) of a neuron. There are many kinds of activation functions. One of the popular activation function is Sigmoid. The neuron which uses sigmoid function as an activation function will be called sigmoid neuron. Depending on the activation functions, neurons are named and there are many kinds of them like RELU, TanH.

If you stack neurons in a single line, it's called a layer; which is the next building block of neural networks. See below image with layers



To predict image class multiple layers operate on each other to get best match layer and this process continues till no more improvement left.

Deep learning not only accelerates the critical task but also improves the precision of the computer and the performance of CT image detection and classification.

In this paper, the problem of classification of benign and malignant is considered. It is proposed to employ, respectively, the convolution neural network (CNN)and deep neural network (DNN). The input data (image data) has a strong robustness on the distortion. The multiscale convolution image feature is generated by setting the convolution kernel size and parameter; the information of different angles is generated in the feature space.

3.1.1 SVM

```
def executeSVM():
    global classifier
    global svm_acc
    text.delete('1.0', END)
    cls = svm.SVC()
    cls.fit(X_train, y_train)
    predict = cls.predict(X_test)
    svm_acc = accuracy_score(y_test,predict) * 100
    classifier = cls
    text.insert(END,"SVM Accuracy : "+str(svm_acc)+"\n")
```

Introduction to SVM

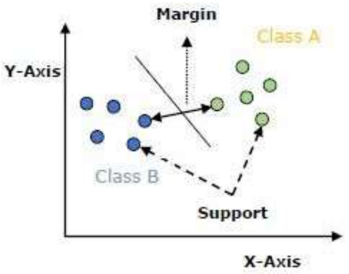
Support vector machines (SVMs) are powerful yet flexible supervised machine learning algorithms which are used both for classification and regression. But generally, they are used in classification problems. In 1960s, SVMs were first introduced but later they got refined in 1990. SVMs have their unique way of implementation as compared to other machine learning algorithms. Lately, they are extremely popular

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 8, August : 2024

because of their ability to handle multiple continuous and categorical variables.

Working of SVM

An SVM model is basically a representation of different classes in a hyperplane in multidimensional space. The hyperplane will be generated in an iterative manner by SVM so that the error can be minimized. The goal of SVM is to divide the datasets into classes to find a maximum marginal hyperplane (MMH).



The followings are important concepts in SVM

- **Support Vectors** − Datapoints that are closest to the hyperplane is called support vectors. Separating line will be defined with the help of these data points.
- **Hyperplane** − As we can see in the above diagram, it is a decision plane or space which is divided between a set of objects having different classes.
- **Margin** − It may be defined as the gap between two lines on the closet data points of different classes. It can be calculated as the perpendicular distance from the line to the support vectors. Large margin is considered as a good margin and small margin is considered as a bad margin.
- The main goal of SVM is to divide the datasets into classes to find a maximum marginal hyperplane (MMH) and it can be done in the following two steps.
- First, SVM will generate hyperplanes iteratively that segregates the classes in best way.
- Then, it will choose the hyperplane that separates the classes corectly.
- Implementing SVM in Python
- For implementing SVM in Python − We will start with the standard libraries import

SVM Kernels

In practice, SVM algorithm is implemented with kernel that transforms an input data space into the required form. SVM uses a technique called the kernel trick in which kernel takes a low dimensional input space and transforms it into a higher dimensional space. In simple words, kernel converts non-separable problems into separable problems by adding more dimensions to it. It makes SVM more powerful, flexible and accurate. The following are some of the types of kernels used by SVM.

Linear Kernel

It can be used as a dot product between any two observations. The formula of linear kernel is as below

K(x,xi)=sum(x∗xi)K(x,xi)=sum(x∗xi)

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 8, August : 2024

From the above formula, we can see that the product between two vectors say & x is the sum of the multiplication of each pair of input values.

Polynomial Kernel

It is more generalized form of linear kernel and distinguish curved or nonlinear input space. Following is the formula for polynomial kernel.

4.RESULTS AND DISCUSSION

Accuracy: The accuracy of a test is its ability to differentiate the patient and

k(X,Xi)=1+sum(X∗Xi)^dk(X,Xi)=1+sum(X∗Xi)^d

Here d is the degree of polynomial, which we need to specify manually in the learning algorithm.

Radial Basis Function (RBF) Kernel

RBF kernel, mostly used in SVM classification, maps input space in indefinite dimensional space. Following formula explains it mathematically

K(x,xi)=exp(−gamma∗sum(x−xi^2))K(x,xi) =exp(−gamma∗sum(x−xi^2))

Here, *gamma* ranges from 0 to 1. We need to manually specify it in the learning algorithm. A good default value of *gamma* is 0.1.

As we implemented SVM for linearly separable data, we can implement it in Python for the data that is not linearly separable. It can be done by using kernels.

Example

The following is an example for creating an SVM classifier by using kernels. We will be using *iris* dataset from *scikit- learn* healthy cases correctly. To estimate the accuracy of a test, we should calculate the proportion of true positive and true negative in all evaluated cases. Mathematically, this can be stated as:

Accuracy = TP + TN TP + TN + FP + FN.

$$\text{Accuracy} = \quad \text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

F1-Score: F1 score is a machine learning evaluation metric that measures a model's accuracy. It combines the precision and recall scores of a model. The accuracy metric computes how many times a model made a correct prediction across the entire dataset.

Precision: Precision evaluates the fraction of correctly classified instances or samples among the ones classified as positives. Thus, the formula to calculate the precision is given by:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\text{Precision} = \frac{\textit{True Positive}}{\textit{True Positive} + \textit{False Positive}}$$

Recall: Recall is a metric in machine learning that measures the ability of a model to identify all relevant instances of a particular class. It is the ratio of correctly predicted positive observations to the total actual positives, providing insights into a model's completeness in capturing instances of a given class.

$$Recall = \frac{TP}{TP + FN}$$

COMPARISION GRAPH

FINAL OUTPUT

FINAL OUTPUT

5.CONCLUSION

In earlier times, the doctor has to do multiple tests in order to detect whether a given patient has lung cancer or not .But this was a very time consuming process. In a diagnosis sometimes a patient has to undergo unnecessary check-ups or different tests to identify the disease of lung cancer. To minimize the process time and unnecessary check-ups there needs to be a preliminary test in which both the patient and the doctor will be notified with the possibilities of lung cancer. Nowadays the machine learning algorithms plays an important role in the prediction and classification of medical data. we can see predicted result as CT-SCAN contains abnormality and in second image we are detecting places were abnormality detected and in third image we extracted all abnormality patches from original image and then displaying.

The application utilizes advanced diagnostic algorithms and machine learning techniques to analyze CT scan data of the lungs. This results in improved accuracy and reliability in the diagnosis of various lung conditions such as lung cancer, pneumonia, and pulmonary embolism. Healthcare professionals can rely on the application's findings to make more informed decisions about patient care and treatment planning.

CT scans can be affected by noise, artifacts, and other imperfections that may hinder accurate diagnosis. The application is designed to handle these challenges and provide robust performance even in noisy environments. By reducing the impact of noise and artifacts, healthcare professionals can obtain clearer and more reliable results, leading to better patient outcomes.

The application automates the analysis of CT scan data, eliminating the need for manual interpretation and reducing the time required for diagnosis. This efficiency allows medical professionals to focus more on patient care and spend less time on the labor-intensive task of reviewing and analyzing CT scans. Quick and accurate diagnosis leads to prompt initiation of treatment, potentially improving patient prognosis.

By providing advanced analysis capabilities, the application enables researchers to extract meaningful insights from large volumes of CT scan data. This can aid in identifying patterns, trends, and correlations, contributing to a deeper understanding of lung diseases and potentially leading to the development of new diagnostic or treatment approaches.

FUTURE SCOPE

As a future work, the experiments could be performed by using Deep CNN architecture for other types of cancer.

REFERENCES

1. DSharma GJindal ‘‘Computer aided diagnosis system for detection of lung cancer in CT scan images International Journal of Computer and Electrical Engineering 3(5): 714-718 2011.

2. Janarthanan R Balamurali R Annapoorani A and Vimala V 2020 Prediction of rainfall using fuzzy logic Materials Today: Proceedings

3. DBhatnagar AKTiwari VVijayarajan AKrishnamoorthy ‘‘Classification of normal and abnormal images of lung cancer IOP Conference Series: Materials Science and Engineering Vol 263 2017

AUTHOR PROFILES

M.philph.d.mlse, miste currently working as an head of department and incharge of pg programs in the department of mca,qis college of engineering and technology, andhra pradesh,ongole

Mrs. Chepuri. Deepti, currently working as an Assistant Professor in the Department of Computer Science and Engineering , QIS College of Engineering and Technology, Ongole, Andhra Pradesh. She did her BTech from Uttar Pradesh Technical University, Lucknow, M.Tech from JNTUK, Kakinada. Her area of interest are Machine Learning, Artificial intelligence, Cloud Computing and Programming Languages.

Ms. Mounika Vadlamudi, currently pursuing Master of Computer Applications at QIS College of Engineering and Technology (Autonomous), Ongole Andhra Pradesh. she completed BSC from Loyola Degree College Vijayawada, krishna, Andhra Pradesh. Her areas of interests are Machine Learng.