International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 – 5161

Available at http://inpressco.com/category/ijcet

Research Article

Lung Cancer Disease Diagnosis using Two-Step Learning Approach

Ms.Nikita Jaywant Waragde and Dr.Prasad Halgaonkar

Department of Computer Engineering Zeal College of engineering and research, pune

Received 10 Nov 2020, Accepted 10 Dec 2020, Available online 01 Feb 2021, **Special Issue-8 (Feb 2021)**

Abstract

The examination in lung infection is the most intriguing investigation zone of expert's in early days. The proposed system is planned to distinguish lung threat in less than ideal stage in two stages. The proposed structure includes various methods, for instance, picture extraction, pre-preparing, paired picture change, thresholding, Division, feature extraction, and neural framework identification.In this examination, we propose both regulated what's increasingly, solo AI frameworks to improve tumor depiction. Our first methodology depends on supervised learning for which we exhibit critical increases with profound learning calculations, especially by using a 3D Convolutional Neural Network and Transfer Learning. Convinced by the radiologists' interpretations of the ranges, we by then advise the most ideal approach to intertwine task subordinate component depictions into a CAD structure by methods for a chart regularized small Multi-Task Learning(MTL) framework. In the ensuing philosophy, we examine a performance learning count to address the confined availability of checked getting ready data, an average issue in therapeutic imaging applications. In our framework we created Lung Cancer identification framework dependent on AI and profound neural system. It diminishes the odds of getting mischief to human lives by early discovery of malignant growth. By and by a couple of structures are proposed and still an enormous number of them are hypothetical arrangement. Convolutional Neural Network based Classification and area game plan of lung tumor.

Keywords: *Convolutional Neural Network, Lung Cancer Disease, Supervised Learning, Unsupervised Learning*

Introduction

Due of enormous pervasiveness of smoking and air contamination around the globe, lung malignancy has gotten one of the most widely recognized and lethal ailment in ongoing decades. It generally requires some investment to imagine and most patients are determined to have the illness inside the age bunch 55 to 65. Early recognizable proof and treatment is the best accessible choice for the tainted individuals. Solid recognizable proof and order of lung malignant growth requires neurotic test, in particular, needle biopsy example and examination by experienced pathologists. In any case, because it incorporates human judgment of a couple of components and a blend of experiences, a decision sincerely strong system is appealing for this circumstance. On-going enhancements in picture planning, structure affirmation, dimensionality reduction and gathering procedures has prepared for substitute ID and request approaches for lung harmful development. Notwithstanding AI draws near, profound learning through confined Boltzmann machine as auto encoders has demonstrated promising accomplishment in order errands in various space including acoustics, supposition grouping, and picture and content acknowledgment. Persuaded by the achievement of profound learning in applicable fields, a profound learning based grouping technique is explored in this work. In which framework, utilized directed learning and unaided learning approaches with NSCLC Radio genomics lung malignancy CT picture dataset. In the wake of getting this DICOM restorative organized pictures those changed over by DICOM converter to PNG position.

Picture securing contains picture perusing by utilizing opencv-python for process it. In the wake of getting picture of lung malignant growth it move into clamor decrease for expelling commotion from it. At that point picture preparing methods applied on it like paired picture change and dim picture transformation followed by division. The picture highlights get gathered and drafted into machine model in preparing period of AI for future forecast of lung malignant growth and stages assessments. We have dealt with directed dataset and utilized profound Convolutional neural system for accomplishing high exactness.

Objective and Scope

1. Main goal behind to build up a framework causes the specialists to cross confirm their analyzed outcomes which gives promising arrangement over existing demise rates.

By utilizing our proposed work attempt to imagine interesting stage and most encouraging answer for early analysis of different ailments.

2. Existing work examination exactness is decreased when the nature of restorative information is fragmented. Also, various areas display exceptional qualities of certain provincial sicknesses, which may debilitate the forecast of infection wrong.

3. So we are giving progressively precise arrangement by utilizing AI and Convolutional neural system to identify maladies and make forecasts.

Problem Statement

- To diminish demise rate by interminable illnesses on the planet.
- To give exceptional answer for lung malignancy ailments.
- To give higher exactness over past research.
- To give most encouraging instrument that can worthy by every one of the specialists.
- Detection of numerous sorts of Diseases.

Review of Literature

Qing Wu and Wenbing Zhao [1] state that An EDM AI calculation with vectored histogram highlights to distinguish SCLC for early malevolent malignancy expectation. While we demonstrate that EDM has sensibly great expectation precision, there is a huge opportunity to get better before our calculation can be utilized in the clinical setting. The extreme objective of this investigation is to build up a clinical basic leadership framework for radiologists to all the more likely anticipate a noxious lung disease from SCLC with computed tomography (CT) imaging. For the future work, we would prepare the proposed strategy with bigger preparing set what's more, more profound system, and consolidate it with convolution neural network, which has been utilized in CT imaging for various applications.

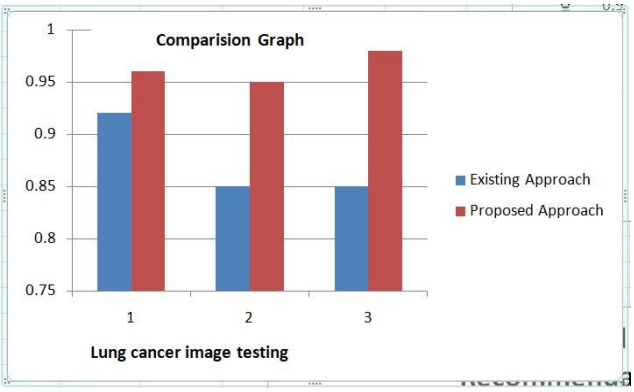
Lilik Anifah et.al [2] proposed that detection of lung cancer utilizing Artificial Neural Network Back-engendering based Gray Level Co-event Matrices (GLCM) highlight. The lung information utilized begins from the Cancer imaging archive Database, information utilized comprised of 50 CT-pictures. CTpicture is gathered into 2 bunches, typical and lung disease. The means of this exploration are: picture pre-processing, locale of intrigue division, highlight extraction, and recognition of lung disease utilizing Neural Network Back-spread. The outcomes demonstrate framework can identify CT-picture of ordinary lung and lung malignancy with exactness of 80%. Early discovery of lung malignant growth will recoup the patient. Instrument used to recognize lung malignant growth is through CT Scan (Computed Tomography).

Prof. Anuradha Deshpande and Dhanesh Lokhande [3] proposed that to locate the beginning times of lung malignancy and increasingly exact outcome by utilizing distinctive strategies like combination, upgrade and division process. Already the vast majority of the malignant growth location strategies relies upon human experience by watching the picture of CT-filter. It will be a bogus discovery of lung malignant growth arrange. Utilizing Image Processing we can rapidly and precisely distinguish tumor of malignant growth. Utilizing Image Processing viable procedures we gather data from complex restorative pictures. In combination procedure, the critical highlights of various unique pictures are consolidated together to acquire the required data in a Fused Image. In medicinal application there are different plans to enhance the substance of picture shape CT and MRI like CT picture examines the denser tissues and MRI filters the delicate tissues, so by joining pertinent data of the two pictures, we get proper data of melded picture. This procedure additionally enhances the nature of the melded picture. Rachid Sammouda [4] introducing for improved technique for Hopfield Artificial Neural Network Classifier show is proposed to section extricated lung locales from human chest Computer Tomography pictures. The pictures are procured utilizing Computer Tomography imaging methods from typical subjects and others as possibility for lung malignant growth determination. A blend of bit-planes of every pixel is utilized to upgrade edges' recognition of lung area flaps. Three indicative guidelines are confirmed too characterized channels of applicant malignant locales from the status of possibility to false or genuine positive status. Abbas Khosravi and Amin Khatami [5] proposed a novel neuralarrange based calculation, which we allude to as entropy debasement technique (EDM), to recognize little cell lung malignant growth (SCLC) from processed tomography (CT) pictures. This exploration could encourage early identification of lung malignant growths. The preparation information and testing information are high-goals lung CT examines given by the National Cancer Institute. We chose 12 lung CT filters from the library, 6 of which are for sound lungs, and the rest of the 6 are examines from patients with SCLC. We arbitrarily take 5 filters from each gathering to prepare our model, and utilized the staying two outputs to test. Our calculations accomplish a precision of 77.8%.

Comparison of Proposed Work

The existing work done for lung cancer detection having less accuracy and less detection rate.In proposed work we are focused on lung cancer detection accuracy by using neural network. The proposed framework introducing a novel Lung discovery and Stage expectation system Supervised and Unsupervised Learning Approaches. Proposed which at first adjusts significant features and a short time later readies a phony neural framework with

these academic features. Exploratory results show the significant taught classifier beats each and every other classifier when arranged with all characteristics and same getting ready tests. It is furthermore demonstrated that the introduction improvement is truthfully gigantic. Portrayal of lung infection using a low people, high dimensional instructive file is attempting a result of lacking guides to get comfortable with a definite mapping among features and class names. In Proposed framework comprise of 2 stages regulated and unaided learning draws near. Current composition generally handles this endeavor through excellent component creation and decision. Profound learning is seen as ready to distinguish the basic structure of information using CNN and different systems. This shows utilization of AI can possibly altogether distinguish and characterize with practically high precision for the low populace in India .High dimensional lung malignant growth informational index without requiring any hand-made, case explicit highlights. High handling velocity improved CNN classifier model.



Graph5.1. comparison graph

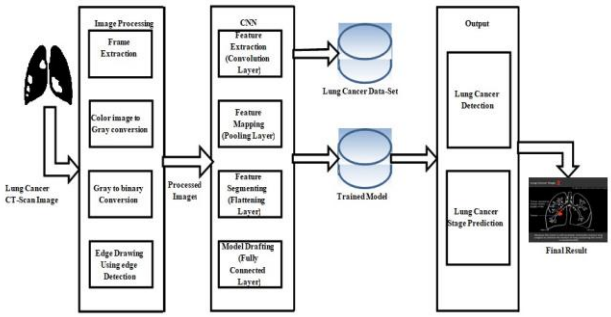


Figure No. 1 System Architecture

In CNN model comprise of following stages

• Image Processing :-

A picture is comprised of RGB hues. Pre-handling unit comprises of commotion evacuation, dark scale transformation, twofold change of pictures followed by include extraction. In future extraction five stages followed in which fingertips look by unpredictability. Next lengthenings of pictures are estimated by considering pixel division just as turn of pictures

• Image Filtering :-

Separating is a procedure to change or improve the picture, for example to feature certain highlights or expel different highlights. It incorporates smoothing, honing and edge upgrade. Picture separating calculations create a yield pixel by watching the area of the info pixel in a picture. Picture sifting calculations are utilized to expel various sorts of clamor from the picture.

• Feature Extraction:-

In include extraction, algorithmic investigation used to discover the element vectors of orderly outcomes consolidates K ebb and flow and arched body calculations. In present work

"K curved frame" calculation which is utilized to distinguish fingertip with more noteworthy precision. In our framework, CNN is utilized for future acknowledgment in which we having the information unit of preparing informational collection of pictures.

• Segmentation :-

Picture division is the path toward allocating a propelled picture into different parts (sets of pixels). All pixels in a region share a commonplace property. Least unpredictable property that pixel can share power. The goal is to unravel and change the depiction of the image into something that is progressively significant and less requesting to separate.

• Edge Detection :-

Edge characterizes the limits between districts in a picture which helps in object location. There are many edge discovery administrators and calculations accessible. Edge Detection Operators and Algorithms utilized in our exploration like arched structure technique.

• Feature Recognition:-

Mind motivated frameworks used to reproduce how people learn. Comprise of information, covered up and yield layers that change the contribution to something that the yield layer can utilize. Great for discovering designs which is perplexing to human for concentrate and show the machine to perceive. CNN assembles their insight by identifying the examples and connections in information and learns (or is prepared) through experience, not from programming.

Mathematical Model

• **System Description:**

S= {I, F, O}

INPUT:

- F=F1,F2,F3...FN Function to execute result
- I=C1,C2,C3... input of systems MRI images and symptoms reports
- O=R1,R2Rn
- I=result access by User
- C1=lung cancer recognition

F:

- F1=Image processing applied on Lung Cancer images
- F2=feature extraction from images **O:**
- R1= model creation from training.
- R2= model based image testing

SPACE COMPLEXITY:

The space complexity depends on Presentation and visualization of discovered patterns. More the storage of data more is the space complexity.

TIME COMPLEXITY:

We are going to use CNN for fast and better recognition with higher accuracy. So time complexity is less. So the time complexity of this algorithm is O(n^n).

Success:

1. High accuracy achieved by using all type of MRI image dataset.
2. User gets result very fast according to their needs.

Failures:

1. Huge database can lead to more time consumption to get the information.
2. Hardware failure.
3. Software failure.

Mathematical Model in Equation format

Notation

Where,

- M= Set of all entities.
- LCIT1= Lung cancer images type 1
- LCIT2= Lung cancer images type 2
- LCITN= Lung cancer images type N
- TLCI=Total Lung cancer images

For calculate total number of users by following equation 1

Total number lung cancer images = Total number lung cancer images type 1+ = Total number lung cancer images type 2+.....+ = Total number lung cancer images type N

∑ TLCI = ∑ LCIT1+ ∑ LCIT2+...+∑ LCITN……equation 1

Algorithms in Pseudo Code

Conventional Neural Network (CNN)

• Image handling utilizing Convolutional neuronal systems (CNN) has been achievement completely utilized in different fields of action, for example, geo strategies, structural designing, mechanics, modern observation, insubordination division, automatics and transport. Picture pre-handling, date decrease, division and acknowledgment are the procedures utilized in overseeing pictures with CNN.

• Each information neuron addresses concealing information in the image, and each yield neuron analyzes to an image. All pictures will be scaled to a similar size (width and tallness) and little to rush to learn. On the spans of the pictures will be resolved on the size of the information vector and the quantity of neurons. The exchange work for this sort of issue is called sigmoid capacity. The pace of learning has values in the range [0.1] and the mistake it is prescribed to have under 0.1.

• Processing of pictures with CNN includes various procedures, for example,

• Image pre-setting up, an action which shows a picture (separate redesign, uproar decline) with unclear estimations from the primary picture. The objective of pictures pre-dealing with CNN involves in improving, restoring or redoing pictures. The settled issues are the cartographic sorts, to update a limit, a gauge work for the diversion of an image.

• Data abatement or feature extraction incorporates expelling different features more diminutive than the amount of pixels in the data window. The movement contains in compacting the image followed by isolating geometric traits (edges, corners, and joints), facial features, etc.

• Segmentation is a division of a picture into locales.

• Recognition includes the assurance of articles in a picture and their characterization.

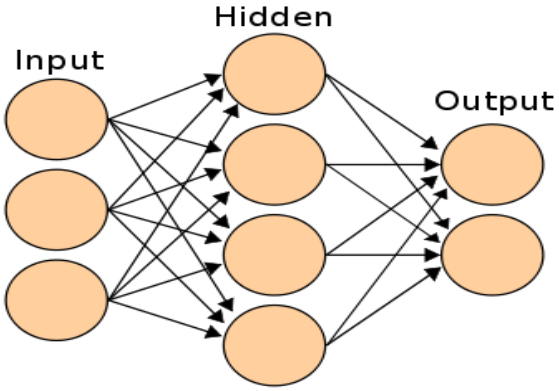


Figure No.2 Block diagram CNN

In CNN consist following 4 steps as shown in figure no.3

- Convolutional Layer
- Pooling
- Flattening
- Fully Connection

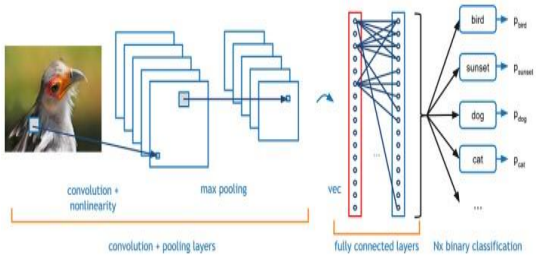


Figure No.3 Steps of CNN

1. Convolutional Layer

In convolution layer, extract different features pixel wise by using feature detectors/kernels. Perform numerous convolutions on input, where each operation

uses a different filter. This results in different feature maps. In the end, we take all of these feature maps and put them together as the final output of the convolution layer.

2. Pooling

In pooling function of pooling is to continuously reduce the dimensionality to reduce the number of parameters and computation in the network. This shortens the training time and controls over fitting Max Pooling extracts out the highest pixel value out of a feature while average pooling calculates the average pixel value that has to be extracted.

3. Flattening

In flattening steps shows that basically here we arrange the pooled feature into a single vector/column as a input for next layer (convert our 3D data to 1D).

4. Fully Connection

In last fully connection Neurons in a fully connected layer has full connections to all the activations in the previous layer.

Combining more neurons to predict more accurately.

Convolutional Neural Network Approach (CNN)

Step 1 - Input hand gesture's
Step 2 - Image capturing by using open-cv
Step 3 - Image processing
Step 4 - Feature Extraction from images
Step 5 - Model generation
Step 6 - Sign recognition
Step 7 - text generation
Step 8 - voice conversion by using GTTs

Tensor Flow

Machine learning is an unpredictable order. Be that as it may, actualizing AI models is far less overwhelming and troublesome than it used to be, because of AI structures.

- Such as Google's Tensor Flow—that facilitate the way toward getting information, preparing models, serving expectations, and refining future outcomes.
- Tensor Flow gives the entirety of this to the software engineer by method for the Python language. Python is anything but difficult to learn and work with, and gives advantageous approaches to express how elevated level reflections can be coupled together. Hubs and tensors in Tensor Flow are Python items, and Tensor Flow applications are themselves Python applications.

Open-CV

- Monty Python's Flying Circus had a "cat detector van" so, in this tutorial, we use Python and Open-CV to make our very own cat detector and recognizer. We also cover examples of human face detection and recognition. More generally, we cover a methodology that applies to training a detector (based on Haar cascades) for any class of object and a recognizer (based on LBPH, Fisher-faces, or Eigen-faces) for any unique objects. We build a small GUI app that enables an LBPH-based recognizer to learn new objects interactively in real time. Although this tutorial uses Python, the project could be ported to Android and IOS using OpenCV's Java and C++ bindings.
- Attendees will pick up involvement with utilizing OpenCV to distinguish and perceive visual subjects, particularly human and creature faces. GUI improvement will likewise be stressed. Participants will be guided toward extra data in books and on the web. There is no proper assessment of participants' work yet participants are welcome to exhibit their work and talk about the outcomes they have accomplished during the session by utilizing various finders and recognizers and various parameters.

Python

- Python interface is being actively developed right now. There are many algorithms and many functions that compose or support those algorithms. Open CV is written natively in C++ and has a template interface that works seamlessly with STL containers.

Image Processing

- Read and Write Images. Detection of images and its features. Detection of shapes like Circle, rectangle etc in a image, Detection of coin in images. Text recognition in images. eg. Reading Number Plates. Modifying image quality and colours.

Comparative Results

Stages of Lung Cancer:-

Stage I : Stage 1 is a part of number staging system and means your cancer is small. it hasn't spread to your lymph nodes and other distinct organ.

- They are of two types :
  - Type A :- Means that the cancer is 2 cm or smaller.
  - Type B :-Means that the cancer is between 2 to 3 cm

Table 9.1: Model Testing

Sr No.	Types	Number of images
1	Lung Cancer images type 1	210
2	Lung Cancer images type 2	175
3	Lung Cancer images type 3	385
4	Lung Cancer images type 4	230

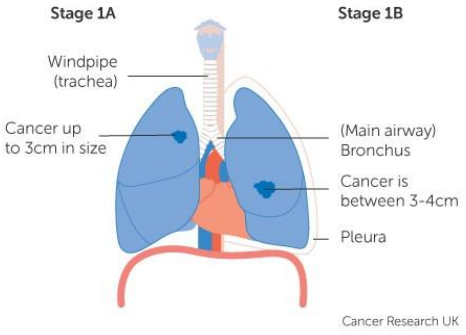


Fig No.4 Lung Cancer Stage 1

▪ **Stage II** : These tumors are larger than the stage 1 tumors and start to spread to nearby lymph nodes or other structures. About 30% of the lungs cancer are diagnosed at these stage,and the treatment success rate is higher than the later stage

.

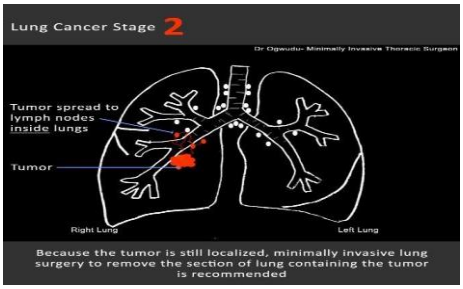


Fig No.5 Lung cancer Stage 2

▪ **Stage III** : In this stage the tumoris more than 3 cm wide and is spread in the lymph nodes and other parts outside the lungs. in this stage the affected lymph nodes are restricted to the same side of the body as the tumor.
▪ **Stage IV** : The cancer is metastasized or spread beyond the lungs into other areas of the body.

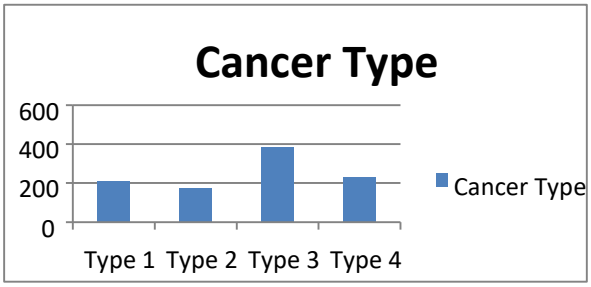
Table 9.2. Lung cancer stage evaluation

Cancer Area(T)	Criteria
T1	< 3cm in diam; T 1a <= 2cm; T 1b > 2cm <= 3cm
T2	> 3cm <= 5cm; T 2a > 3cm <= 4cm; T 2b > 4cm <= 5cm
T3	>5cm<=7cm
T4	Any Size greater than above

In our experimental setup, in table no.8.1, shows total number of 1000 images of lung cancer. In our project contains mainly Lung cancer images type 1, Lung cancer images type 2, Lung cancer images type 3, and Lung cancer images type 4. In this project consist of 210 numbers of images of lung cancer type 1, 175 numbers of images of lung cancer type 2, 385 numbers of images of lung cancer type 3 and 230 numbers of images of lung cancer type 4.

Results

From above data, as shown in graph 9.1, the total numbers of images of lung cancer type 1 were 210, total numbers of images of lung cancer type 2 were 175, total numbers of images of lung cancer type 3 were 385 and total numbers of images of lung cancer type 3 were 230.



Graph 10.1: Total number of cancer type

Conclusion

Hence we actualized the Lung malignancy location framework to precisely assessment of the phases with 2 stages learning approaches like supervised learning and unsupervised learning. Profound neural system approach is utilized to accomplish more prominent precision in discovery of lung malignancy and precisely anticipate stages. We demonstrated that utilization of AI can possibly fundamentally distinguish and group with nearly exactness for the low populace, high dimensional lung disease dataset without requiring any handmade, case explicit highlights.

Future Work

Future work will be based on real time lung cancer image dataset with considering high accuracy over proposed work.

Acknowledgment

This work is supported in a lung cancer disease diagnosis using two-step learning approach research field in India. Authors are thankful to Faculty of Engineering and Technology (FET), Savitribai Phule Pune University, Pune for providing the facility to carry out the research work.

References

[1] Qing Wu and Wenbing Zhao "Small-Cell Lung Cancer Detection Using a Supervised Machine Learning Algorithm" International Symposium on Computer Science and Intelligent Controls2017
[2] Lilik Anifah, Haryanto,Rina Harimurti, "Cancer lung detection on CT scan image using ANN backpropagation based gray level co occurrence matrix feature." 978-1-5386-3172-0/17/ 2017 IEEE
[3] Prof. Anuradha Deshpande, Dhanesh Lokhande, "Lung cancer detection with fusion of CT and MRI image using image processing." (IJARCET) Volume 4 Issue 3, March 2015
[4] Rachid Sammouda, "Segmentation and analysis of CT chest images for early lung cancer detection." Global Summit on Computer & Information Technology 978-1-5090-2659-3/17 2017 IEEE
[5] Abbas Khosravi, Amin Khatami, "Lung cancer classification using deep learned features on low population dataset." Canadian Conference on Electrical and Computer Engineering (CCECE) 978-1-5090-5538-8/17 2017 IEEE