International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-4, November 2019

A Deep Convolutional Neural Network Framework for Lung Cancer Detection

Hemalatha Eedi

Abstract*: Exact identification of pulmonary nodules with high sensitivity and specificity is basic for programmed lung malignancy analysis from CT scans. In fact, many deep learning-based algorithms gain incredible ground for improving the exactness of nodule recognition; the high false positive rate is yet a difficult issue which restricted the programmed determination. We propose a novel customized Deep Convolutional Neural Network (DCNN) architecture for learning high-level image representation to achieve high classification accuracy with low variance in medical image binary classification tasks. Moreover, a High Sensitivity and Specificity system is introduced to eliminate the erroneously recognized nodule competitors by following the appearance changes in consistent CT slices of every nodule. The proposed structure is assessed on the open Kaggle Data Science Bowl (KDSB17) challenge dataset. Our strategy can precisely distinguish lung nodules at high sensitivity and specificity and accomplishes 95 % sensitivity.

Keywords*: Deep Convolutional Neural Network, Kaggle Data Science Bowl, Lung Cancer, Sensitivity and Specificity.

I. INTRODUCTION

This, as per the World Health Organization (WHO) the lung malignancy is the second most regular disease in the two people. Prostate disease is progressively normal in men and chest harmful development in women [1]. Almost 13% of new malignant growths are incorporates lung diseases. According to American Cancer Society [2] the assessed lung disease for the year 2019 are around 228,150 new instances of lung cancer (116,440 in men and 111,710 in ladies) and around 142,670 passing from lung cancer (76,650 in men and 66020 in women). Cancer are of various sorts like lung, colon, bosom and prostate tumours. Every year, a bigger number of people kick the container in perspective on lung cancer development than other kind.

Thinking about this measurement and the likelihood of increment the existence quality or solution for early recognition, it is critical to contribute on frameworks for early lung malignant growth location. Normally when lung malignancy ended up symptomatic it is ahead of time organize, so to contribute on hazard gatherings screening could diminish the quantity of passing [3]. One can expand the symptomatic exhibition and the survival of patients by partner a MATLAB framework to the standard indicative system. Various ways to excess with MATLAB frameworks have been connected utilizing neural systems [4] indicating promising outcomes. This medicinal field of research with picture investigation for early recognition of malignant growth can be regularly ordered in two sorts of assignment: division; to decide a suspicious area which distinguished as disease and grouping; that utilization to conclusion and appraisal of a sort of malignant growth.

There has been an extending excitement among the specialist in helpful picture to make lung tumour growth analysis techniques subject to significant learning strategies is an improved type of neural frameworks, which contains a couple of layers to deliver high-organize features from its data and a while later, draws out the foreseen a motivating force on the most elevated purpose of the framework. Among profound learning methods, convolutional neural systems (CNNs) [5] have been generally connected in PC vision assignments. The last consequences of convolution system have demonstrated its shape in various pictures. As of late, the connected profound learning systems are developed additionally in restorative picture investigation task. The profound learning method, proposed to improve the exhibition of medicinal pictures [6], have been connected either by adjusting the design of the current profound learning systems, or proposing new ones. To validate the performance of DCNN, we test it on KDSB17 [7] and compare the effectiveness with other dataset competitors.

II. LITERATURE SURVEY

A considerable lot of the current framework indicates different strategies for recognizing the lung carcinoma. One of the proposed networks by Sandeep Kumar Saini [8] recommends the utilization of ACO method for upgraded characterization and fluffy rationale for the testing of the highlights separated for its precision of dangerous image. In this framework the lung CT picture is connected to pre-processing and afterward includes extraction is finished utilizing the Binary order.

Comparably Zakaria Suliman Zubi [9] proposes the strategy of summed up Rule Mining and Neural Networks to recognize the trademark highlights of the medicinal pictures and group them into destructive and non-malignant cells. The CT check chest films which are put away in tremendous sight and sound databases for a therapeutic intention is connected to some

Manuscript published on November 30, 2019.
* Correspondence Author
Hemalatha Eedi*, Department of Computer Science and Engineering, JNTUH College of Engineering Hyderabad, Hyderabad, India. Email: hemamorarjee@jntuh.ac.in

picture acknowledgment techniques to remove helpful learning and standards from the database. These standards are given to the neural system to characterize the database and help the specialists to settle on a patient's condition.

The different systems that can be utilized for the discovery of lung malignant growth, for example, Artificial Neural Network [10], Naïve Bayes Classifier, Decision-Tree and If-Then Rules.

Choice tree is a prescient model utilized for statistics data mining and machine learning [11]. This calculation chips away at top-down methodology and parts the arrangement of things at each progression. The Artificial Neural Network is gathered hubs called neurons and the yield of every neuron is processed by non direct capacity of the aggregate of its info. The outcomes are gotten after characterization and the forecast lung malignancy framework is utilized for early discovery which can spare a patient's life. Another proposed structure that performs Association Rule mining assessment on lung malignant growth information to perceives the hotspots in the infection data, where the survival time of patient is by and large higher or lower than the ordinary survival time over the entire dataset.

The SEER (Surveillance, Epidemiology and End Result) [12] is just the thorough wellspring of populace-based data that incorporates phase of malignant growth at the season of determination and patient's survival information in the United States. The essential piece of SEER is the Quality Control and culmination of information that has been accounted for.

The following proposed network is Apriori calculation is for learning affiliation principles and intended to work on databases which gathers client's profile with successive thing sets. It uses base up methodology where continuous subsets are broadened one thing at once and the information tried on gathering of competitors

III. PROPOSED NETWORK

In this area, we present the customized DCNN arrange engineering in detail in Fig.1.We initially portray preprocessing stage with generally structure of DCNN and after that clarify our misfortune capacity used to prepare the DCNN model.

The proposed DCNN system portrays pre-preparing stage with generally structure of DCNN and afterward clarifies our misfortune capacity used to prepare the model. Here presents lung disease discovery dependent on chest CT pictures utilizing CNN. From the start lung districts are separated from CT picture and in that area to get tumors every cut is sectioned. This fragmented tumor areas are utilized to prepare engineering of CNN. After that CNN is utilized to test the patient pictures as shown in Fig. 1. The fundamental goal of this examination is to distinguish whether the tumor present in a patient's lung is harmful or benevolent.

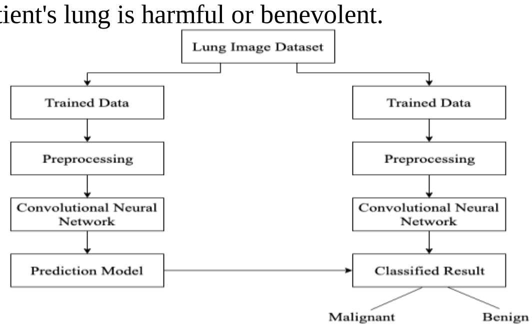

Fig. 1. Block Diagram of Proposed Network

3.1 Preprocessing

During preprocessing we use sifting method for smoothing, honing and improving system. It might be connected by utilizing:

1. Spatial Domain
2. Frequency Domain

In preprocessing stage, the middle channel is utilized to reestablish the picture under test by limiting the impacts of the corruptions. The middle channel basically replaces every pixel esteem with the middle estimation of its neighbors including itself and the pixel esteems that is unique in relation to their neighbors will be wiped out.

As the first information picture is size of 512 × 512 pixels and subsequent to preprocessing the size of picture will be 28 × 28 pixels appeared in Fig. 2(a) and 2(b).

Fig. 2(a). Input Image

Figure 2(b). Preprocessed Image

3.2 The Overall Network

The proposed DCNN configuration principally involves the going with layers: three convolution layers which seek after two max-pooling layers, and two totally connected layer with two fragile max units.

As showed up in Fig. 3 the framework begins with two convolution layers, in which the convolution layer instated with feature map, height, width of the image with data size of 28 × 28 pixels. The central convolution layer contains 40-part maps with the convolution bit of 5 × 5.

The consequent convolution layer contains 50 component maps with the convolution bit of 3 × 3. The piece size for max pooling layers is 2 × 2 and the stroll of 1 pixel, and the totally related layer makes a yield of 10 estimations. These 10 yields are then passed to another totally related layer containing 2 sensitive max units, which address the probability that the image is containing of the lung harmful development or not. Note that each convolution layer in our DCNN model is trailed by a revised direct unit (ReLU) layer to make their yields.

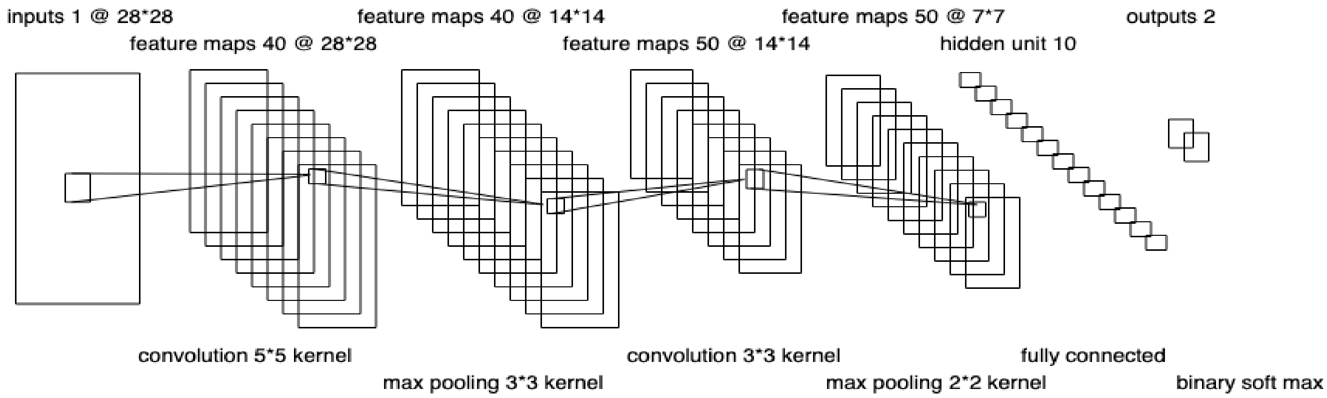

Fig. 3. Scheme of our Proposed DCNN Architecture

3.3 Loss Function

SoftMax loss and cross-entropy loss terms are utilized reciprocally. The SoftMax classifier is a direct classifier that uses the cross-entropy [13] misfortune work. the inclination of the above capacity advises a SoftMax classifier how precisely to refresh its loads utilizing some streamlining like slope plummet. The SoftMax function part essentially standardizes our system expectations with the goal that they can be deciphered as probabilities. When our system is anticipating a likelihood dissemination over marks for each information, the log misfortune is comparable to the cross entropy between the genuine name dispersion and the system expectations. Rather than choosing one greatest worth, it breaks the entire (1) with maximal component getting the biggest bit of the dissemination, yet other littler components getting some of it too.

Cross entropy shows the separation between what the model accepts the yield distribution to be, and what the original distribution is. Cross entropy measure is a generally utilized option of squared error. Thus, it is utilized as a loss function in neural systems which have SoftMax enactments in the yield layer.

The loss could be calculated by,

`

$$L = e^{a_i} / \sum_{i=1}^{k} e^{a_i} \tag{1}$$

Where L is approximated maximum-function, a is the activation function and K is the number of classes (which is equal to 1 at this work). And finally, the gradients are computed by standard backpropagation of the error [5].

3.4 Training CNN

Back-propagation calculation [14] is utilized to prepare the Deep CNN to recognize lung tumors in CT picture of size 28×28×14. It comprises of two stages. In the main stage, a CNN comprises of numerous volumetric convolutions, corrected straight units (ReLU) and max pooling layers is utilized to separate important volumetric highlights from information.

The subsequent stage is the classifier. It has various FC and limit layers, trailed by a Soft Max layer to play out the abnormal state thinking of the neural system. During preparing, the arbitrary sub-volumes removed from the CT pictures of the preparation set and are standardized by a gauge of the typical dispersion of the pixel esteems from the dataset.

IV. PERFORMANCE MEASUREMENT

The presentation of a strategy, in therapeutic picture investigation, is ordinarily estimated by particularity, affectability, and F1 score.

Affectability estimates the extent of real positives tests that accurately distinguish, in which the level of harmful knob that is effectively named destructive. In this manner, it is figured by the accompanying definition:

$$Sensitivity = \frac{TP}{TP + TN} \tag{2}$$

Where TP (True positive) is the quantity of knobs which have been effectively recognized and FN (false negative) is the quantity of knobs which have been identified by the technique. Interestingly, explicitness estimates the extent of distinguished negatives tests, in which the rate without malignant knob is accurately delegated non dangerous.

In this manner, specificity is computed as:

$$Specificity = \frac{TN}{TN + TP} \tag{3}$$

where TN (True negative) is the quantity of non-disease patients which have been effectively characterized, and FP (false positive) is the quantity of non-malignancy patients which have been wrongly delegated malignancy.

And, F1-score measures the average F1 score through different class labels which are computed as:

$$F1 = 2 \times \frac{PPV \times TPR}{PPV + TPR} \tag{4}$$

Where, $PPV = \frac{TP}{TP+FP}$ and $TPR = \frac{TP}{TP+FN}$

The cross-entropy and log loss are gently change according to setting, while in AI they carry on in same way during mistake rate figuring somewhere in the range of 0 and 1.

we can compose cross-entropy as:

$$Loss(p,q) = -\sum p_i \log q_i = -y \log \hat{y} - (1-y)\log(1-\hat{y}) \tag{5}$$

Where, p = set of true labels
q = set of prediction
y = true label

$\hat{y}$ = predicted prob

We must show that,

$$\text{if mini batches} \begin{cases} < 1 \text{ , cancer class} \\ = 1 \text{ , Non cancer class} \end{cases}$$

Need to evaluate the graphical representation that how loss error decreases during the optimizing network parameter and how accuracy increases by using mini batches.

V. EXPERIMENTAL EVALUATION

We assessed the DCNN model on Kaggle Data Science Bowl 2017 informational collection, which as of late have been discharged for a challenge on lung malignant growth pictures. We likewise contrasted our outcome and various contenders by displaying their outcomes.

5.1 Data set for training and testing

The data about patients with CT sweep pictures contained in the KDSB17. The pictures of this informational index are in size of 512 × 512 pixels. This dataset incorporates CT outputs of lung malignant growth patients and furthermore non patients. We utilized 100% of the pictures for preparing, 70% utilized for cross-approve set and test our model. We have actualized DCNN by utilizing MATLAB 2016a with Deep Learning tool compartment. We re-scaled every one of the pictures of this informational collection to the size of 28 × 28 pixels before preparing stage, on account of memory space constraint on GPU.

5.2 Implementation details

Preparing information pictures are arbitrarily isolated into little clusters. The model performs forward engendering on every short cluster and figures the yield and misfortune. At that point, back spread is utilized to process the slopes on this cluster, and system loads are refreshed. We perform stochastic angle drop to perform weight refreshes, and the misfortune capacities is limited by stochastic inclination plummet with the size of group 30 which has been done on 1000 cycles on the example. Likewise, we try a couple hyper parameters and pick the learning rate 0.01 and utilize a force of $\mu = 0.9$.

5.3 Output Result Analysis

The neural system dependent on convolutional and middle sifting has been executed in MATLAB and the framework is prepared with test informational indexes for the model to comprehend and acclimate the lung disease. An example picture has been sustained as a contribution to the prepared model and the model at this stage can tell the nearness of malignancy. The procedure includes the encouraging the information picture, preprocessing, highlight extraction, distinguishing the malignant growth and show the outcomes to the client. If there should be an occurrence of the harm is available, a message showing the nearness of will be shown on the screen alongside the given info picture as shown in Fig. 4(a) and Fig. 4(b).

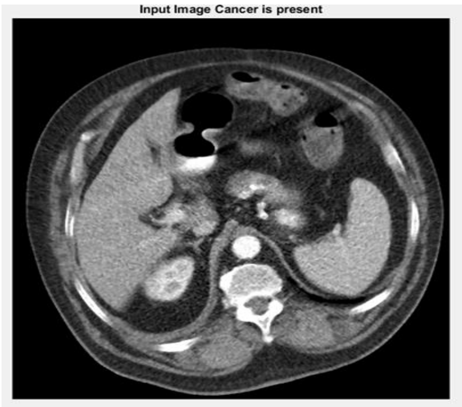

Fig 4(a). Output Image for Cancer cell

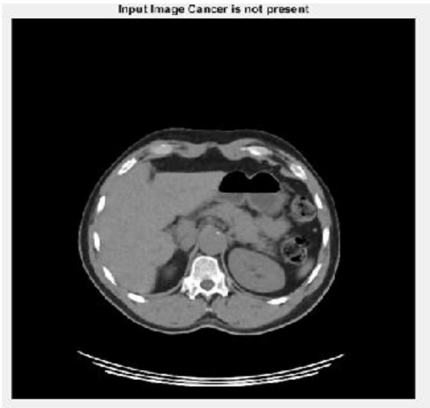

Fig 4(b). Output Image for Non-Cancer cell

The identification of lung malignant growth cells utilizing DCNN displayed by sorts of learning, for example Introduce the quantity of age, inclination minute, concealed neurons and learning rate. A portion of the parameter utilized for preparing the model of the neural system is appeared in Table-I.

Table-I: FC layer with computing the class score

Parameter	Value
Epoch	50
Learning Rate	0.01
Gradient moment	0.9
Hidden Neurons	100

CNN has two layers, for example, 2 convolution layers and 2 sub inspecting layer which is utilized to expand the precision of location. The perplexity network parameters got from CNN yield are given in Table-II.

Table-II: Confusion matrix for the proposed method in lung cancer diagnostic task

Number of testing images=3500	Predicted class Yes	Predicted class No	Predicted Result
Actual class Yes	TP = 2500	FP =70	Accuracy = 94%
Actual class No	FN = 130	TN = 800	Error= 5.7%
Actual Result	Sensitivity = 95%	Specificity =91.9%	

Table-III reports the whole outcomes with three referenced measurements think about the aftereffects of particularity, affectability, and F1-score

Retrieval Number: D9764118419/2019©BEIESP
DOI:10.35940/ijrte.D9764.118419
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

with LUNA16 group. High Sensitivity/low Specificity Test is called mammography test.
It is useful for identifying real instances of the infection with decreased particularity. F measure gives us the trial of exactness, on the off chance that its worth ranges to 1, at that point it is flawless accuracy or Recall (it takes both FP and FN with real positive).

Table -III: The result of our Proposed System

	Sensitivity	Specificity	F-measure
LUNA 16	0.92	0.95	0.067
KDSB 17	0.95	0.919	0.798

The log loss of Lung cancer detection with the proposed convolutional neural network-based method was compared with that obtained by previous works in Table-IV.

Table-IV: Comparison of Log Loss

Name	Log –Loss
LUNA 16	0.47
grt 123	0.39
NLST DS	0.38
Ours	0.36

Figure 4 shows precision and loss of DCNN during preparing stage which outlines how misfortune blunder diminishes during the advancing system parameters and how exactness increments are a direct result of utilizing little clump rather than full-group in profound neural systems, for example, DCNN.
The quantity of test pictures chosen to get ideal outcome for precision. The high number of CT information enabled us to utilize a moderate small-scale clump size and manage preparing procedure to increase fitting rate and effectiveness.

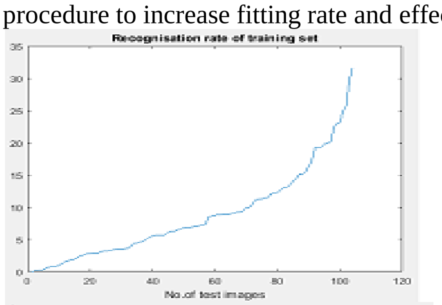

Figure 4.a The recognition rate of cross validation set

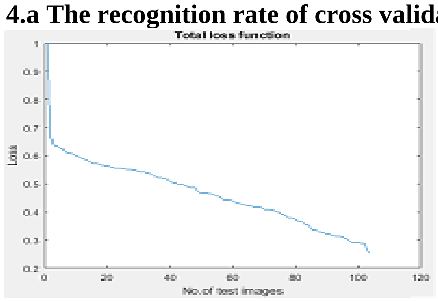

Figure 4.b Total loss function with respect to mini batches

VI. CONCLUSION

The modified possess profound convolutional neural system design for double characterization of lung disease to recognize the danger tissues present in the info lung CT output picture. The picture with various size, state of the lung harmful tissues has sustained at the information layer for preparing the framework. The customized framework can distinguish the nearness and nonattendance of dangerous cells with sensitivity of about 95%. The log loss of Lung tumor growth identification with the proposed convolutional neural network-based strategy was contrasted. The general precision of the framework can be improved utilizing 3D Convolutional Neural Network and furthermore by improving the concealed neurons with profound system.

VII. FUTURE SCOPE

Consequently, structured this system with the assistance of Le-Net utilizing up to 7 layers for some example pictures as a result of framework execution. If picture tests will increment up to lakhs, then we can go for multilayer neural system intended to perceive visual example straightforwardly from pixel picture with insignificant pre-processing. The systems might be Alex Net, ZF Net, Google Net, VGG Net , ResNet and so forth.

REFERENCES

1. W. H. Organization et al. Global health observatory (gho) data 2015. Retrieved, 3:23, 2015.
2. R. L. Siegel, K. D. Miller, S. A. Fedewa, D. J. Ahnen, R. G. Meester, A. Barzi, and A. Jemal. Colorectal cancer statistics, 2017. CA: a cancer journal for clinicians, 67(3):177–193, 2017.
3. A. Jemal, P. Vineis, F. Bray, L. Torre, and D. Forman. The cancer atlas. American Cancer Society, 2016.
4. R. S. Shriwas and A. D. Dikondawar. Lung cancer detection and prediction by using neural network. International Journal of Electronics & Communication (IIJEC), 3(1), 2015.
5. A. Teramoto, T. Tsukamoto, Y. Kiriyama, and H. Fujita. Automated classification of lung cancer types from cytological images using deep convolutional neural networks. BioMed Research International, 2017, 2017.
6. D. Kumar, A. Wong, and D. A. Clausi. Lung nodule classification using deep features in ct images. In Computer and Robot Vision (CRV), 2015 12th Conference on, pages 133–138. IEEE, 2015.
7. Kaggle. Data science bowl 2017, 2017.
8. Sandeep Kumar Saini, Gaurav, AmitaChoudhary, Detection Of Lung Carcinoma using fuzzy Logic and ACO Techniques, IJERT, Vol.3 Issue 8, pp.903-906,2014.
9. ZakariaSulimanZubi and Rema AsheibaniSaad," Using Some Data Mining Techniques for Early Diagnosis of Lung Cancer", Rescent researches in Artificial Intelligence, Knowledge Engineering and data Bases, Libya,2007.
10. F. Taher and R. Sammouda. Lung cancer detection by using artificial neural network and fuzzy clustering methods. In GCC Conference and Exhibition (GCC), 2011 IEEE, pages 295–298. IEEE, 2011.
11. V. Krishnaiah, Dr. G. Narsimha, Dr. N. Subhash Chandra, 2013, Diagnosis of Lung Cancer Prediction System Using Data Mining Classification Techniques, International Journal of Computer ScienceandInformayion Technologies, Vol.4(1),2013,39-45.
12. Yu M, Reiter JP, Zhu L, Liu B, Cronin KA, Feuer EJR. Protecting Confidentiality in Cancer Registry Data with Geographic Identifiers. Am J Epidemiol 2017 Jul 1;186(1):83-91.
13. L. Bottou. Stochastic gradient tricks. Neural Networks, Tricks of the Trade,Reloaded, 7700:430–445, 2012.
14. M. J. DCruz, M. A. Jadhav, M. A. Dighe, M. V. Chavan, and J. Chaudhari. Detection of lung cancer using backpropagation neural networks and genetic algorithm.

AUTHORS PROFILE

Ms. Hemalatha Eedi, working as Assistant Professor in the Department of Computer Science and Engineering, Jawaharlal Nehru Technological University Hyderabad College of Engineering Hyderabad since 2006. Her areas of interest include Parallel and Distributed Computing, Cloud Computing, IoT, Machine Intelligence and Deep Learning.