A Novel Approach to Lung Cancer Classification using Deep Learning Technique

Preeti Varshney1 :
CS Dept.
ABES Engineering College
Ghaziabad, India

Mudit Sharma3 :
CS Dept.
ABES Engineering College
Ghaziabad, India

Priyanshu2 :
CS Dept.
ABES Engineering College
Ghaziabad, India

Poonam Rana4 :
CS Dept.
ABES Engineering College
Ghaziabad, India

Abstract **—** In this research paper, there is a review of previous successful researches that had been done in the disease prediction & classification of deep learning. Nowadays, deep learning is all the rage, as healthcare centres and companies across industries seek to advanced computational techniques to find useful information across huge swaths of data. Deep learning use in many diseases' diagnosis like breast cancer, lung cancer, Prostate cancer, skin cancer, etc. with the help of methods to detect like X-Rays, CT scan,2-D CNNs,3-D CNNs, etc. Cancer is the leading cause of deaths worldwide. Both persons making observations and science, medical experts are facing the questions of fighting diseased growths. For saving the lives of many, early cancer detection is the priority. Typically, for these types of a cancer diagnosis visual examination and manual techniques are used. This hands-controlled sense given of medical images demands high time using up and is high with a tendency to errors.

Keywords**—** Cancer Prediction, Deep Learning Techniques, Convolutional Neural Networks, Stacked Autoencoders.

I. INTRODUCTION

Lung cancer is one of the killer diseases due to its aggressive nature and delayed detection at advanced stages. In harmony with the American diseased growths society, 96,480 deaths are looked to because of, about skin diseased growths, 142,670 from breathing part diseased growths, 42,260 from chest diseased growths, 31,620 from prostate 1 diseased growth, and 17,760 deaths from brain diseased growths in 2019. Early detection significantly improves the chances of survival and prognosis. Typically, for these types of cancer diagnosis visual examination and manual techniques are used. Deep learning has more chances of producing going straight to something from uncooked images the high-level point pictures of. In addition to deep learning, giving clear, full picture processing units (GPU) are also being used in parallel, for point extraction and image being seen. For example, CNN has been able to detect cancer with promising performance. Detection of lung nodules is time-consuming due to the volume of data involved, and also suffers from inter-radiologist variance. The common methods to detect cancer malignant nodules are chest radiographs (X-ray) and computer tomography scans. It is also possible that the existence of benign nodules leads to erroneous decisions. A CT scan can contain millions of pulmonary voxels, and lung nodules are relatively small. Deep learning will help in the reduction of misdiagnosis and false-positive results in early-stage lung cancer diagnosis.2-D Convolutional Neural Networks (CNNs) or a combination of 2-D CNNs using multiple views have been previously used in lung nodule detection.3-D CNNs which use the full 3-D nature of the input data vs 2D CNNs, specifically in lung nodule detection. This paper utilizes a 3-D convolution neural network (CNN) to detect the tumours found in the lungs as malignant and benign via deep / machine learning. We are using this method as it is the most effective as compared to the other methods. Mainly the CNN is image analysis we have used 3D customized mixed link network architectures for lung nodule detection and classification respectively. By using deep learning, we can improve the clinical result outcomes predictions which will be more accurate than any other method. It will be of low cost and minimal human input we can just put the command and AL will do all the work easily. Visualizing model decision making to increase radiologists' trust and improve adoption.

- (Hinton et al., 2006 describes the deep weighing network (DBN)) it is a multilayer generative model where each layer encodes statistical dependencies among the units in the layer unelevated it.
- It is trained to (approximately) maximize the possibility of its training data.
- DBNs have been powerfully used to learn high-level structure in a wide range of domains, including handwritten digits
- human motion capture data (Taylor et al., 2007). We build upon the DBN in this paper considering we are interested in learning a generative model of images that can be trained in a purely unsupervised manner.

- However, DBN was with a good outcome still scaling them to realistic-sized (e.g., 200x200 bit of picture) images being hard for two reasons.
- First, images are high-dimensional, so the algorithms must scale gracefully and be computationally tractable plane when unromantic to large images.
- Second, objects can towards at wrong-headed locations in images; thus, representations should be invariant at least to local translations of the input.
- We write these issues by incorporating translation invariance. Like LeCun et al. (1989) and Grosse et al. (2007), we learn full-length detectors that are shared among all locations in an image, considering features that capture useful information in one part of an image can pick up the same information elsewhere.

Furthermore, researches had taken place that improves the efficiency and our model is precisely based on the 3-D CNN the best method to detect Lung Cancer which uses the full 3-D nature of the input-data specifically in lung nodule detection

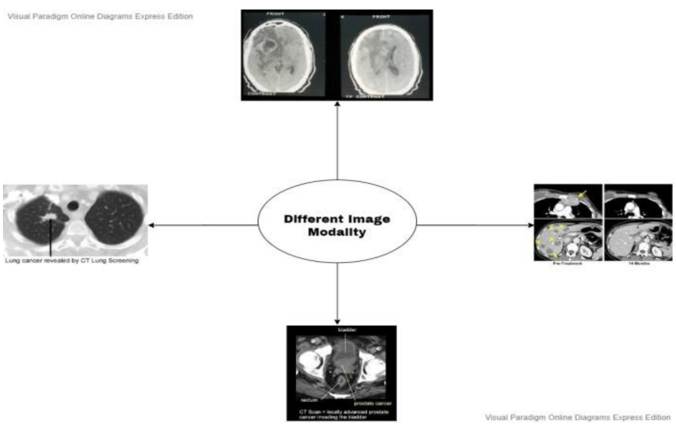


Fig.1: Different types of Images Modalities used for detecting various types of Cancer

II. Literature Review:

i) Lung cancer-

One of the world's most common causes of lung death is lung cancer. If the disease is a small, diffuse tumour, a combination of operating, percutaneous, and surgical therapies can support several patients. Unfortunately, a diagnosis of advanced clinical disease, nodal development, and/or metastatic disease occurs later in 75 percent of lung cancers, since there are few without any symptoms in the early stages of the disease. The overall survival rate of patients diagnosed with lung cancer is 15 percent [88], according to Australian studies. Several researchers have documented their work using the LIDC / IDRI Database in the field of lung nodule detection and classification. The database was composed of 240000 plus nodule images.

By removing discriminative features from alternately stacked layers, Multi-Convolution Neural Networks (MCNN) is used to capture nodular heterogeneity. Lung nodule screening and annotation are used for the assessment of the proposed LIDC-IDRI process. In the MCNN model, three CNNs are used in this process, in which parallel nodule patches of different sizes are assembled as inputs. Using the LIDC database, the precision of the segmentation process is 97 percent. A strategy for pulmonary nodule detection was suggested by Setio et al., using a multi-view Convolution network for model training. For candidate nodule detection, three fused algorithms, i.e., large-solid, sub-solid, and solid, accurately detect all suspected nodules. On the publicly accessible LIDC-IDRI dataset, the suggested framework is trained and validated. The sensitivity level achieved by the research work is 85.4% at 1 and 90.1% at 4 false-alarm. respectively. Dou et al. used a new technique to decrease false-positive in automated detection of the nodule by using 3-DCNNs from volumetric CT scans. This technique has been authenticated in the challenge of LUNA16, where they achieved the highest CPM score decreasing of the false alarm track.

ii) Breast cancer-

Breast cancer occurs in the cells of the breast and is the most common cancer in women after skin cancer in the world. Both men and women may suffer from breast cancer, but it is far more prevalent in women. Machine-assisted systems for the initial discovery of breast cancer are not only the latest advancement in medical imaging, but also improve radiologists' diagnostic skills. Mammography, tomography, Breast Ultrasound (BUS), MRI, CT scans and even deeper PET are the most popular instruments used for breast cancer diagnostics. The breast is usually counted as the human body's oversensitive organ, so only some of these procedures are recommended, which depends on the state of the patient and the status of the tumour. Mammography at an early stage of breast cancer is considered a low-cost and safe treatment, but it is unsuccessful in the dense breast of a young woman. In order to stop unnecessary biopsy, the BUS treatment is considered to be supportive of mammograms.

Several breast imaging datasets are publicly accessible, including DDSM, MIAS, WBCD, BCDR and NBIA, etc. Different pre-processing operations are carried out before segmentation following image acquisition, such as pectoral muscle removal and removal of artefacts, etc. The segmentation method is the most critical step of the machine-assisted system to increase accuracy and minimise the presence of abnormality by false positives. Several studies have suggested the GLCM approach for explaining texture-based characteristics.

Similarly, LBP is another impressive mechanism used to separate benign masses from malignant ones for texture extraction.

In the Convolutionary Neural Network architecture, the input layer, output layer and other hidden layers are fully connected layers known as transforming and pooling. A CNN-based approach for detecting breast carcinoma was suggested by Abdel-Zaher and Eldeib using an unmonitored pathway network of deep-faith beliefs followed by a route of backward propagation. The Wisconsin Breast Cancer Dataset (WBCD) was used for research and claimed 99.68 percent precision. Sun et al.] also suggested a breast cancer classification model. Using a deep convolutionary neural network (CNN), a graph-based semi-serviced learning (SSL) scheme was developed. For parameter training and fine-tuning, CNN usually uses large amounts of labelled data, and only a small portion of labelled data in the training set is required from the proposed arrangements. There were four modules in the diagnostic device: data assessment, assignment selection, data co-training and CNN.

Fig.2: Deep Learning Approach for Predicting Breast Cancer

iii)<u>Skin cancer</u>-

For the last few decades, machine-assisted systems using dermoscopic images have been introduced to assist the clinical decision of dermatologists and to recognise highly suspicious cases. Intelligent systems may also be used by novice clinicians as an additional method to obtain an initial assessment and improve the process of patient follow-up. These systems are loosely divided into two main groups concerning meaningful feature extraction from dermoscopic images, in which one class used diagnostic medical procedure and automatically extracted the same medical features i. In addition, another class is focused on machine learning to identify statistical patterns and to apply image attributes, i.e., texture and colour characteristics. In most of the work, the emphasis is on improving machine learning techniques with advanced extraction of features, such as the ABCD rule, a 3-point checklist. Therefore, DCNNs accomplished major outcomes in the field of medical imaging, through which features are generated directly from the images.

<u>iv)Prostate cancer</u>-

The second leading cause of male cancer deaths is prostate cancer. The rate of death caused by prostate cancer can be effectively decreased by early cancer detection. Due to high and multi-resolution prostate cancer MRIs, appropriate diagnostic systems and tools are needed. In the past, computer-assisted diagnosis (CAD) programs have been developed by researchers to help the radiologist diagnose anomalies. In this research paper, we used novel Machine Learning strategies for detecting prostate cancer, such as the Bayesian method, Support Vector Machine (SVM) kernels: polynomial, radial base function (RBF), and Gaussian and Decision Tree. In addition, various techniques for extracting features are suggested to boost the efficiency of detection. The techniques for evoking characteristics are based on the characteristics of texture, morphological, scale-invariant feature transformation (SIFT), and elliptic Fourier descriptors (EFDs). The output was evaluated using Machine Learning Classification techniques based on single as well as mixed features. In terms of receiver operating curve (ROC) and specificity, sensitivity, positive predictive value (PPV), negative predictive value (NPV), false-positive rate (FPR), output was assessed and Cross-validation (Jack-knife k-fold) was performed. SVM Gaussian Kernel has the highest accuracy of 98.34 percent with 0.999 AUC, based on single function extraction strategies. SVM Gaussian kernel with texture + morphological, and EFDs + morphological features offer the highest accuracy of 99.71 percent and AUC of 1.00 when using a combination of features extracting strategies.

III. PREVIOUS APPROACHES USED IN PREDICTION AND CLASSIFICATION OF CANCER DISEASE USING DEEP LEARNING TECHNIQUES:

There have been several approaches to learning deep networks:

i)Albert chon, Niranjan Balachandar, Peter Lu -- The altogether contributed in a research paper. They use 3D CT scans using segmentation, normalization, down sampling , and zero-centring. Our initial approach was to simply input the pre-processed 3D CT scans into 3D CNNs, but the results were poor, so we needed additional pre-processing to input only regions of interests into 3D CNNs.

Their result for vanilla CNN is dropout with 0.2 probability after each conv layer during training, Adam Optimizer with learning rate = 0.0003

Result for 3D Google net architecture is dropout with 0.3 probability after each conv and inception layer during training, Adam Optimizer with learning rate = 0.0001

ii) Wafaa Alakwaa ,Mohammad Nassef, Amr Badr -- They have used CAD systems for lung cancer which have the following pipeline: image pre-processing, detection of cancerous nodule candidates, nodule candidate false positive reduction, malignancy prediction for each nodule candidate, and malignancy prediction for overall CT scan. These pipelines have many phases, each of which is computationally expensive and requires well-labelled data during training.

The initial approach was to simply input the pre-processed 3D CT scans into 3D CNNs, but the results were poor. So, an additional pre-processing was performed to input only regions of interests into the 3D CNNs. Accuracy of model is 86.6%, Mis-classification rate is 13.4%, False positive rate is 11.9%, and False Negative is 14.7%. Almost all patients are classified correctly.

iii) Sarfaraz Hussein,Kunlin Cao,Qi Song UlasBagci -- They used the 3D CNN trained on Sports-1M dataset [15] which had 487 classes. We fine-tuned the network using samples from lung nodule dataset. In order to generate the binary labels for the six attributes and the malignancy, we used the centre point and gave positive (or negative) labels to samples having scores greater (or lesser) than the centre point.

Method Accuracy mean score difference

GIST feature with LASSO 76.83% 0.6753

3D CNN MTL with Trace norm 80.08% 0.6259

Table1: Comparison of Various works used in Predicting Lung Cancer on Various Datasets using Deep Learning Techniques:

S. No	Author Names	Dataset	Task	Performance Metrics
1.	Albert Chon NiranjanBhalachandar Peter Lu	Luna16 dataset	Accurately predict the patient label (cancer or non-cancer)	83%
2.	WafaaAlakwaa Mohammad Nassef Amr Badr	Luna16 dataset	Detection of the infected lung nodules through 3-D CNN	81%
3.	Sarfaraz Hussein Kunlin Cao Qi Song UlasBagci	Sports-1M dataset	Detection of the malignancy lung nodules through 3-D CNN	91.26%

IV. PROPOSED APPROACH FOR PREDICTION OF LUNG CANCER

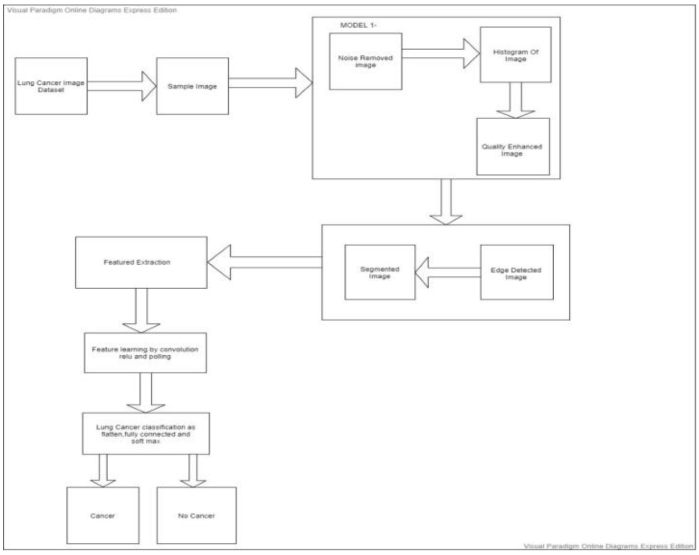
There are various steps involved in this proposed Methodology:

Step1: Data Pre-processing on Lung Cancer Dataset Images: Firstly, all the noises and irregularities from the images should be removed.

Step2: After enhancing the quality of images, we should perform Image Segmentation by applying various images Segmentation Technique:

Step3: After performing Image Segmentation on Lung Cancer Datasets images, we should perform Feature Extraction using Deep Learning Techniques either by Using Convolutional Neural Networks.

Step4: After feature extraction the classification step can be performed whether the person is benign or malignant.



CONCLUSION

In this paper various approaches used in prediction and Classification of various types of cancer like Lung Cancer, Breast Cancer, Skin Cancer, Pro-state Cancer have been described and one Methodology have been proposed for predicting Lung Cancer and Classification on Lung Cancer Datasets. In the future work this methodology will be implemented using Deep Learning Convolutional Neural Networks and the various Parameters like True positive, True Negative, False Positive, False Negative, Precision, Recall, F1-score, accuracy and Ruc Curve will be calculated.

REFERENCES

[1] Lee, H., Grosse, R., Ranganath, R. and Ng, A.Y., 2009, June. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In *Proceedings of the 26th annual international conference on machine learning* (pp. 609-616).

Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B. and Sánchez, C.I., 2017. A survey on deep learning in medical image analysis. *Medical image analysis*, *42*, pp.60-88.

Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y. and Alsaadi, F.E., 2017. A survey of deep neural network architectures and their applications. *Neurocomputing*, *234*, pp.11-26.

Fakoor, R., Ladhak, F., Nazi, A. and Huber, M., 2013, June. Using deep learning to enhance cancer diagnosis and classification. In *Proceedings of the international conference on machine learning* (Vol. 28). New York, USA: ACM.

Deng, L., 2014. A tutorial survey of architectures, algorithms, and applications for deep learning. *APSIPA Transactions on Signal and Information Processing*, *3*.

Tan, J., Ung, M., Cheng, C. and Greene, C.S., 2014. Unsupervised feature construction and knowledge extraction from genome-wide assays of breast cancer with denoising autoencoders. In *Pacific Symposium on Biocomputing Co-Chairs* (pp. 132-143).

Yuan, X., Xie, L. and Abouelenien, M., 2018. A regularized ensemble framework of deep learning for cancer detection from multi-class, imbalanced training data. *Pattern Recognition*, *77*, pp.160-172.

Hu, Z., Tang, J., Wang, Z., Zhang, K., Zhang, L. and Sun, Q., 2018. Deep learning for image-based cancer detection and diagnosis− a survey. *Pattern Recognition*, *83*, pp.134-149.

Xiao, Y., Wu, J., Lin, Z. and Zhao, X., 2018. A deep learning-based multi-model ensemble method for cancer prediction. *Computer methods and programs in biomedicine*, *153*, pp.1-9.

Burt, J.R., Torosdagli, N., Khosravan, N., RaviPrakash, H., Mortazi, A., Tissavirasingham, F., Hussein, S. and Bagci, U., 2018. Deep learning beyond cats and dogs: recent advances in diagnosing breast cancer with deep neural networks. *The British journal of radiology*, *91*(1089), p.20170545.

Chon, A., Balachandar, N. and Lu, P., 2017. Deep convolutional neural networks for lung cancer detection. *Standford University*.

Hussein, S., Cao, K., Song, Q. and Bagci, U., 2017, June. Risk stratification of lung nodules using 3D CNN-based multi-task learning. In *International conference on information processing in medical imaging* (pp. 249-260). Springer, Cham.

Alakwaa, W., Nassef, M. and Badr, A., 2017. Lung cancer detection and classification with 3D convolutional neural network (3D-CNN). *Lung Cancer*, *8*(8), p.409.