Performance of VGG-19 Convolutional Neural Network Model Based Lung Cancer Classification on Computed Tomography

Dr.Amjad Khan
Associate Professor Dept. of Artificial Intelligence & Machine Learning Bapuji Institute of Engineering & Technology, Davangere,Karnataka,India
Affiliated to Visvesvaraya Technological University Belagavi,Karnataka,India,

***Abstract*— The heterogeneity of the lung nodules and the complexity of the surrounding environment have made the robust nodule detection challenging task in identification of lung cancer. The survival rate of lung cancer depends upon early identification of lung nodules which is an efficient way to minimize the death rate of patients. The proposed method for Lung nodule classification from CT images developed using VGG-19 convolutional neural networks. This method eliminates the need of manual feature extraction as depicted in the feedback of previous work. The network is fed with raw lung CT images from publicly available LIDC-IDRI dataset. Here, the lung images are classified into two classes such as benign and malignant. This classification is achieved with the help of VGG-19 convolutional neural network. This method successfully classified the lung CT images into two classes and achieved 86% accuracy with comparatively less false positive rates.**

Keywords— Lung Cancer, VGG-19, Computed Tomography, Classification

I. INTRODUCTION

Identification of lung cancer is an efficient way to minimize the death rate of patients. It is a vital step to screen out the computed tomography (CT) images for pulmonary nodules for the diagnosis of lung cancer. Hence suitable mechanism should be adopted to detect and identify this disease in the initial stage to save the life of large number of peoples suffering from lung cancer. If it is detected and identified in primary stage then survival rate of many number of patients can be improved. Later after disease identification, by providing proper diagnosis can reduce the death rate of patients. So in order to avail a suitable and instantaneous outcome the importantly, applying recent techniques of machine leaning in the medical image processing field by enhancing the amount of duplication for the methods use can increase the accuracy of the classification. Therefore proper timely detection and identification in the prior stage will definitely improve the level of survival and can decrease the death rate.

The medical images taken in most of the earlier studies comprise of computed tomography (CT), magnetic resonance, and mammography images. The expert doctor of this domain uses these images for analysis to detect and identify the various levels of lung cancer by using suitable techniques. The different laboratory and clinical steps are being used including chemical treatment to destroy or stop the duplications of malignant cell, targeted therapy and also radiotherapy. All these procedures adopted to identify and detect the cancer diseases are lengthy, costlier and more painful for the patients. Thus, to overcome all these problems suitable machine learning techniques for processing these medical images were used which comprise of CT scan images. CT scan images are preferred compared to other images because CT images are less noisy as compared to MRI and X-Ray reports.

II. LITERATURE REVIEW

Disha Sharma & Gagandeep Jindal [2] has proposed an automated arrangement for identifying the lung tumor by using Computed Tomography (CT) images. The methods such as binary image slicing, Erosion, wiener filter were used to extract rigion of interest from CT image. It is shown that the system produces an sensitivity of 90% with a false positive of 0.05 per image. Besides the tumor size of diameter 3mm is determined to identify lung nodule in primary stage thereby patient's survival rate increases. Farzad Vasheghani Farahani *et al.* [3] introduced a system for the early identification of lung nodule using CT images. In this research first the lung is segmented using region growing and thresholding technique and then features were extracted CT scans such as circularity, eccentricity, compactness. In the classification phase features were used as a data for each classifier (KNN, SVM and MLP) and ensemble system, in which classifier makes its own decision and at the end majority voting is used for the combine decision of this system. Jin, Zhang and Jin [5] proposed a model which uses Convolutional neural network as classifier to identify the lung nodules. This model achieves an accuracy of 84.6%, sensitivity of 82.5% and specificity of 86.7%.

Hence quality of diagnosis increases from the large dataset. Ryota Shimizu *et.al,* [6] proposed a system to detect malignancy of lung based deep neural network. The learning model uses urine to detect different substances. The model provides an accuracy of 90% while detecting malignancy of lung but it does not determine the category or nature of cancer. Po-Whei Huang *et.al,* [7] presented a system which achieves an accuracy 0f 83.11% with the ROC area as 0.8437. Here the system classifies malignant tumor and benign tumor from given CT images using support vector classifier based on a number of fractal based features collected from fractional Brownian motion model. Vaishali C.Patil [8] has presented a lung tumor recognition using CT images. To detect disease malignancy, the computer aided design system was used. Image processing techniques were used to eliminate noise from CT images. After segmentation, a variety of classifiers such as Artificial Neural Networks and Support vector machine were used to determine different stages of lung cancer to enhance efficiency and to minimize error rate. Ailton Felix *et.al,* [9] proposed a 3D CAD System to extract texture features and 3D Margin Sharpness Features from LIDC dataset. This system classify small pulmonary nodules with diameters between 3-10mm. In this task they used three Machine Learning Algorithm: Random Forest, Multilayer Perceptron and K-Nearest Neighbor. Sri Widodo *et.al*, [11] proposed a Principal Component Analysis for classification of pulmonary nodules and artery automatically on chest CT scan image. In this study they consider 3 steps. The first step is lung organ segmentation using Active Appearance Model, the second step is segmentation of nodule using the morphological method and the third step is to classify pulmonary nodules and artery using PCA technique. Experimental result shows that the accuracy of the classification system is 90%. Ravindranath K [12] gives early detection of lung malignancy which includes detection of an uncertain tumor. The nodule is then classify according to the different stages of the disease. The detecting stage includes image pre-processing and segmentation which increase the accuracy by using statistical classifiers, SVM and fuzzy logic classifier. Difference in the level of intensity detects the normal and abnormal tumor at the early stage. The detected tumor is then classified using neural network classifiers, which distinguish the normal and abnormal lung malignancy. Rui Xu *et.al*, [13] introduced a deep Convolutional neural based system for the lung segmentation in CT scans for both mild and severe diseases. It's not easy task for radiologist to detect lung diseases which have complex opacities in ROI. So Deep-CNN model is introduced for lung segmentation. In this complex opacities is considered as a texture based problem in which pixel is classified as one inside or outside in ROI. This problem is solved by using CNN based model. This system takes 42 computed tomography images with severe lung disease and 7 Computed Tomography images with a mild lung cancer which includes six kinds of opacities. The jacquard index of this model is more than the mostly used lung segmentation method. In this research various classifiers are used such as Naive Bayes classifier, decision trees, SVM, k-nearest neighbors, logistic regression. In the lung sound data decision tree classifier and SVM classifiers gives the great accuracy of decisions and results. It is assumed that with the increase in the input data, accuracy also increases. Pratiksha Hattikatti [15] proposed Convolutional Neural Network for finding the range of the lung texture pattern of diseases from computed tomography images. The term called interstitial lung disease includes various disease related to lung, characterization of lung tissue is an important element of CAD system for determining the interstitial lung diseases. The Convolutional neural network achieves accuracy of 94% with high sensitivity and same data is passed through SVM classifier which gives the accuracy of 86% only. It concludes that CNN gives more accurate results for interstitial lung diseases. Pouria Moradi [17] proposed 3D Convolutional Neural Network, which can reduce the false positive rate and provide high sensitivity in detecting lung cancer. The main objective of this research is to improve classification accuracy. Researcher achieved 91.23% accuracy for 3.99 false positive per scan using the method for combining different classifiers. Sakif Rahman *et.al,* [18] proposed a lung cancer identification and prognosis method using deep neural network (DNN) which reduce the time complexity and increase accuracy.

III. METHODOLOGY USED

A. VGG-19 Network Model

The proposed lung cancer Identification system using VGG-19 Model shown in the figure-1 is mainly divided into two parts. In the first part, we are doing preprocessing before feeding the images and then were identifying the nodule that is used to train to ultimately classify the CT input images as malignant or benign for lung cancer to achieve the result. Figure 1.1 presents the algorithm of VGG-19 based lung cancer classification.

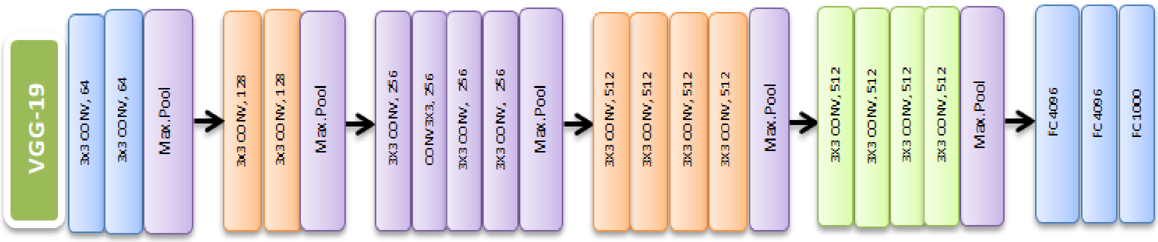

Figure 1: VGG-19 Network Model

Algorithm

1. **Acquire the images of lung cancer containing both diseased and non-diseased images and also, using existing dataset and augmentation technique.**
2. **Preprocess all the images for resize the images based on the algorithm /technique used**
3. **Assign the class labels to the images that are benign and malignant.**
4. **Categorize the images among training and testing dataset selecting from all the class labels.**
5. **Train the VGG19 network model with the help of training images**
6. **Test the VGG19 network model with the help of testing images**
7. **Calculate the various performance measure parameters**
8. **Validate the performance of the proposed model and compare the results with the other state-of-the-art approaches.**

Figure.1.1 Algorithm VGG-19 Based Lung Cancer Classification

The proposed VGG19 architectural block diagram shown in figure.1 mainly consists of the following layers: sixteen convolutional layers which follows, five max- pooling layers and one fully-connected layer. As shown in Figure2, the network begins with a convolution layer, in which the first convolution layer takes the image with input size of 224 × 224 pixels. The second convolution layer consists of 32 feature maps with the convolution kernel of 3 × 3. The kernel size for max pooling layers is 2 × 2 and the stride of 2 pixels, and the fully-connected layer generates an output of 1024 dimensions. After applying these architectures, some images detected with cancerous nodules and some identified as non-cancerous.

B. Training VGG-19 Network Model

Back-propagation algorithms are used to train the VGG-19 using CT images of size 224x224x3. It consists of two phase, training phase and testing phase. In the first phase trained using CT images are being used to train the network for the classification of lung as either cancerous or non-cancerous. In the testing phase an image unknown to the network is applied as input to classify as cancerous or noncancerous. For minimum loss of features images are trained and tested in the DICOM format itself by modifying the network parameters. The proposed designed network accuracy can be achieved by suitable evaluation.

VI. RESULTS AND DISCUSSIONS

The dataset used in the research work belongs to LIDC-IDRI which is the Lung Image Database Consortium image collection (LIDC-IDRI) consists of diagnostic and lung cancer screening thoracic computed tomography (CT) scans with marked-up annotated lesions. In the initial blinded-read phase, each radiologist independently reviewed each CT scan and marked lesions belonging to one of three categories ("nodule > or =3 mm," "nodule <3 mm," and "non-nodule > or =3 mm"). The inputs are the image files that are in DICOM format, it is important to note that in order to preserve the original values of the DICOM images as much as possible; no scaling was applied to the CT images of the dataset. The result of classification depicted in figure-3. The large dimension images cannot be fed directly into Convolutional neural network architecture because of the limit on the computation power. Hence it is preprocessed to reduce the size of the input data and thus segmenting the images into equal size and format. Also dataset used was of size 15 GB but a subset of the images around 2.5GB were used for training and testing usage.[5][6].

Results of VGG-19

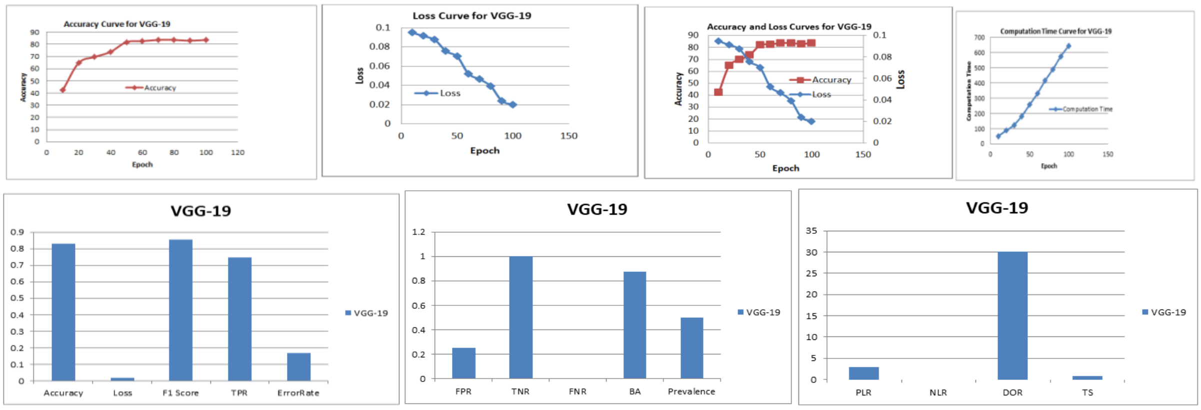

Figure.3 Results of Classification

In our research work, Lung nodule classification has been implemented in MATLAB 2018b and the dataset used for training and testing purposes are taken from LIDC-IDRI to get familiarize with lung cancer. Here the images samples are used to feed the network model which is able to detect and identify the presence of cancer that is cancerous images and Non-Cancerous Images. As it is observed from the results that as training proceeds further classification accuracy will be increases with increase in the computation time, thereby decreases the percentage of loss as shown in above output graphs and accuracy obtained were compared with the work of earlier research papers [19][20].

V. CONCLUSION AND FUTURE WORK

In this research work, VGG-19 Convolutional neural networks for classifying the CT images of lung into cancerous and non-cancerous were used. Thus preprocessing has been done before applying input CT images to network model to make equal sizes and format of the images. The dataset used in our research work belongs to LIDC dataset. Hence an accuracy of 86% is achieved with comparatively less false positive rates. VGG-19's strength lies in its ability to identify diverse and wide ranging objects in images but it stumbles when faced with less diversity and a more subtle classification problem.

REFERENCES

[1] Makde et al., Deep Neural Network Based Classification of Tumorous and Non-tumorous Medical Images, Information and Communication Technology for Intelligent Systems (ICTIS 2017) - Volume 2, Smart Innovation, Systems and Technologies 84, DOI 10.1007/978-3-319-63645-0 22.

[2] Hiroyuki Sugimori, Classification of Computed Tomography Images in Different Slice Positions Using Deep Learning, Hindawi Journal of Healthcare Engineering Volume 2018, Article ID 1753480, 9 pages https://doi.org/10.1155/2018/1753480,pages1-9.

[3] Joel Than Chia Ming1, Norliza Mohd Noor2, Omar Mohd Rijal3 , Rosminah M. Kassim4 , Ashari Yunus5., Lung Disease Classification using Different Deep Learning Architectures and Principal Component Analysis, 2018 2nd International Conference on BioSignal

Analysis, Processing and Systems (ICBAPS), 978-1-5386-1278-1/18/187-190 ©2018 IEEE.

[4] J. T. C. Ming, O. M. Rijal, R. M. Kassim, A. Yunus, and N. M. Noor, "Texture-based classification for reticular pattern and ground glass opacity in high resolution computed tomography Thorax images," in Biomedical Engineering and Sciences (IECBES), 2016 IEEE EMBS Conference on, 2016, pp. 230–234.

[5] M. Awais, H. Muller, and F. Meriaudeau, "Classification of SDOCT images using Deep learning approach," in 2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), 2017, pp. 3–6.

[6] R. Hooda, S. Sofat, S. Kaur, A. Mittal, and F. Meriaudeau, "Deep learning: A potential method for tuberculosis detection using chest radiography," in 2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), 2017, pp. 497–502.

[7] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv Prepr. arXiv1409.1556, 2014.

[8] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu et al., "Convolutional neural networks for medical image analysis: full training or fine tuning," IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1299–1312, 2016.

[9] Yuechao Zhang, Jianxin Zhang, Lasheng Zhao1, Xiaopeng Wei1, Qiang Zhang, Classification of Benign and Malignant Pulmonary Nodules Based on Deep Learning, 2018 5th International Conference on Information Science and Control Engineering,978-1-5386-5500-9/18/156- 60, ©2018 IEEE, DOI 10.1109/ICISCE.2018.00042.

[10] S.G. Armato, G. McLennan, L. Bidaut, M.F. McNitt-Gray, C.R. Meyer, A.P. Reeves, et al., The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): a completed reference database of lung nodules on CT scans, Medical Physics 38 (2) (2011) 915–931, https://doi.org/10.1118/1.3528204.

[11] J.-P. Charbonnier, E.M. van Rikxoort, A.A.A. Setio, C.M. Schaefer-Prokop, B. van Ginneken, F. Ciompi, Improving airway segmentation in computed tomography using leak detection with convolutional networks, Medical Image Analysis 36 (2017) 52–60, https://doi.org/10.1016/j.media.2016.

[12] Ravindranath K , K Somashekar, "Early Detection of lung cancer by nodule extraction – A Survey", International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT), 2017.

[13] Rui Xu, Jiao Pan, Xinchen Ye, Yasushi Hirano, Shoji Kido, Satoshi Tanaka "A Pilot Study to Utilize a Deep Convolutional Network to Segment Lungs with Complex Opacities" Chinese Automation Congress (CAC), 2017

[14] Wei Chen, Haifeng Wei , Jiawei Sun ,Xu Qiao , Boqiang Liu, "Hybrid Segmentation Network for Small Cell Lung Cancer Segmentation" IEEE Access, vol. 7, pp. 75591 - 75603, June 2019.

[15] Moradi, P. & Jamzad, M., "Detecting Lung Cancer Lesions in CT Images using 3D Convolutional Neural Networks". 4th International Conference on Pattern Recognition and Image Analysis, 2019.

[16] Md. Sakif Rahman, Pintu Chandra Shill, Zarin Homayra, "A New Method for Lung Nodule Detection Using Deep Neural Networks for CT Images". International Conference on Electrical, Computer and Communication Engineering (ECCE), 2019.

[17] Pathan, A. and Saptalkar, B.K. (2012) Detection and Classification of Lung Cancer Using Artificial Neural Network. International Journal on Advanced Computer Engineering and Communication Technology , 1, 2278-5140.

[18] Dandil, E., Cakiroglu, M., Eksi, Z., Özkan, M., Kurt, Ö.K. and Canan, A. (2014) Artificial Neural Network-Based Classification System for Lung Nodules on Computed Tomography Scans. 6th International Conference of Soft Computing and Pattern Recognition, Tunis, 11-14 August 2014, 382-386.

https://doi.org/10.1109/SOCPAR.2014.7008037

[19] Qaisar Abbas, "Nodular-Deep: Classification of Pulmonary Nodules using Deep Neural Network," International Journal of Medical Research & Health Sciences", vol. 6, pp. 111-118, 2017.

[20] Worawate Ausawalaithong, Arjaree Thirach, Sanparith Marukatat, Theerawit Wilaiprasitporn, "Automatic Lung Cancer Prediction from Chest X-ray Images Using Deep Learning Approach," arXiv:1808.10858v1 [eessIV], August 2018.

[21] Brett A. Simon, Gary E. Christensen, […], and Joseph M. Reinhardt. "Computed Tomography Studies of Lung Mechanics", American Thoracic Society, 2005.

[22] S. M. Salaken, A. Khosravi, A. Khatami, S. Nahavandi, and M.A. Hosen, "Lung cancer classification using deep learned features on low population dataset," In Electrical and Computer Engineering (CCECE), IEEE 30th Canadian Conference, pp.1−5 , IEEE, 2017.

[23] Siddharth Bhatia, Yash Sinha and Lavika Goel, "Lung Cancer Detection: A Deep Learning Approach", Soft Computing for Problem Solving, Advances in Intelligent Systems, 2019.

[24] Yutong Xie ,Yong Xia , Jianpeng Zhang ,Yang Song , Dagan Feng , Michael Fulham , Weidong Cai, "Knowledge-based Collaborative Deep Learning for Benign-Malignant Lung Nodule Classification on Chest CT" IEEE Trans. Med. Imag., vol. 38, no. 4, pp. 991 - 1004, April. 2019.

[25] M. Gomathi , Dr. P. Thangaraj "Automated CAD for Lung NoduleDetection using CT Scans" 2010 International Conference on Data Storage and Data Engineering.

[26] Song QZ, Zhao L, Luo XK, Dou XC. Using deep learning for classification of lung nodules on computed tomography images. J Healthc Eng 2017;2017. http://dx.doi.org/10.1155/2017/8314740