2023 International Conference on Artificial Intelligence and Smart Communication (AISC) | 979-8-3503-2230-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/AISC56616.2023.10085058

Prediction of Lung Cancer using Convolution Neural Networks

Aanchal Vij
School of Computing Science and Engineering Galgotias University
Greater Noida, India
aanchal.vij04@gmail.com

Kuldeep Singh Kaswan
School of Computing Science and Engineering Galgotias University
Greater Noida, India
kaswankuldeep@gmail.com

***Abstract*—Lung cancer is one of the deadly diseases whose prediction is required to reduce the death rate. So, Artificial intelligence is used on CT scan images are used for achieving better accuracy in an automated manner. Deep Learning is one of the emerging trends for predicting values. Convolution Neural Networks is one of the deep learning algorithms which implemented to sample produces better outcomes as compared toother machine learning algorithms. In this paper, the data-set hasbeen taken with 1000 images of chest scans for different types of lung cancers such as Adenocarcinoma, Large Cell Carcinoma andSquamous Cell Carcinoma. Multiple machine learning algorithmshas been compared and then it has been confirmed that CNN is one of the best among all to check the accuracy of the prediction. The paper includes VGG - 16 implemented on data set having different types of Lung Cancer and thus helping to check the severity and precautions for the same in a distinct manner.**

***Index Terms*—Neural Networks, Deep Learning, Lung Cancer,VVG -16, Non-Small lung Cancer**

I. INTRODUCTION

Lung Cancer has a severe effect on generation with the growing death rates. [1] It has second largest death among all the cancer types.As per the American Cancer Society, it has been surveyed that Lung cancer needs to be diagnosedat early stage so as to achieve the recovery else recovery is suspected while prediction in late stages. [2] There are multipleways to detect Lung Cancer with medical Images such as CT Scans, MRI, X- rays among which CT Scan is consideredone of most efficient one among these. But instead of manual checkup, the automatic detection can help in providing better results. Convolution Neural Networks is considered one of thebest methods to work on images for prediction. So, the paper includes multiple CT Scan images of multiple cancer types such as Lung adenocarcinoma, large cell undifferentiated and squamous cell carcinoma. Every type has its own attributes anddreadful outcomes which has been provided below.

A. Lung Adenocarcinoma

Lung adenocarcinoma is the most common type of lung cancer [3] [4], accounting for 30% of all cases overall and approximately 40% of all non-small cell lung cancer occurrences. Adenocarcinoma can be found in a variety of cancers, including breast, prostate, and colorectal cancer. Adenocarcinomas of the lung are found in the glands that secrete mucus and help us breathe. Coughing, hoarseness, weight loss, and weakness are all symptoms [5].

B. Large-Cell Undifferentiated Carcinoma

Large-cell undifferentiated carcinoma lung cancer expands and spread very quickly and it can be observed anywherewithin the lung. This type of lung cancer typically accounts for10 to 15% of all non-small lung cancer cases [6]. Large-cell undifferentiated carcinoma grows and spreads rapidly [7].

C. Squamous Cell Carcinoma

Squamous Cell lung cancer occurs internally in the lung, in which the relatively large bronchi connect the trachea to the lung, or in one of the main airway branches [8]. Squamouscell lung cancer accounts for approximately 30% of all non-small cell lung cancers [17] and is commonly associated with smoking. There are multiple ways to detect Lung Cancer with medical Images such as CT Scans, MRI, X- rays among which CT Scan is considered one of most efficient one amongthese. But instead of manual checkup, the automatic detection can help in providing better results. Convolution Neural Networks is considered one of the best methods to work on images for prediction among all the existing methods [9].

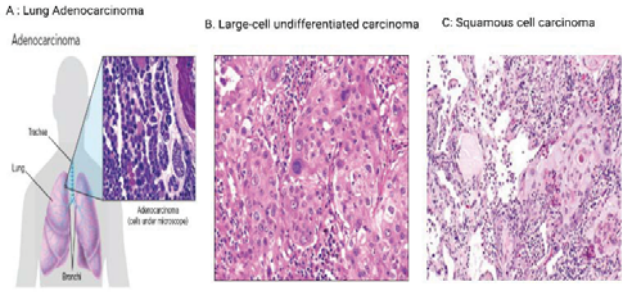

Fig. 1. Pathology images of different type of Lung Cancer [10]

Figure 1 shows the various pathology images of different types of lung cancer described above it. This paper has been drafted after analysing multiple methods to detect lung cancer and thus claims CNN (VGG 16) is one of the best methods achieving 77.62% accuracy.

II. RELATED WORK

Multiple researchers surveyed and resulted to some outcomes such as C. Yao et *al.* [11] designed a CNN model taking self-made data set achieved 90% accuracy but the validity of data and accuracy on real time data sets remained doubtful.So, the method needs to be tested at several other data sets to check its reliability.

In the same manner M. Norouzi [1] claimed that nanotech- nology is one of the efficient methods to be applied for therapiesfor lung cancer rather than complicated chemo-therapies andit can also reduce the amount of toxicity. The survey hasproved nanotechnology a great method to be considered with conditions applied because patient selection and combinations of multiple treatment provides an advanced enhancement.

Multiple papers have been surveyed in J. Wang et *al.* [9]and verified that false positive rates of those are quite high which decrease the accuracy, so to overcome the problem, 3 dimensional - convolutional neural network, VGG 16, Alex Net and Multi Crop Net L.ye et *al.* [12] advanced from 2-dimensional architecture which extracted 8.28% which is relatively low when compared with other algorithms.

A data set of 1000+ images has been taken in J. Wang et al. [13] with stage T1a-3N0M0, NSLC where it has been claimed that non-small lung cancer is more dangerous than small lung cancer, the death rate compared, the former is considered more harmful.

Survival rate has been checked by A. Agaimy et *al.* [6] after the first diagnosis of non-small lung cancer to check the criticality of it. The data contains both the genders which includes 8 males and 6 females under the age group of 52to 85 years (median 60). The maximum survival rate afterthe diagnosis was of one patient who was having cerebral metastasis, alive even after 45 months.

VGG 16 implementation is also done in paper S. T. M. Sheriff et *al.* [14] where the result only reveals the presence of lung cancer or not with not so good accuracy, so the inspiration has been taken from the paper to implement notjust on cancerous images but also the types of lung cancer so as to check that how critical the consequences would be and what precautions need to take care.

Multiple CNN models has been checked in M. Phankokkruad et *al.* [15] such as VGG16, ResNet50V2, and DenseNet201 which are based on transfer learning. Every model predicted its accuracy provided as 62%, 90%, and 89%, respectively.

A faster R - CNN model is proposed (Faster Reception - convolutional neural network) in Z.Xiao et *al.* [16] producing 92.8% as per the research conducted claiming it one of the bestmodels to predict lung cancer nodules using CT scan images but again this models only predicts the presence of lung cancer not the specific lung cancer type which provides a descriptive knowledge.

Real data has been taken from KI Hospital, Iraq having multiple images and is being tested on Matlab GUI by D. N. Anwer and S. Ozbay, [17] and checked that either the 7-layer model can work on real data or not which proved accurate as model is working perfectly. Comparison with similar work has been also done with the confusion matrix provided.

III. Methodology Used

A data set has been created with CT - Scan images of Chest which has been classified into categories with respect to images. Since Lung cancer has multiple types, so various images are taken of Adenocarcinoma, Large Cell Carcinoma, Squamous Cell Carcinoma and 1 folder of normal CT-Scan images (normal). These lung cancer types have its own severity and consequences, so predicting the correct lung cancer type instead of broad lung cancer prediction so as to have distinctive confirmation so as to take appropriate preventive methods for the same. Multiple methods have been implemented on the dataset and the best among them is figured out. The best method which has been surveyed after all the methods analyzed is VGG 16 (Visual Geometry Group -16) [13]

A. Introduction to VGG 16

VGG represents for Visual Geometry Group, and it is a multilayer deep Convolutional Neural Network (CNN) architecture. The term “deep” relates to the number of layers in VGG-16 or VGG-19, which have 16 or 19 convolutional layersrespectively.

The VGG architecture serves as the foundation for cutting- edge object recognition models. The VGGNet, which wascreated as a deep neural network, outperforms baselines ona variety of tasks and data sets in addition to ImageNet. Furthermore, it is still one of the most widely used image recognition architectures today. [18] VGG 16 is top 5 models accuracy models which achieves about 92.7 % accuracy. Figure 2 describes the diagrammatic representation of VGG model having 16 deep convolutional layers.

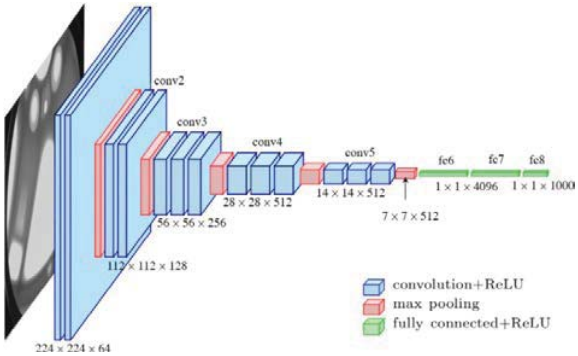

Fig. 2. Figure 2: Diagrammatic Representation of VGG 16 [19]

The figure 2 represents the view of Visual geometry group with multiple deep layers in it. Since the model is having 16 layers, so it having sixteen multiple layers in it.

After multiple researches and comparisons done, it has beendetected that VGG 16 is one of the efficient methods to be worked on image data set for image classification so a model has been designed to provide better results using Convolution Neural Networks (VGG - 16). The model provides 77.62% accuracy after loading the best weight. It has a sequence of steps and a flow which needs to be followed, the flow chart describing the flow of model created has been described in figure 3. Once the model is finalized and data set is compiled over it, it is being fed to VGG 16 where images are being undergone 16 deep layers so as so to check maximum classification factors and to provide maximum accuracy with less false rate. The below given figure 3 shows the flowchart of the model and figure 4 describes the flow of the model with respectto VGG- 16. As the dataset is having 4 types of images, thus all are being fed and dividing them in 3 parts namely Train, test and validate with 70%, 20% and 10% respectively which then lead to the generation of results with multiple factors calculated such as accuracy, precision, recall, auc, f1 score and loss.

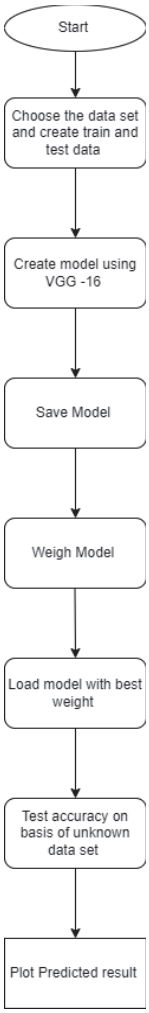

Fig. 3. Figure 3: The flow of Model Completion [20] There are different design aspects of the model which has been provided in figure 4 provided below:

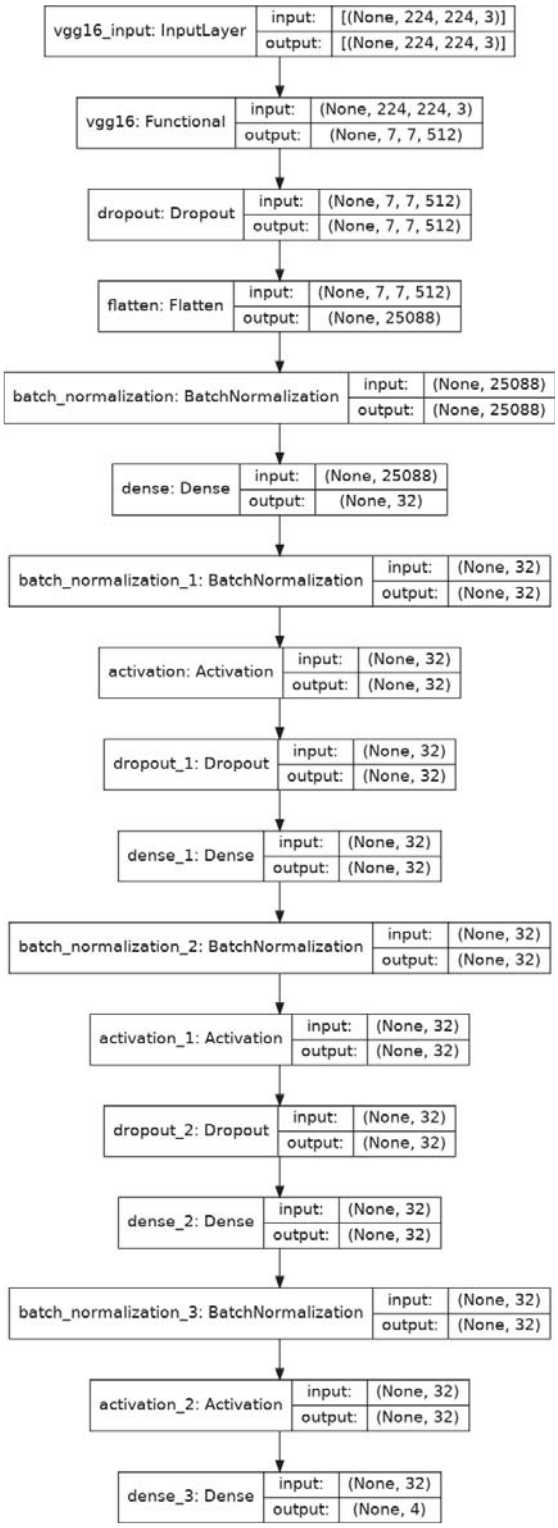

Fig. 4. Figure 4: VGG Model Flow

B. Data-Set Used

The data- set used consists of 1000 images of CT scan. There are 988 images in .png format and 12 images in .jpeg format. The distribution among Train, Test and Validate is 70%, 20% and 10% respectively. The image count is 613, 315 and 72. The data is thus trained and tested, resulted in multiple outcomes. The images contain multiple cancer typesto not to just check that either the image is cancerous or notbut also the type of lung cancer so as the specific result predict the seriousness of the cancer and thus check the preventive measures after that.

IV. RESULTS AND FACTS OBTAINED

The data-set when applied VGG-16 algorithm to check the accuracy and other metrics, resulted in the provided outcomes: The best result predicted with the model is provided in table 2

TABLE 2: RESULTS GENERATED

Parameter	Prediction %ge
Accuracy	77.62%
Precision	60.38%
Recall	30.48%
AUC	75.08%
F1 Score	40.67%
Loss	1.05%

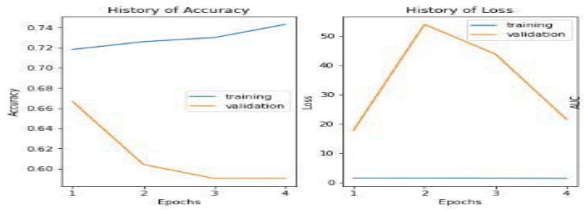

Fig. 5. Graph of Accuracy and loss with Epochs and Metrics with the Model

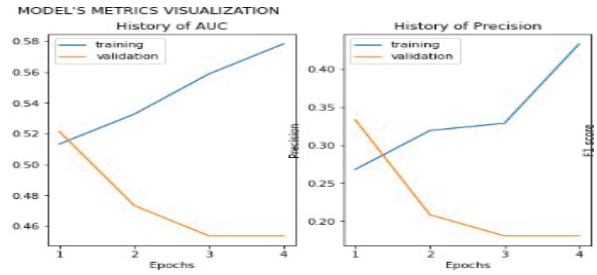

Fig. 6. Graph of AUC and Precision with Epochs and Metrics with the Model

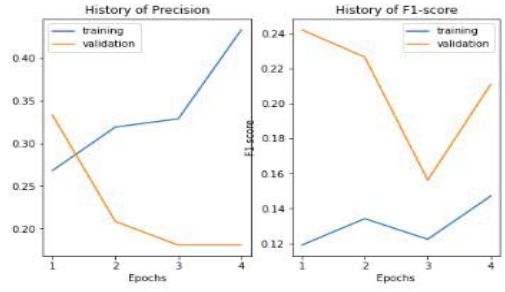

Fig. 7. Graph of Precision and F1 score with Epochs and Metrics with the Model

The figure 5 predicts accuracy and loss, figure 6 predicts AUC and precision and figure 7 predicts precision and F1 score. The uphill and downfall has been shown in the figures provided.

V. FUTURE SCOPE

The model produces distinct result for the types of non-smalllung cancer which provides a better picture to check the threat and chose the diagnosis, but the model achieves only 77.62% which is not much effective and can be enhanced using multiple other algorithms so as to have efficient algorithm.

REFERENCES

[1] M. Norouzi and P. Hardy, "Clinical applications of nanomedicines in lung cancer treatment," *Acta Biomaterialia*, vol. 121, pp. 134–142, 2021.

[2] P. H. Viale, "The american cancer society's facts & figures: 2020 edition," *Journal of the Advanced Practitioner in Oncology*, vol. 11,no. 2, p. 135, 2020.

[3] D. G. Beer, S. L. Kardia, C.-C. Huang, T. J. Giordano, A. M. Levin,
D. E. Misek, L. Lin, G. Chen, T. G. Gharib, D. G. Thomas, *et al.*, "Gene-expression profiles predict survival of patients with lung adeno- carcinoma," *Nature medicine*, vol. 8, no. 8, pp. 816–824, 2002.

[4] C.-R. Guo, Y. Mao, F. Jiang, C.-X. Juan, G.-P. Zhou, and N. Li, "Com- putational detection of a genome instability-derived lncrna signaturefor predicting the clinical outcome of lung adenocarcinoma," *Cancer Medicine*, vol. 11, no. 3, pp. 864–879, 2022.

[5] M. A. Gillette, S. Satpathy, S. Cao, S. M. Dhanasekaran, S. V. Vasaikar,
K. Krug, F. Petralia, Y. Li, W.-W. Liang, B. Reva, *et al.*, "Proteogenomic characterization reveals therapeutic vulnerabilities in lung adenocarci- noma," *Cell*, vol. 182, no. 1, pp. 200–225, 2020.

[6] A. Agaimy, O. Daum, M. Michal, M. W. Schmidt, R. Stoehr, A. Hartmann, and G. Y. Lauwers, "Undifferentiated large cell/rhabdoid carcinoma presenting in the intestines of patients with concurrent or recent non-small cell lung cancer (nsclc): clinicopathologic and molecular analysis of 14 cases indicates an unusual pattern of dedifferentiated metastases," *Virchows Archiv*, pp. 1–11, 2021.

[7] K. I. Tosios, V. Papanikolaou, D. Vlachodimitropoulos, and N. Goutas, "Primary large cell neuroendocrine carcinoma of the parotid gland. report of a rare case," *Head and Neck Pathology*, pp. 1–8, 2021.

[8] B.-Y. Wang, J.-Y. Huang, H.-C. Chen, C.-H. Lin, S.-H. Lin, W.-H. Hung, and Y.-F. Cheng, "The comparison between adenocarcinoma and squamous cell carcinoma in lung cancer patients," *Journal of cancer research and clinical oncology*, vol. 146, no. 1, pp. 43–52, 2020.

[9] S. Li, P. Xu, B. Li, L. Chen, Z. Zhou, H. Hao, Y. Duan, M. Folkert, J. Ma,
S. Huang, *et al.*, "Predicting lung nodule malignancies by combining deep convolutional neural network and handcrafted features," *Physics inMedicine & Biology*, vol. 64, no. 17, p. 175012, 2019.

[10] images from biorender, "www.biorender.com," 2022.

[11] C. Yao, Y. Qu, B. Jin, L. Guo, C. Li, W. Cui, and L. Feng, "A convolutional neural network model for online medical guidance," *IEEEAccess*, vol. 4, pp. 4094–4103, 2016.

[12] L.-Y. Ye, X.-Y. Miao, W.-S. Cai, and W.-J. Xu, "Medical image diagnosis of prostate tumor based on psp-net+ vgg16 deep learning network," *Computer Methods and Programs in Biomedicine*, vol. 221, p. 106770, 2022.

[13] J. Wang, B. Wang, J. Bi, and K. Li, "Prognostic significance of microvascular invasion and microlymphatic permeation in non-small-cell lung cancer," *European Journal of Cardio-Thoracic Surgery*, vol. 43,no. 6, pp. 1269–1269, 2013.

[14] S. T. M. Sheriff, J. V. Kumar, S. Vigneshwaran, A. Jones, and J. Anand, "Lung cancer detection using vgg net 16 architecture," in *Journal of Physics: Conference Series*, vol. 2040, p. 012001, IOP Publishing, 2021.

[15] M. Phankokkruad, "Ensemble transfer learning for lung cancer detection," in *2021 4th International Conference on Data Science and Information Technology*, pp. 438–442, 2021.

[16] Z. Xiao, B. Liu, L. Geng, J. Wu, and Y. Liu, "Detection of pulmonary nodules based on reception and faster r-cnn," in *Proceedings of the*

2019 8th International Conference on Computing and Pattern Recognition, pp. 160–166, 2019.

[17] D. N. Anwer and S. Ozbay, "Lung cancer classification and detection using convolutional neural networks," in *Proceedings of the 6th International Conference on Engineering & MIS 2020*, pp. 1–8, 2020.

[18] A. Srinivasulu, K. Ramanjaneyulu, R. Neelaveni, S. R. Karanam, S. Ma-jji, M. Jothilingam, and T. R. Patnala, "Advanced lung cancer predictionbased on blockchain material using extended cnn," *Applied Nanoscience*,pp. 1–13, 2021.

[19] Y. Lu, H. Liang, S. Shi, and X. Fu, "Lung cancer detection using a dilated cnn with vgg16," in *2021 4th International Conference on SignalProcessing and Machine Learning*, pp. 45–51, 2021.

[20] M. Arju and A. Rahman, "Convolutional neural network-based image classifier for breast cancer histopathology images," 2019.

[21] J. H. Lee, Y. C. Yoon, H. S. Kim, M. J. Cha, J.-H. Kim, K. Kim, and H. S. Kim, "Obesity is associated with improved postoperative overall survival, independent of skeletal muscle mass in lung adenocarcinoma,"*Journal of Cachexia, Sarcopenia and Muscle*, vol. 13, no. 2, pp. 1076– 1086, 2022.

[22] B. Gundogdu, D. Gurel, and E. C. Ulukus, "Biomarkers in pulmonary carcinomas," in *Biomarkers in Carcinoma of Unknown Primary*, pp. 99–128, Springer, 2022.

[23] N. N. S. M. Marzuki, I. S. Isa, N. K. A. Karim, I. L. Shuaib, Z. H. C. Soh, and S. N. Sulaiman, "Demarcation of lung lobes in ct scan images for lung cancer detection using watershed segmentation," in *Proceedingsof the 2020 12th International Conference on Computer and AutomationEngineering*, pp. 70–74, 2020.