Grenze International Journal of Engineering and Technology, June Issue

Lung Cancer Detection using CNN

Shobhar[1], Aswin R B[2], Rohan Gupta[3], Varun N Sharma[4] and Nalini N[5]
[1-5]Department of Computer Science and engineering, Nitte Meenakshi Institute of Technology- Affiliated to Visvesvaraya Technological University, Bangalore, India
Email: shobha.p@nmit.ac.in

***Abstract*— Lung cancer remains a sizable global health concern, challenging beforehand detection for improved treatment outcomes. This paper explores the transformative potential of Deep Learning techniques, exclusively convolutional neural networks (CNNs), in the accurate identification and classification of lung cancer using computed tomography (CT) scans. A comprehensive literature survey outlines diverse methodologies, including CNNs, support vector machines (SVMs), and feature extraction, employed in the pursuit of precise lung cancer diagnosis. Our proposed methodology involves preprocessing CT scan images, feature extraction, and the application of a CNN model trained on a dataset of lung CT scans. The CNN model aims to categorize images into distinct types of lung cancer, proposing a promising avenue for enhanced diagnostic accuracy. Results and analysis showcase the proficiency of our CNN model, supported by key performance indicators. As we delve into the future of cancer diagnostics, the integration of advanced machine learning with medical imaging emerges as a beacon of hope for quick and personalized lung cancer detection, surfacing the way for improved patient outcomes.**

***Index Terms*— Accuracy, Cancer Classification, CT scan. Convolutional Neural Networks (CNNs), Deep Learning (DL), Diagnostic Accuracy, Early Detection, Machine Learning (ML), Medical Imaging. Lung Cancer.**

I. INTRODUCTION

In a thorough review of lung cancer detection methods, the author explores diverse approaches, emphasizing DL techniques on CT images. Various studies showcase high accuracy rates, up to 93%, achieved through preprocessing, segmentation, and classification steps. CNN is recognized for their effectiveness in medical image tasks. The proposed methodology involves CNN-based classification of lung CT scans into specific cancer types, backed by comprehensive evaluation metrics. Results display CT images of normal lungs and different cancer types. The paper concludes by underscoring the potential impact of integrating advanced ML into clinical practice, anticipating future research directions and acknowledging implementation challenges.

Lung melanoma is the largely common cancer-related cause of death worldwide, and the odds of successful treatment are considerably increased by early detection. Lung cancer identification and diagnosis have shown promise for DL techniques. These methods entail utilizing a ML model that has been trained on a enormous gathering of medical pictures, such as computed tomography (CT) scans, to identify anomalies in fresh images. CNNs can be trained on a dataset of CT scans to detect lung cancer. The trained CNNs can then be used to classify different subdivisions of the scan as normal or abnormal. CNN may be trained to identify CT scan patterns, such as nodules or masses, that are suggestive of lung cancer. Other DL techniques for lung cancer diagnosis include

Grenze ID: 01.GIJET.10.2.39

generative models, which may create new CT scans from a given input and identify any anomalies. A set of CT scans can be used to train these models, which can then be applied to create fresh, synthetic CT scans that highlight potential lung cancer-bearing areas in the image. This may be employed in the future to increase the precision and speed of diagnosis.

II. Literature Survey

Various methodologies employing DL and ML techniques for early detection and categorization of lung cancer are comprehensively outlined. These approaches encompass preprocessing, segmentation, attribute extraction, and classification, showcasing promising results in accuracy and potential clinical applications. Significance is placed on the role of CNNs and various algorithms in enhancing diagnostic accuracy and efficiency in lung cancer detection. The author provides a strategy for using DL to classify the high-risk region on Computed Tomography (CT) images and identify the type of lung cancer. The technique can be used with either the original dataset or a modified dataset and is intended to achieve high precision and generalizability. The protocol contains step-by-step instructions for standardizing data, preparing data, implementing models, displaying findings using heatmaps, and performing statistical testing with Python or Origin software. This protocol keeps a promising method for employing DL to precisely detect lung cancer and high-risk regions on CT scans [1]. The construction and analysis of the IQ- OTH/NCCD lung cancer dataset, which includes more than 1100 CT scan pictures of both healthy and tumor- free chest scan from two Iraqi hospitals, are presented in this proposed work. The article suggests a computer system employing image-processing and computer-vision techniques to identify lung cancer in the dataset. Three preparation phases are included in the system: picture enhancement, image segmentation, and feature extraction methodsThe evaluation of several SVM kernels and feature extraction methods yields the greatest accuracy of 89.8876% [2].

The author explores the challenges associated with understanding medical images and highlights the efficiency of the CNNs in overcoming these challenges. CNNs have demonstrated their use as effective instruments for understanding images, often surpassing human experts in this domain. The author has extensively examined the various applications of the CNNs in the interpretation of medical images, involving tasks such as image classification, segmentation, localization, and detection. The applications of the CNNs in medical image analysis of the brain, breast, lung, and other organs have been extensively and meticulously assessed after providing a concise overview of CNNs and their various frameworks. The purpose of this paper is to encourage researchers in the field of medical image understanding to utilize CNNs extensively in their research and diagnostic work. This is because the effectiveness of image understanding carried out by skilled medical professionals is constrained by the scarcity of human experts, their fatigue, and the imprecise estimation procedures. involved. The essay provides a comprehensive evaluation of the challenges that CNNs encounter when analyzing medical images, including a critical examination of these issues [3].

The author talks about an image processing-based method for the early identification and categorization of lung cancer using CT images is proposed [4,14]. This paper focuses on the identification and categorization of various lung cancer subtypes from computed tomography (CT) scan pictures. The use of techniques, such as feature extraction, feature fusion using the Local Binary Pattern (LBP) and the Discrete Cosine Transform ML (DCT), and classification using the SVM and K-Nearest Neighbours (KNN), is offered as a novel approach. In comparison to existing methodologies, the proposed technique achieves an accuracy of 93% for SVM and 91% for KNN, showing encouraging results for the diagnosis of lung cancer [5].

Pre-processing, segmentation, feature extraction using different parameters, and categorization of cancer as benign or malignant make up the system's four steps. The system's diagnostic precision in identifying lung cancer was determined to be 83.33%. For prompt treatment and intervention, early lung cancer detection is essential [6]. The identification of lung cancer, a prominent cause of death worldwide, is the subject of this chapter. The work uses DL methods to identify and categorize lung cancer nodules as Adenocarcinoma, Large Cell Carcinoma, or Squamous Cell Carcinoma using the VGG16, INCEPTIONV3, and RESNET50 models. CT scans, or computed tomography, are used to collect vital data for diagnosis. To increase the accuracy of cancer detection using these DL techniques, the study offers performance evaluation metrics. Improved cancer treatment outcomes are largely dependent on early and precise identification [7].

This paper focuses on the utilization of Low Dose CT (LDCT) scan pictures for the early diagnosis of lung cancer. Lung nodule detection using LDCT scans is successful, however, the image quality is inferior to that of conventional CT scans. CNN, a ML and DL technique, are used in the suggested methodology for feature selection and classification. The study emphasizes CNN's potential for identifying and categorizing nodules. The report also

examines datasets that are now accessible, potential future applications, and difficulties in the field, offering new directions for researchers working on LDCT-based lung cancer early detection [8,16].
Using CT scan image processing, the research suggests a novel method for the identification and categorization of lung cancer. The authors use a variety of preprocessing methods to enhance and smooth images. The lung tumor region is then separated from the lung mask using thresholding and edge detection techniques. For the extracted region of interest (ROI), geometrical parameters like area, eccentricity, perimeter, compactness, and circularity are computed. Based on common training features, the authors use a SVM to categorize the test features as benign or malignant. The findings show that the suggested method detects and categorizes lung cancer nodules with a certain level of accuracy [9]. Using CT scan image processing, the research suggests a novel method for the identification and categorization of lung cancer. The authors use a variety of preprocessing methods to enhance and smooth images. The lung tumor region is then separated from the lung mask using thresholding and edge detection techniques. For the extracted region of interest (ROI), geometrical parameters like area, eccentricity, perimeter, compactness, and circularity are computed. Based on common training features, the authors use a SVM to categorize the test features as benign or malignant. The findings show that the suggested method detects and categorizes lung cancer nodules with a certain level of accuracy [10,15].
The use of a convolutional neural network (CNN) to create a lung cancer detection system is discussed in the paper. The photos were pre-processed and segmented by the authors using a dataset from the Cancer Imaging Archive website. The photos were cropped to a 100 × 100 size. A training dataset was used to train the CNN model, and a testing dataset was used to evaluate its performance. When the model's performance was compared to that of pre-trained models like VGG16, VGG19, and ResNet50, the authors' model fared better. The shortcomings of the current approach are also highlighted in the research, including the difficulty in obtaining CT scan pictures, the poor validity accuracy, and the requirement for increased specificity and reduced false positive rates [11].
In this paper, an automated method for spotting lung cancer in CT scan pictures is presented. CT scans are frequently used to identify lung cancer, a common and fatal disease. However, inaccuracies and delays in detection may occur when interpreting CT scan pictures visually. The suggested approach entails median filtering for image preprocessing, mathematical morphological operations for lung region of interest segmentation, computing of geometric characteristics, and support vector machine classification. With this method, lung cancer will be more accurately and effectively identified in CT scan images [12].
The early identification of cancer using cutting- edge technology and computer-aided systems is the main topic of this paper. The objective is to detect lung cancer at an early stage using ML methods, notably CNN. Due to their lower noise levels compared to other imaging modalities, CT pictures are recommended. The suggested method calls for median filtering to enhance image quality before using the AlexNet CNN architecture to extract and classify features. The classification layer assesses whether the supplied image has a benign or malignant tumor, or whether it is normal, while the CNN's layers extract various low-level and high-level information. The goal of this technology is to increase the precision and effectiveness of detecting lung cancer in its early stages [13].

III. Proposed Methodology

The dataset which consists of images is read in the first step and pre-processed. The preprocessing usually includes ensuring that all the images are annotated, following that, the images are resized accordingly and the necessary features are extracted and stored. The CNN algorithm is then applied to the extracted features and a model is made that can be used for prediction. CNN models consist of multiple layers of interconnected neurons that perform convolutional operations on input data. These convolutional layers are designed to detect local patterns and features in the input, such as edges and corners, and combine them to form higher-level representations. The output from the convolutional layers is typically passed through one or more fully connected layers, which perform classification or regression tasks. With the help of this model, the classification can be done [Figure 1].
The dataset used for the Convolutional Neural Network (CNN) model comprises of CT scan pictures specifically of the lungs. The photos have undergone pre-processing to eliminate noise and standardize pixel intensities. The images are monochromatic, with dimensions of, for instance, 256 x 256 pixels. The dataset has been partitioned into training, validation, and testing subsets. The primary objective of the CNN model is to accurately categorize and classify newly acquired CT scan images. The goal will be accomplished by training the model utilizing convolutional layers, pooling layers, and fully connected layers. The output layer will produce a probability distribution for each input image, indicating the likelihood of belonging to each class. The model will undergo training using the backpropagation algorithm, which is a supervised learning technique, in order to decrease the

discrepancy between the anticipated and actual class labels. The model's performance will be assessed using measures like as accuracy, precision, recall, and F1 score. This model has the potential to assist medical practitioners in accurately identifying CT scan pictures, hence aiding in the detection and diagnosis of lung disorders.

The dataset which consists of images is read in the first step and pre-processed. The preprocessing usually includes ensuring that all the images are annotated, following that, the images are resized accordingly and the necessary features are extracted and stored. The CNN algorithm is then applied to the extracted features and a model is made that can be used for prediction. CNN models consist of multiple layers of interconnected neurons that perform convolutional operations on input data. These convolutional layers are designed to detect local patterns and features in the input, such as edges and corners, and combine them to form higher-level representations. The output from the convolutional layers is typically passed through one or more fully connected layers, which perform classification or regression tasks. With the help of this model, the classification can be done.

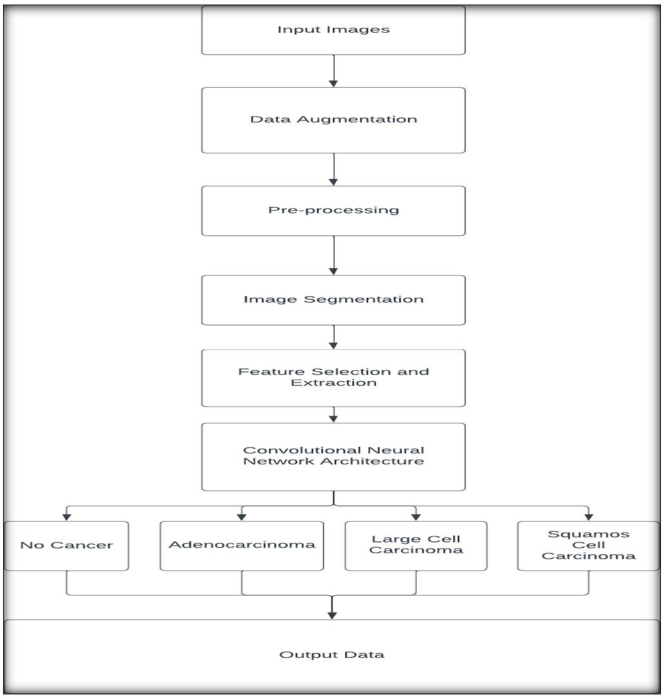

Figure 1: Proposed System

The dataset used for the CNN model comprises of CT scan pictures specifically of the lungs. The photos have undergone pre-processing to eliminate noise and standardize pixel intensities. The images are monochromatic, with dimensions of, for instance, 256 x 256 pixels. The dataset has been segregated into training, validation, and testing subsets. The primary objective of the CNN model is to accurately categorize and classify newly acquired CT scan images. The goal will be accomplished by training the model utilizing convolutional layers, pooling layers, and fully connected layers. The output layer will produce a probability distribution for each input image, indicating the likelihood of belonging to each class. The model will undergo training using the backpropagation algorithm, which is a supervised learning technique, in order to decrease the discrepancy between the anticipated and actual class labels. The model's performance will be assessed using measures like as accuracy, precision, recall, and F1 score. This model has the potential to assist medical practitioners in accurately identifying CT scan pictures, hence aiding in the detection and diagnosis of lung disorders.

IV. Results And Analysis

The result obtained shows the classification of detection of Cancer. Figure 2 shows a CT scan of the lungs without cancer appearing as a clear and uniform view of the lung tissue, with no visible abnormalities or tumors. The scan shows a detailed view of the internal structures of the lung, such as the airways and blood vessels, and is considered normal and healthy.

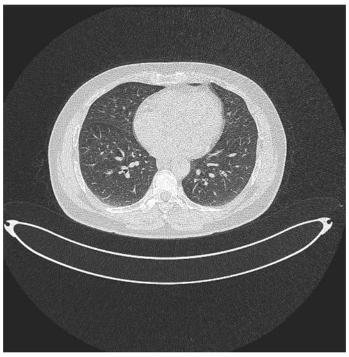

Figure 2: Normal Cas

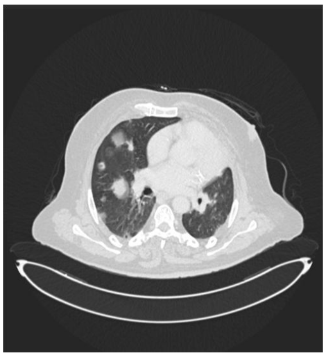

Figure 3: Malignat Case

Figure 4: Benign Case

Figure 3 shows a CT scan image of a lung with adenocarcinoma which has a localized mass or nodule, appearing as a darker region as seen on the image. The appearance of the mass would be indicative of cancer commonly originating in glandular cells. Figure 4 shows a CT scan image of a lung with large cell cancer showing a localized mass or nodule within the lung tissue, appearing as a darker region. The mass is relatively large and irregular in shape, with poorly defined borders. The radiologist would evaluate the appearance of the mass to determine the stage and severity of the cancer. Fig 4 shows a CT scan image of a lung with squamous cell carcinoma that would show a localized mass or nodule within the lung tissue, appearing as a darker region on the image. The mass would have a relatively well-defined border and may show signs of central cavitation.

TABLE II: MODEL EVALUATION METRIC.

Metric	Value	Explanation
Accuracy (Overall)	92.59%	The proportion of correctly classified samples in the validation set.
Loss (Overall)	26.5910	The overall loss calculated during model evaluation on the validation set.
Training Accuracy	100.00%	The accuracy achieved on the training dataset after completion of training.
Training Loss	0.0000	The loss calculated on the training dataset after completion of training.
Validation Accuracy	92.59%	The accuracy achieved on the validation dataset during the final epoch of training.
Validation Loss	26.5910	The loss calculated on the validation dataset during the final epoch of training.

The application of a CNN DL model for the detection of four different types of lung cancer holds tremendous promise in revolutionizing cancer diagnostics and personalized treatment strategies. By leveraging the power of advanced ML algorithms and medical imaging data, our study aims to improve the accuracy and efficiency of lung cancer classification. The integration of DL techniques with medical imaging not only enables early detection but also facilitates the identification of specific cancer subtypes, leading to more tailored treatment approaches. Moreover, the potential integration of the CNN model into routine clinical practice can enhance the overall quality of care for lung cancer patients. This research paves the way for further investigation and validation on larger datasets, ultimately shaping the future of cancer detection and management through the seamless fusion of cutting-edge technology and medical expertise. In this paper, we present the outcomes and performance metrics of the Lung Cancer Detection system. The results highlight the effectiveness of the Convolutional Neural Network (CNN) in accurately classifying lung cancer types [Table I], including Adenocarcinoma, Large Cell, Normal, and Squamous Cell Carcinoma. Screenshots and visual representations of the project in action will be attached to showcase the user interface and the system's response to various inputs.

It also includes a comparative analysis with existing benchmarks or similar studies in the field of lung cancer detection. User feedback or responses from stakeholders, if applicable, will be included to provide insights into the system's usability, user satisfaction, and practical implications in a real-world healthcare setting. The discussion will cover both the strengths and limitations of the Lung Cancer Detection system, addressing any challenges faced during the implementation and potential areas for improvement. The chapter concludes with a summation of the obtained results and their significance in the context of healthcare practices.

V. CONCLUSIONS

Utilizing a CNN DL model for identifying four distinct lung cancer types shows significant potential in transforming cancer diagnosis and tailoring treatment approaches. Our research seeks to enhance the precision and effectiveness of lung cancer classification by harnessing advanced machine learning algorithms and medical imaging data. The amalgamation of DL methods with medical practices underscores our commitment to advancing diagnostic capabilities.

Appendix A: Data set

Gathered during a three-month period in fall 2019 from specialized hospitals, the Iraq-Oncology Teaching Hospital/National Center for Cancer Diseases (IQ-OTH/NCCD) lung cancer dataset comprises 1190 CT scan images representing 110 cases, distinguishing between normal, benign, and malignant classifications. The images, originally in DICOM format, were obtained using a Siemens SOMATOM scanner with specified CT protocol settings, including 120 kV, 1 mm slice thickness, and a window width ranging from 350 to 1200 HU. These cases, encompassing a range of demographics and regions in Iraq, obtained institutional review board approval, with written consent waived, ensuring de-identification before analysis.

REFERENCES

[1] Liuyin Chen, et al, "A DL based CT image analytics protocol to identify lung adenocarcinoma category and high-risk tumor area", Star protocol, CellPress open access, 2022.

[2] Hamdalla F.Kareem, et al, " Evaluation of SVM performance in the detection of lung cancer in marked CT scan dataset", donesian Journal of Electrical Engineering and Computer Science Vol. 21, No. 3, March 2021, pp. 1731~1738 ISSN: 2502-4752, DOI: 10.11591/ijeecs.v21.i3.pp1731-1738 2021.

[3] D. R. Sarvamangala and Raghavendra V. Kulkarni, "Convolutional neural networks in medical image understanding: a survey",Published online: 3 January 2021.

[4] Gagan Thakral et al, "Proposed methodology for Early Detection of Lung Cancer with low-dose CT Scan using Machine Learning", 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON), 26-May 2022.

[5] Jashasmita Pal et al, "DETECTION OF LUNG CANCER USING CT-SCAN IMAGE - DL APPROACH", 2022 OITS International Conference on Information Technology (OCIT), 2022.

[6] Qurina Firdaus et al, "Lung Cancer Detection Based On CT-Scan Images With Detection Features Using Gray Level Co-Occurrence Matrix (GLCM) and Support Vector Machine (SVM) Methods", 2020 International Electronics Symposium (IES), 2020.

[7] Amjad Rehman et al, "Lung Cancer Detection and Classification from Chest CT Scans Using Machine Learning Techniques", 2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA), 2021.

[8] Gaurav Thakur et al, "Lung Cancer Segmentation and Detection using Cuckoo Search Evolutionary Algorithms in Early Stage", 2021 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON) Pune, India. Oct 29-30, 2021.

[9] Nusraat Nawreen, et al, "Lung Cancer Detection and Classification using CT Scan Image Processing", International Conference on Automation, Control, and Mechatronics for Industry 4.0 (ACMI),2021.

[10] S.A.D.L.V. Senarathna, et al, "Lung Cancer Detection and Prediction of Cancer Stages Using Image Processing",3rd International Conference on Electrical, Control and Instrumentation Engineering (ICECIE),2021.

[11] Maithily Marathe, et al, "Detection of Lung Cancer using CT Scans with DL Approach",7th International Conference on Communication and Electronics Systems (ICCES),2022.

[12] Nidhi S. Nadkarni, et al, "Detection of Lung Cancer in CT Images using Image Processing ", Proceedings of the Third International Conference on Trends in Electronics and Informatics,2019.

[13] Sushama Garud, et al, "Lung Cancer Detection Using CT Images and CNN Algorithm ", International Conference on Advances in Computing, Communication, and Control (ICAC3),2021.

[14] Jusman, Y., Ningrum, R. O., & Nurkholid, M. A. F.. Leukemia Cell Image Classification Using CNN: AlexNet and GoogLeNet. In 2023 8th International Conference on Electrical, Electronics and Information Engineering (ICEEIE) (pp. 1-6). IEEE 2023.

[15] Siddiqui, E. A., Chaurasia, V., & Shandilya, M. Detection and classification of lung cancer computed tomography images using a novel improved deep belief network with Gabor filters. Chemometrics and Intelligent Laboratory Systems, 235, 10476 2023.

[16] Kerpel-Fronius, A., Megyesfalvi, Z., Markóczy, Z., Solymosi, D., Csányi, P., Tisza, J. & Bogos, K. (. HUNCHEST-II contributes to a shift to earlier-stage lung cancer detection: final results of a nationwide screening program. European Radiology, 1-9 2023.