



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 5, May 2023

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com



Predicting Lung Cancer Using Convolutional Neural Network Algorithm

Dr. R Senthamil Selvi¹, Aruna A P², Keerthika E³, Mahalakshmi P⁴, Manimozhi N⁵

Associate Professor, Dept. of Computer Science and Engineering, Saranathan College of Engineering, Tiruchirappalli, Tamil Nadu, India¹

Student, Dept. of Computer Science and Engineering, Saranathan College of Engineering, Tiruchirappalli, Tamil Nadu, India²⁻⁵

ABSTRACT: Lung cancer is one of the most deadly types of cancer causing approximately 2.1 million deaths annually. Early detection of lung cancer is difficult and medical imaging techniques, such as chest images, are typically used for lung cancer screening. Deep learning techniques, including Convolutional Neural Networks (CNNs), have shown promise in medical imaging analysis, particularly for lung cancer detection using chest images. The proposed model for lung cancer detection involves two sequential CNN layers with distinct filter sizes followed by two fully connected layers. The model captures local and global features in the input images, and the fully connected layers provide a mapping between the learned features and the output labels. The model's performance is evaluated using metrics such as accuracy and loss values on a publicly available lung dataset with disease labels. The experimental results demonstrate that the proposed model achieves high accuracy in lung cancer detection. The precise and efficient detection of lung cancer using chest X-ray images could have significant implications for the medical community. The proposed model could potentially be used as a tool to assist radiologists in lung cancer screening and early detection, potentially saving lives and improving patient outcomes.

KEYWORDS: Deep Learning, Convolution Neural Networks, CT images

I.INTRODUCTION

The cells that make up the lungs are where lung cancer develops. Many additional cancers, including breast and kidney cancers, have the ability to spread (metastasize) to the lungs. This is so because the location of the original tumour determines what type of cancer it is and how it is treated. Doctors only diagnose lung cancer after carefully examining CT images, which takes a lot of time and is not always correct. Modern optimisation methods and image processing strategies were needed to produce imagery that was as accurate, functional, and efficient as feasible. With the help of the suggested technology, clinicians will be better able to recognise lung nodules early on and learn more about the internal structure. Some flaws in the identification of lung cancer are explored here as part of the contribution. To correctly separate nodules of various sizes and forms, the suggested partitioning method just needs a few parameters. Lung cancer, commonly referred to as lung carcinoma, is a malignant tumour that is characterised by unchecked cell proliferation in lung tissues. To prevent this tumour from metastasizing to other body parts, it must be treated. Carcinomas are the most common type of lung cancer. Small-cell lung cancer and non-small-cell lung carcinoma are the two primary kinds. 85% of lung cancers are primarily caused by long-term cigarette use. About 10-15% of instances involve people who have never smoked but are brought on by radon gas, asbestos, secondhand smoke, and air pollution. The traditional approaches for identifying the presence of lung cancer include computer tomography (CT) and radiography. Biopsy, which is typically carried out via bronchoscopy or CT scan, is used to confirm the diagnosis. Lung cancer accounts for the majority of cancer-related deaths among men. Therefore, it is crucial to develop a new, reliable approach to detect lung cancer at an earlier stage. Nearly 1100 lung image samples have been collected for examination in the current investigation.

II. LITERATURE SURVEY

[1] Deep convolutional neural networks (DCNN) were proposed by Mohanapriya, N., and B. Kalaavathi (2019) for the classification and identification of lung tumours from CT scan pictures. In this study, a DCNN-based architecture is suggested for categorising lung tumours. If the tumour is benign, the likelihood of malignancy is low, and early-stage cancer may be treatable. The likelihood of cancer increases when a tumour develops into a malignant one. Convolutional, pooling, and fully connected layers are just a few of the hidden layers found in deep neural networks. Weight sharing and local connectivity define the convolutional layer.

[2] SACHIN BHAT (2020) suggested using convolutional neural networks to classify and identify lung nodules. For the categorization and identification of nodules, several machine learning (ML) algorithms have been developed. But the choosing of features in ML algorithms is a significant drawback. This could result in limited sensitivity, which would reduce accuracy. However, DL algorithms might be the best option because they automatically learn characteristics from the unprocessed image data without being explicitly provided.

[3] A method for identifying lung cancer from pathological photos using a convolutional neural network (CNN) model is suggested in the publication "Detection of Lung Cancer from Pathological Images Using CNN Model" by Mengshi Chen et al. (2021). The study's findings demonstrated that the suggested CNN model beat other machine learning models in terms of precision, recall, and F1 score and obtained high accuracy. According to the study, CNN models may enhance the precision and effectiveness of lung cancer identification from pathological images, which could have substantial effects on the early detection and management of the disease. In conclusion, the paper offers a viable CNN-based technique for lung cancer identification from pathological pictures. The study's findings imply that the suggested strategy has the potential to increase the precision of lung cancer detection and may be a useful tool for early diagnosis.

[4] A deep learning-based technique for the detection and classification of lung cancer was proposed in the publication "Deep learning for lung Cancer detection and classification" by A. Asuntha and Andy Srinivasan (2021). Area under the receiver operating characteristic curve (AUC-ROC), accuracy, sensitivity, and specificity are some of the metrics the authors use to assess the effectiveness of their model. Additionally, they assess how well their model performs in comparison to other machine learning models.

[5] The study's findings demonstrated that the suggested CNN model beat existing machine learning models in terms of sensitivity, specificity, and AUC-ROC and attained high accuracy. According to the study, CNNs in particular have the potential to increase the precision and effectiveness of lung cancer detection and categorization, which could have important ramifications for the disease's early diagnosis and treatment.

III. PROPOSED METHODOLOGY AND DISCUSSION

The proposed system for lung cancer detection uses a Sequential CNN architecture, which is a type of deep learning algorithm that has shown great promise in medical imaging analysis. The existing systems detect lung cancer using X-ray images which are not good for early diagnosis because of low resolution. These difficulties are overcome by using CT images of the lung. Sequential CNN is used here because it is a popular and effective approach for image classification tasks. The architecture consists of stacking convolutional layers, pooling layers, and dense layers to learn hierarchical representations of the input images. The output of each layer is passed as input to the next layer, allowing the model to learn complex features from the input images. The sequential nature of the CNN ensures that the model can learn features in a hierarchical and ordered manner, which is important for image classification tasks. Additionally, the use of the ImageDataGenerator from Keras helps to efficiently preprocess the image data and generate batches of augmented images for training the model. The architecture of the proposed system is shown below fig III(a).

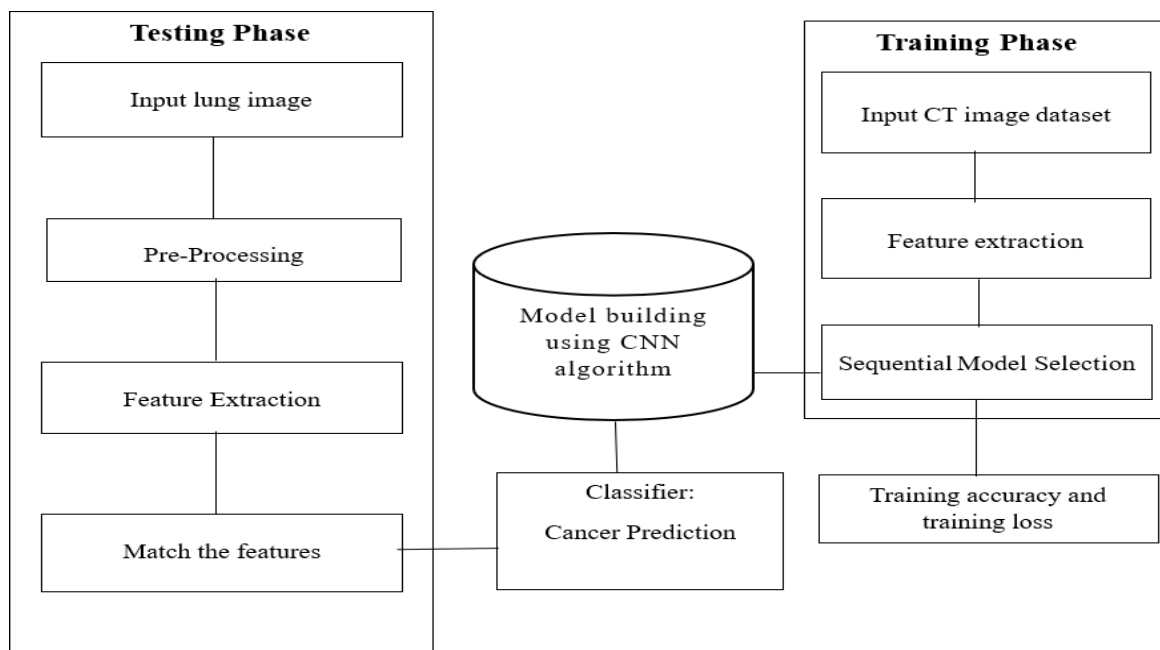


Fig III(a) Proposed System Architecture

COLLECTION OF DATASET

The datasets are gathered in this stage from the Kaggle website, where the images are categorized as benign, malignant, and normal. The feature extraction modules go over this data in more detail and fine-tune them to increase performance.

PRE-PROCESSING

The ImageDataGenerator class in Keras is used to perform data pre-processing and augmentation on image data before feeding it into a neural network. Rescaling is a common practice that normalizes pixel values, and the flow_from_directory method loads and augments images from a specified directory. The class_mode parameter is used to specify the label type, and the classes parameter is used to specify the categorical labels. The ImageDataGenerator also provides various data augmentation options, such as random transformations, to increase the diversity of the training data. Finally, the fit_generator method is used to train the model on the augmented images generated by the ImageDataGenerator.

FEATURE EXTRACTION

The model is created using the Sequential class from Keras, which allows us to build a sequential stack of layers.

The convolutional layers use filters to extract features from the input images. In this case, the first convolutional layer has 16 filters, with a kernel size of (3,3) and a 'relu' activation function. The subsequent convolutional layers gradually increase the number of filters to 32, 64, and 64, respectively.

Each convolutional layer is followed by a MaxPooling2D layer, which downsamples the feature maps by selecting the maximum value within each pooling window.

After the last convolutional layer, a Flatten layer is added to transform the 2D feature maps into a 1D vector, which can be connected to the dense layers.

The first dense layer has 128 neurons with a 'relu' activation function, which learns complex patterns and



representations from the flattened features. The final dense layer has 3 neurons, corresponding to the three classes: 'Benign', 'Malignant', and 'Normal'. The activation function is set to 'Softmax', which outputs probabilities for each class, allowing the model to make a prediction of the most likely class for a given input image.

The ReLU activation function returns the input value itself for any positive input value, and returns 0 for any negative input value, allowing the network to learn non-linear patterns in the data.

$$f(x) = \max(0, x)$$

The Softmax function maps the output values to a probability distribution over the classes, which is useful for classification tasks, allowing the model to predict the most likely class for a given input image.

$$f(x_i) = \frac{\exp(x_i)}{\sum(\exp(x))}$$

```

Model: "sequential_1"
-----
Layer (type)                Output Shape                Param #
-----
conv2d_5 (Conv2D)           (None, 198, 198, 16)       448
max_pooling2d_5 (MaxPooling (None, 99, 99, 16)         0
2D)
conv2d_6 (Conv2D)           (None, 97, 97, 32)         4640
max_pooling2d_6 (MaxPooling (None, 48, 48, 32)         0
2D)
conv2d_7 (Conv2D)           (None, 46, 46, 64)         18496
max_pooling2d_7 (MaxPooling (None, 23, 23, 64)         0
2D)
conv2d_8 (Conv2D)           (None, 21, 21, 64)         36928
max_pooling2d_8 (MaxPooling (None, 10, 10, 64)         0
2D)
conv2d_9 (Conv2D)           (None, 8, 8, 64)           36928
max_pooling2d_9 (MaxPooling (None, 4, 4, 64)           0
2D)
flatten_1 (Flatten)         (None, 1024)                0
dense_2 (Dense)             (None, 128)                 131200
dense_3 (Dense)             (None, 3)                   387
-----
Total params: 229,027
Trainable params: 229,027
Non-trainable params: 0
    
```

Fig III(b) Summary of Sequential CNN

The summary of the CNN model architecture fig III(b), displaying the details of each layer in the model and the total number of trainable parameters. It gives an overview of the model's structure and the output shape of each layer in the form of (batch_size, height, width, channels).

TRAINING THE DATA

Input CT image dataset: The first step in building a CNN model for cancer prediction is to gather a large dataset of CT images of the lung. This may involve collecting images from medical databases or creating a custom dataset by capturing and annotating images of the lung using medical imaging devices.

Feature Extraction: Once the dataset is collected, the images are pre-processed and passed through the



CNN model to extract features from each image. This involves selecting and tuning the CNN model architecture to optimize feature extraction performance.

Sequential Model Selection: The extracted features are then used to train a sequential model that can accurately classify the images as benign, normal, or malignant. This involves selecting and tuning the hyperparameters of the model to optimize its accuracy and performance.

Training accuracy and training loss: During training, the model is evaluated on a validation set to measure its accuracy and performance. This involves calculating the training accuracy and training loss, which are used to adjust the model's parameters and improve its performance.

TESTING THE DATA

Input lung image: The first step is to input a CT image of the lung into the system.

Pre-processing: The input image is pre-processed to standardize the size, colour, and orientation of the image. This may include techniques such as normalization, resizing, and filtering to enhance the quality of the image and reduce noise and distortion.

Feature Extraction: The pre-processed image is then processed through a CNN model that extracts features from the image. This involves passing the image through multiple convolutional layers that identify patterns and features in the image. The extracted features are then used to classify the image as either benign, normal, or malignant.

Match the features: The extracted features are compared to a set of pre-trained features to determine the most likely classification for the input image. This involves using a classifier algorithm that assigns a probability to each classification based on the similarity between the extracted features and the pre-trained features.

PERFORMANCE EVALUTION

Accuracy

Accuracy (ACC) is found as the fraction of total number of perfect predictions to the total number of test data.

It can also be represented as $1 - \text{ERR}$. The finest possible accuracy is 1.0, whereas the very worst is 0.0.

$$\text{ACCURACY} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FN} + \text{FP})} \times 100$$

True positive (TP): number of true positives - perfect positive prediction

False positive (FP): number of false positives - imperfect positive

prediction True negative (TN): number of true negatives - perfect negative

prediction False negative (FN): number of true negatives - imperfect

negative prediction

Loss Probability

The categorical_cross entropy loss function measures the difference between the predicted and actual probability distributions of the target classes. It is used for multi-class classification problems where each instance can belong to only one class.

The formula for categorical_cross entropy loss is:

$$\text{LOSS} = - \left(\frac{1}{N} \right) * \sum (y * \log(y_hat))$$

where N is the number of instances, y is the one-hot encoded true class label vector, and y_hat is the predicted class probabilities vector.

IV. EXPERIMENTAL RESULTS

On the test dataset, the model had an accuracy of 0.99 and a loss of 0.01 for a result range of 0 to 1.

As we have employed sequential CNN, we can add additional layers to CNN to improve accuracy and lower analysis loss rates. Thus, a product with greater accuracy is the end result. The accuracy and loss of the

model increased consistently throughout the course of the 25 epochs, according to visualizations of the training process. Fig IV(a) and (b) depict the plot for model accuracy and loss vs epochs respectively.

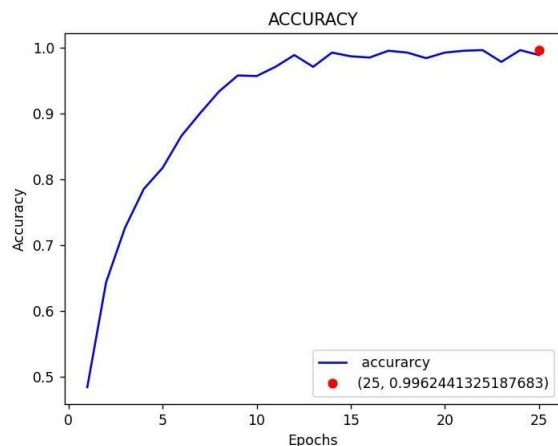


Fig IV(a) Plot for model Accuracy Vs Epochs

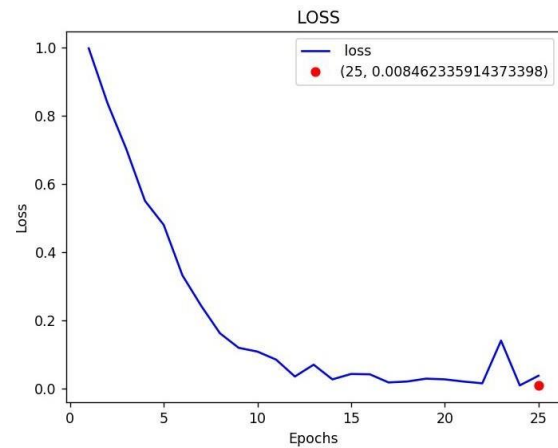


Fig IV(b) Plot for model Loss Vs Epochs

V. CONCLUSION

The major purpose of this project was to prevent inaccurate manual evaluation by humans. We can obtain great accuracy using CNN. And early detections are a possibility. Additionally, it will save the doctors' time. The main issue is that we require many datasets for training. Additionally, it's crucial to maintain the privacy of those datasets. Despite these obstacles, CNN has a lot of promise in accurate prediction and diagnosis. The accuracy will rise, and fatality rates will drop, as algorithm usage is further improved.

REFERENCES

- [1] A. Asuntha, Andy Srinivasan(2020) 'Deep learning for lung Cancer detection and classification', Multimedia Tools and Applications. DOI:10.1007/s11042-019-08394-3.
- [2] Chen, M., Huang, S., Huang, Z., & Zhang, Z. (2021). Detection of Lung Cancer from Pathological Images Using CNN Model. 2021 IEEE International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI). DOI: 10.1109/CEI52056.2021.9470915.
- [3] Emine Cengil, Ahmet Çinar (2019) 'A Deep Learning Based Approach to Lung Cancer Identification', IDAP. DOI:10.1109/IDAP.2018.8620723.
- [4] Jun Gao, Qian Jiang, Bo Zhou, Daozheng Chen (2020) 'Lung Nodule Detection using Convolutional Neural Networks with Transfer Learning on CT Images', Computers in Biology and Medicine. Volume 24, issue 6, 2021 DOI: 10.2174/1386207323666200714002459
- [5] Mengshi Chen, Siming Huang, Ziwei Huang, Zexuan Zhang (2021) 'Detection of Lung Cancer from Pathological Images Using CNN Model', IEEE International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology(CEI).DOI: 10.1109/CEI52496.2021.9574590
- [6] N.Mohanapriya, B.Kalaavathi, T.Senthil Kumar (2019) 'Lung Tumor Classification and Detection from CT Scan Images using Deep Convolutional Neural Networks (DCNN)', International Conference on Computational Intelligence and Knowledge Economy (ICCIKE). 2019, pp. 800-805, DOI: 10.1109/ICCIKE47802.2019.9004247.
- [7] Prenitha Lobo, Sunitha Guruprasad (2018) 'Classification and Segmentation Techniques for Detection of Lung Cancer from CT Images', International Conference on Inventive Research in Computing Applications (ICIRCA).DOI:10.1109/ICIRCA.2018.8597273
- [8] Sachin Bhat, R Shashikala, Sandesh Kumar, K Gururaj (2020) 'Convolutional Neural Network approach for the Classification and Recognition of Lung Nodules', 4th International Conference on Electronics, Communication



and Aerospace Technology (ICECA).DOI: 10.1109/ICECA49313.2020.9297626

- [9] Jun Gao, Qian Jiang, Bo Zhou, Daozheng Chen (2020) 'Lung Nodule Detection using Convolutional Neural Networks with Transfer Learning on CT Images', Computers in Biology and Medicine.
- [10] Tasnim Ahmed, Mst.Shahnaj Parvin (2020) 'Lung Cancer Detection Using CT Image Based on 3D Convolutional Neural Network', Journal of Computer and Communications. Vol.8 No.3.
- [11] Thaseen, M., UmaMaheswaran, S. K., Naik, D. A., Aware, M., Pundhir, P., & Pant, B. (2022). A Review of Using CNN Approach for Lung Cancer Detection Through Machine Learning. 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE).
- [12] Wadood Abdul (2020) 'An Automatic Lung Cancer Detection and Classification (ALCDC) System Using Convolutional Neural Network', 13th International Conference on Developments in eSystems Engineering (DeSE).DOI: 10.1109/DeSE51703.2020.9450778.
- [13] Wang, X., Yang, Q., Cao, Y., & Shi, Y. (2021). A deep learning approach to improve lung cancer detection with PET/CT imaging. IEEE Journal of Biomedical and Health Informatics, 25(2), 758-766.
- [14] Vijay L.Agrawal, Sanjay V.Dudul (2020) 'LIME-Enabled Investigation of Convolutional Neural Network Performances in COVID-19 Chest X-Ray Detection', IEEE Canadian Conference on Electrical and Computer Engineering (CCECE) ,Vol.38.



INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.423



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com



www.ijirset.com

Scan to save the contact details