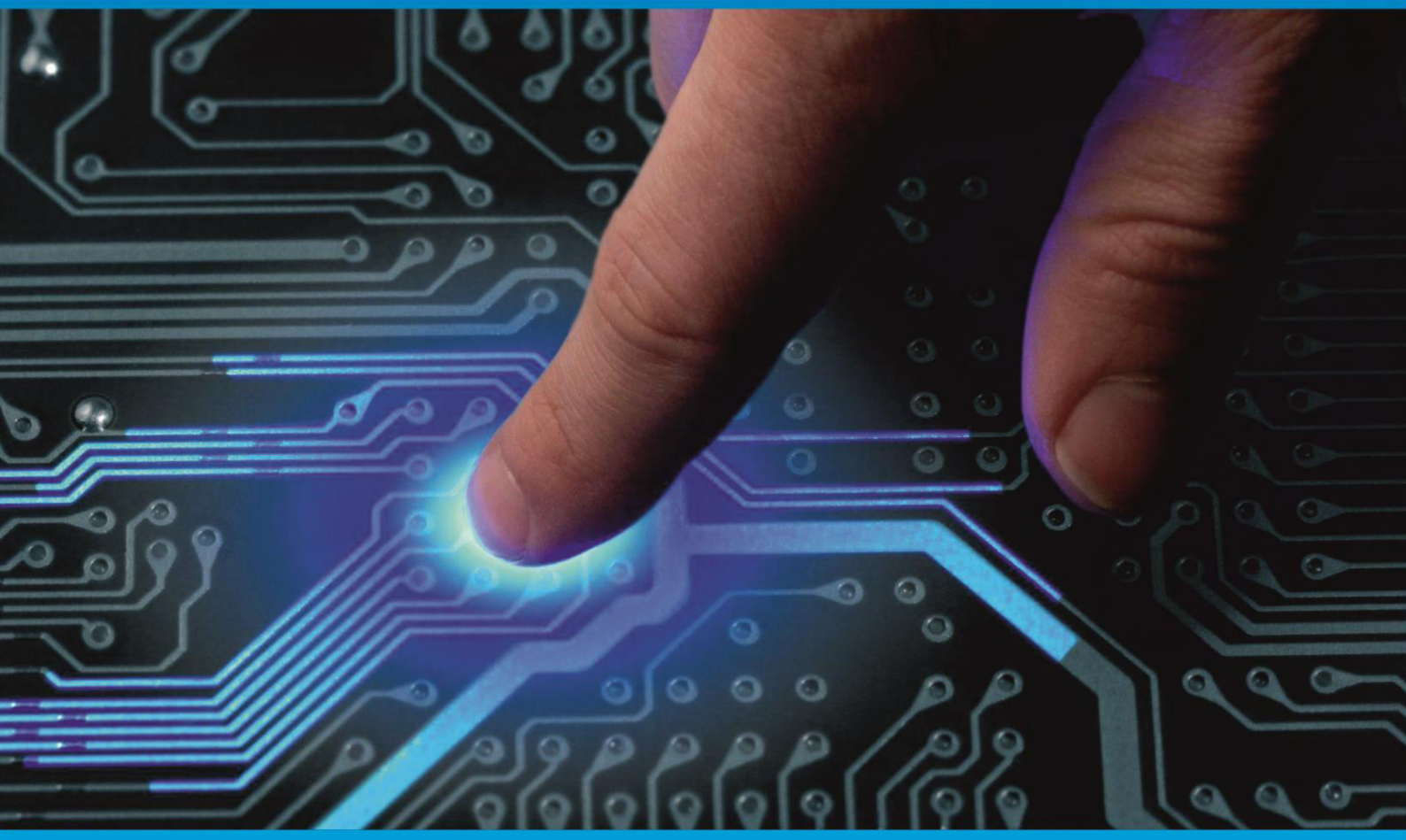
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 7, July 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.165

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007053 |

Lung Cancer Prediction in CT-Scan Images using Deep Learning Techniques

Sneha Tiwari, Stanzin Yangdol, Varshini KS, Yashaswini K, Kavitha K

Dept. of Computer Science and Engineering, Bangalore Institute of Technology, Bangalore, India

ABSTRACT: This study has been undertaken to analyse the performance of different deep learning methods in detecting the lung cancer in CT-Scan images.Lung cancer is a prominent cause of cancer related mortality across the globe. It is very essential to diagnose the cancer in early stages in order to treat and cure the patients. The accurate diagnosis of lung cancer is a key point to make a proper treatment planning to cure and improve the existence of the patients with lung cancer disease. Here, in this paper, the aim will be focused on predicting the cancer and classifying it as Benign or Malignant using segmentation, feature extraction as well as various deep learning techniques.

KEYWORDS: Lung Cancer; Segmentation; Deep Learning

I. INTRODUCTION

Lung cancer is a harmful disease which is also a second leading cause for cancer deaths globally. The uncontrolled growth and spread of abnormal cells leads to cancer which can result in death if the spread is not controlled by taking necessary actions. Lung cancer is the most common cancer worldwide, killing over 1.7 million people per year [2]. Symptoms do not usually occur until the cancer is advanced, and may include persistent cough, sputum streaked with blood, chest pain. The main factors causing lung cancer include cigarette smoking and passive smoking as per world health organization. This causes a serious breathing problem in both inhale and exhale part of the chest.

Though there are many high tech medical facilities for diagnosis and effective medical treatment, the number of deaths due to lung cancer is increasing day by day in all age groups as compared to other cancers and the mortality rate is not yet controlled up to a good extent [1]. Therefore it is highly necessary to diagnose the cancer at early stages for effective treatment. Lung cancer can be detected by using different types of medical images such as computed tomography (CT), chest X-Ray, MRI Scans and so on. The CT images can be cancerous and non-cancerous. In case of cancerous type it can be Malignant or Benign. A Benign tumor has distinct, smooth and regular borders which do not spread to other tissues of our body. Whereas a Malignant tumor has a irregular borders and grows faster than a Benign tumor and rapidly spreads to other parts of the body.

In medical fields now a days Machine learning plays a crucial role for prediction and detection of various diseases at early stages. The various algorithms of machine learning and deep learning make it easier for prediction and classification problems. The crucial application of machine learning is known to be feature extraction that extracts the attributes which are the real information containers that contributes in identifying the diseases. Artificial Neural Networks provide effective methods to deal with the problems associated with images. In the proposed work deep learning algorithms like Convolutional Neural Networks and Recurrent Neural Networks are used for classification.

II. RELATED WORK

This section presents the existing methods and relevant approaches which are surveyed as follows.

Joana Morgado, Tania Pereira and Francisco Silva [3] proposed a system for Lung cancer prediction based on EGFR mutation status. In this system, a variety of linear, nonlinear, and ensemble predictive classification models such as SVM, Elastic Net and logistic regression along with several feature selection methods, were used to classify the binary outcome of wild-type or mutant EGFR mutation status.

HengYu ,ZhiquingZhuo, and Qiming Wang[12] proposed a method to detect Lung cancer in CT images using Adaptive Hierarchical Heuristic Mathematical Model. They have used AHHMM model and designed the deep neural network (DNN) framework to extract high-level features directly from the information and applied the K-means clustering algorithm which has given the accuracy of 90%.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007053 |

Dakhaz Mustafa Abdullah &NawzatSadiq Ahmed [5] proposed a system using various Machine Learning techniques for Lung cancer detection. They have used watershed algorithm for segmentation and Alexnet model for classification of lung cancer using CT-scan images with 92% accuracy.

Kubra Tuncal, Boran Sekeroglu, and Cagri Ozkan [6] proposed a sytem for Lung Cancer Incidence prediction using Machine Learning.They performed comparative study of Support Vector Regression, Backpropagation and Long-Short Term Memory for lung cancer incidence prediction, in which Support Vector Regression performed superior results than other considered algorithms.

Dr. Prabhpreet Kaur and Vikas [7] proposed a system that uses Chi-square feature selection method. The author has performed comparative analysis between SVM and Random Forest with and without feature selection (Chi-square).

H. Xie, D. Yang, N. Sun, Z. Chen, and Y. Zhang [8] proposed a system for Automated pulmonary nodule detection in CT images using deep convolutional neural networks. This system is a novel DCNN approach, consisting of two stages, in first stage nodule candidates are identified with R-CNN trained using online hard negative mining. Second, false positive reduction is performed by 3D DCNN classifiers trained on difficult examples produced during candidate screening with 86% accuracy.

Smita Raut and Shraddha Patil [9] proposed a system that makes use of digital image processing techniques followed by machine learning algorithm that is C 4.5 to discover the tumor in the CT images.

ChinmayiThallam, AarshaPeruboyina and Sagi Sai Tejasvi Raju [10] proposed a system for early stage lung cancer detection.The study involves comparing various classification and ensemble models such as SVM, K-Nearest Neighbor, Random Forest, Artificial Neural Networks and a hybrid model, Voting classifier to evaluate their performances.

Christopher Joy Mathew and Ashwini Maria David [11] proposed a method for lung cancer screening. This system evaluates the role of artificial intelligence and its efficiency in lung cancer screening and classification of nodules.Deep learning techniques that is CNN allows automatic characterization and classification of nodules with 97% accuracy.

T. Arumuga Maria Devi and V. I. Mebin Jose [4] proposed a Three Stream Network Model for Lung Cancer Classification in the CT Images. The proposed architecture has a three-stream network to extract the manual and automated features from the images. Among these three streams, automated feature extraction as well as the classification is done using residual deep neural network and custom deep neural network which gave accuracy of 98%.

III. SYSTEM ANALYSIS AND DESIGN

A. *Existing system and their drawbacks*

To develop an appropriate application, the existing lung cancer applications are reviewed. The drawbacks of existing systems are:

- Models exist to predict the presence of cancer but systems which can detect the type of lung cancer accurately is necessary.
- Standard image processing algorithms for lung cancer detection have some difficulties, such as time consuming, sufficient domain knowledge. For example most of the existing systems use DNN algorithm for classification which take more time to train the model.
- AI assisted diagnosis systems mostly focus on the detection of cancer but not the type (Malignant, Benign).

B. *System Design*

In this section, we will discuss the Architectural System design of our proposed system and Description of the various modules involved. Module and their description are listed below:

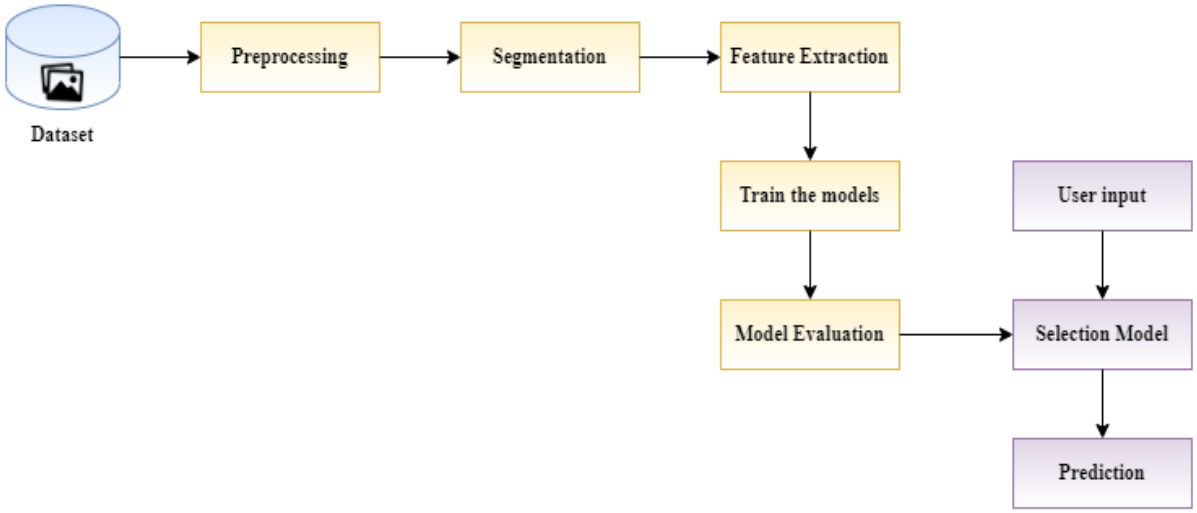
1. *Data Collection:*The CT scan image dataset is obtained from National Centre for Cancer Diseases which contains around 1000 CT-scan images with 3 cases - Benign, Malignant and Normal. Each image in the dataset is of resolution 512 x 512.

2. *Data Preprocessing and Segmentation:*For the images in the original dataset, preprocessing such as grey scale conversion and noise removal is done. We are performing segmentation to extract and separate the foreground region from background region.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007053 |



3. *Feature Extraction:*Geometrical features such as Area, Perimeter and eccentricity of nodules can be considered. These features are extracted by detecting the key points and descriptors using KAZE feature detection method.

4. *Building and Training the Module:*Different Deep learning algorithms such as CNN, CNN_LSTM and RNN are used to build the model and performing comparative analysis among them to select the best model. The model is built using the layers provided in the 'keras' library.

5. *Model Selection:*The performance of the trained models will be evaluated based on the accuracy and the model with best results will be selected for prediction of cancer.

6. *Prediction of Lung Cancer:*The built model will predict whether the person has lung cancer or not from the CT-Scan image and if the cancer is present predicts the type of cancer.

IV. ALGORITHMS

Algorithm design refers to a method or a mathematical process for problem-solving and engineering algorithms. The design of algorithms is part of many solution theories of operation research, such as dynamic programming and divide-and-conquer.

Purpose: To extract features from a set of images and classify the type of cancer present.
Input: A set of CT-Scan images.
Output: Classification as Malignant, Benign and Normal Cases.

Algorithm:

Step 1: Acquire and read the dataset images.

Step 2: For each image in the dataset,

- Perform pre processing such as grey scale conversion and noise removal using gaussian blur.
- Segmentation using Otsu thresholding to extract the foreground region.
- Extract the feature values using KAZE feature detection method that extracts key points and descriptors.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007053 |

Store the feature values of all the images in a csv file

Step 3: Build a model using deep learning algorithms that are CNN, CNN_LSTM, RNN by adding the layers such as convolution, maxpooling, dense, dropout and lstm layers which are available in keras.

Step 4: Calculate accuracy and plot the graphs for all the three algorithms. Select the one which gives higher accuracy.

Step 5: Save the selected model to deploy in the application.

Step 6: Predict cancer by using the trained model for given input CT-scan image.

V. RESULTS

Given below are the graphs depicting the accuracy of different deep learning models.

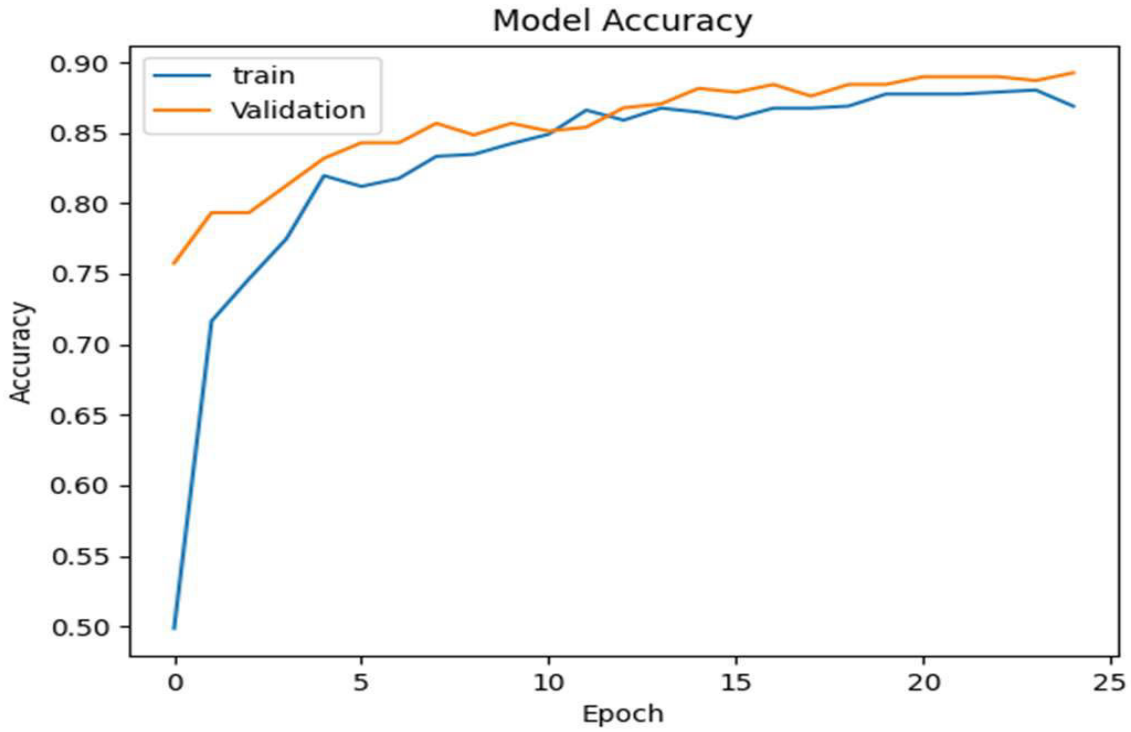


Figure 1: Snapshot of Accuracy for CNN

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007053 |

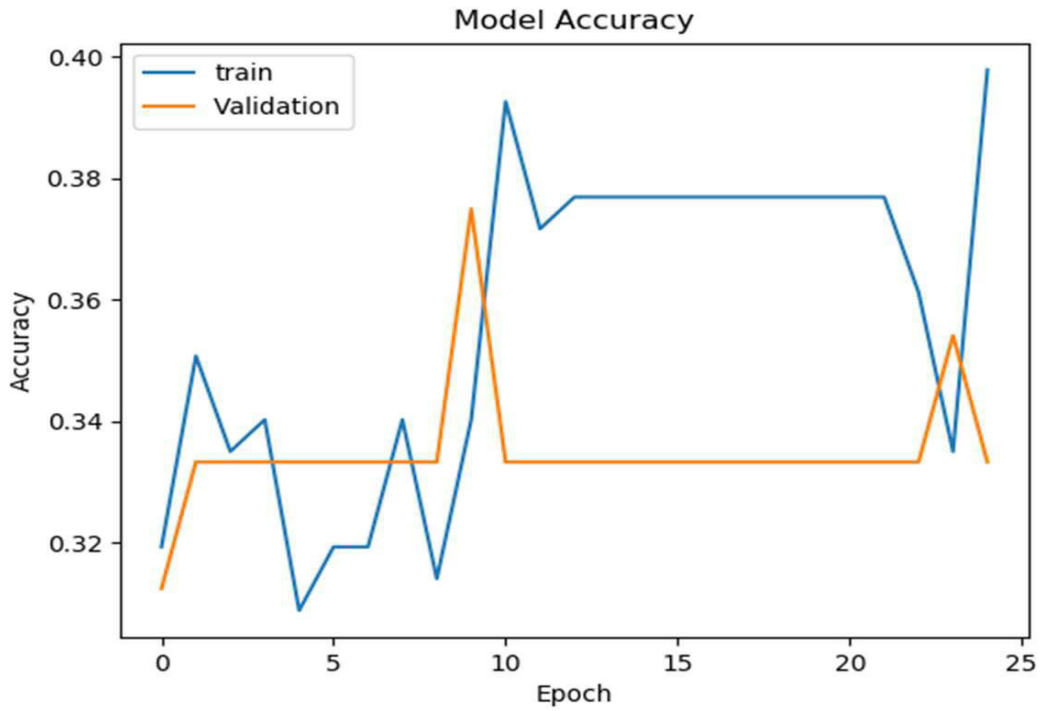


Figure 2: Snapshots of Accuracy for CNN_LSTM

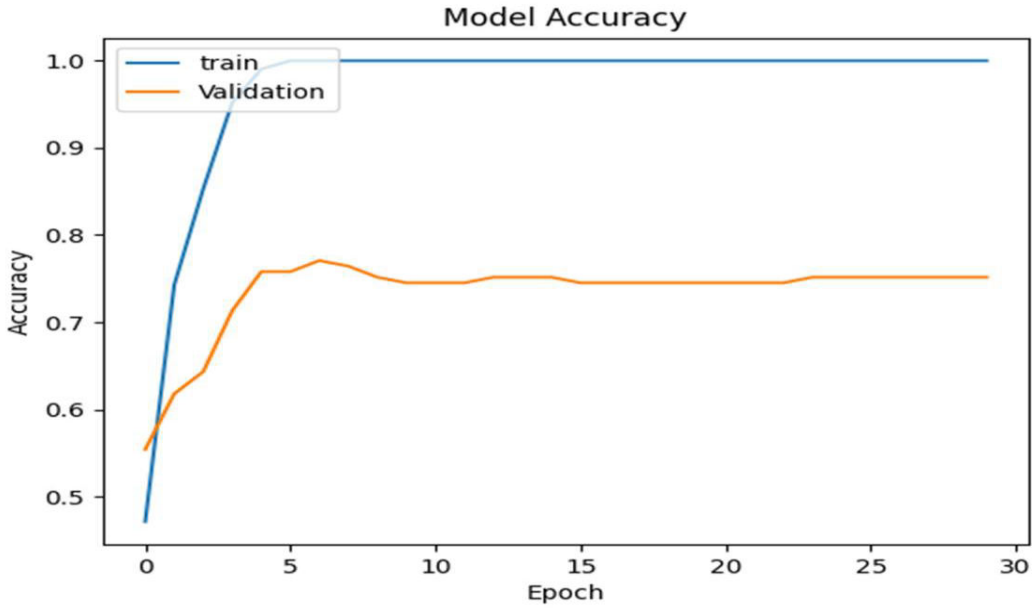


Figure 3: Snapshots of Accuracy For RNN

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007053 |

VI. APPLICATIONS

- The lung Cancer prediction model can be deployed in hospitals to assist doctors to identify lung cancer. The prediction can be taken as a second opinion by the doctors.

- The main application of the proposed lung cancer prediction system is that it can identifythetypeofcanceri.e.,MalignantandBenigninasinglemodelunlikeexisting system which detect only the presence ofcancer.

VII. CONCLUSION AND FUTURE WORK

This project proposes a model to help doctors in early diagnosis of lung cancer and when lung cancer is diagnosed at an early stage, it would be beneficial because the medication will then be initiated to prevent disease from having aharmful result. It provides an efficient method for classifying the types of lung cancer using Deep Learning Techniques such as CNN, CNN_LSTM andRNN.

The application for predicting Lung Cancer from CT-scan images can be further improved by training the model by exploring morenumber of CT images. Features can be extracted more precisely to improve the accuracy of the model so that false negatives can be avoided.

REFERENCES

[1] Patra, R. (2020). Prediction of Lung Cancer Using Machine Learning Classifier. In: Chaubey, N., Parikh, S., Amin, K. (eds) Computing Science, Communication and Security. COMS2 2020. Communications in Computer and Information Science, vol 1235. Springer, Singapore. https://doi.org/10.1007/978-981-15-6648-6_11

[2] Mustafa Abdullah, D., Mohsin Abdulazeez, A., & Bibo Sallow, A. (2021). Lung cancer Prediction and Classification based on Correlation Selection method Using Machine Learning Techniques. *Qubahan Academic Journal*, *1*(2), 141–149.

[3] Morgado J, Pereira T, Silva F, Freitas C, Negrão E, de Lima BF, da Silva MC, Madureira AJ, Ramos I, Hespanhol V, Costa JL, Cunha A, Oliveira HP. Machine Learning and Feature Selection Methods for *EGFR* Mutation Status Prediction in Lung Cancer. *Applied Sciences*. 2021.

[4] Arumuga Maria Devi, T. and Mebin Jose, V. I.. "Three Stream Network Model for Lung Cancer Classification in the CT Images" *Open Computer Science*, vol. 11, no. 1, 2021, pp. 251-261. https://doi.org/10.1515/comp-2020-0145

[5] Dakhaz Mustafa Abdullah &NawzatSadiq Ahmed, "A Review of most Recent Lung Cancer Detection Techniques using Machine Learning", International Journal of Science and Business (IJSAB), vol 5,2021.

[6]Kubra Tuncal, BoranSekeroglu, and Cagri Ozkan, "Lung Cancer Incidence Prediction Using Machine Learning Algorithms," Journal of Advances in Information Technology, Vol. 11, No. 2, pp. 91-96, May 2020. doi: 10.12720/jait.11.2.91-96

[7] Vikas, Dr. Prabhpreet Kaur, "Lung Cancer Detection Using Chi-Square Feature Selection and Support Vector Machine Algorithm", International Journal on Advanced Trends in Computer Science and Engineering, vol.10, ISSN,2021.

[8] H. Xie, D. Yang, N. Sun, Z. Chen, and Y. Zhang, "Automated pulmonary nodule detection in CT images using deep convolutional neural networks", International Conference on Emerging Smart Computing an Informatics (ESCI), ISSN, Jan2019.

[9] Smita Raut , Shraddha Patil , Gopichand Shelke, "Lung Cancer Detection using Machine Learning approach",International Journal of Advanced Scientific Research and Engineering trends, vol. 6, ISSN,2021.

[10] C. Thallam, A. Peruboyina, S. S. T. Raju and N. Sampath, "Early Stage Lung Cancer Prediction Using Various Machine Learning Techniques," *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, 2020, pp. 1285-1292, doi: 10.1109/ICECA49313.2020.9297576.

[11] Christopher Joy Mathew, Ashwini Maria David., "Artificial Intelligence and its future potential in lung cancer screening", EXCLI Journal, vol.10, ISSN,2020.

[12] H. Yu, Z. Zhou and Q. Wang, "Deep Learning Assisted Predict of Lung Cancer on Computed Tomography Images Using the Adaptive Hierarchical Heuristic Mathematical Model," in *IEEE Access*, vol. 8, pp. 86400-86410, 2020, doi: 10.1109/ACCESS.2020.2992645.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details