www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

Vol. 03, Issue 04, April 2023, pp : 483-488

e-ISSN : 2583-1062

Impact Factor : 5.725

LUNG CANCER DETECTION USING DEEP LEARNING

Gandhikota Uma Mahesh[1], Kondamuri Prasanna Devi[2],

Gopi Surya Deepthi[3], Kakada Ganesh[4]

[1,2,34]Aditya Engineering College, Surampalem, A.P, India.

ABSTRACT

Lung cancer is one of the most common and deadly diseases. In fact, it is the most prevalent cancer and the one with the highest mortality rate. Early detection and diagnosis of lung cancer is crucial for improving the patient's chances of survival. Our system was successful in detecting lung cancer in a large number of cases. We believe that our system will be a valuable tool for clinicians and lung cancer patients alike. CT scans are the most common type of imaging used to detect lung cancer. However, CT scans are a challenging task, and they are prone to subjective variability. In this paper, we developed a computer-aided diagnosis system that uses deep transfer learning to handle the scarcity of data. We also designed a CNN model and a MobileNet model to detect lung cancer.

Key Words– Medicinal, Deep Learning, MobileNet model, Convolutional Neural Network.

1. INTRODUCTION

Lung cancer is considered one of the most deadly diseases due to its high mortality rate. Lung cancer among all other types of cancer is the most common and deadly, occurring in both men and women. In lung cancer, a tumour is created by the growth of aberrant cells. Cancer cells spread swiftly as a result of blood flow through the lymphatic fluid found in lung tissue. Cancer cells frequently move to the centre of the chest due to normal lymph flow. Metastases occur when cancer cells migrate to different tissues. Cancer must be diagnosed as early as possible because it tends to spread and cannot be cured if it spreads too far. Lung cancer is difficult to identify since symptoms appear late in the disease's progression, and it is nearly hard to save a person's life at that point. Imaging modalities are used to obtain images of the lungs for assessment. The imaging modalities such as computed tomography (CT), positron emission tomography (PET), magnetic resonance imaging (MRI), and x-rays are used. The CT imaging technique is the most common of the methods mentioned because of its ability to produce an image that excludes overlapping structures. Interpreting and recognizing cancer is difficult for doctors. Lung cancer can be diagnosed by using image processing and deep learning methods. These methods can increase accuracy, as it is difficult to detect tumours and determine their shape, size, and location. Saving a lot of time is made possible by prompt detection. And the patient can benefit from this time by receiving therapy sooner.To categorise tumours into one of the two groups, malignant or benign, in this research, pre-processing (resizing and rescaling) and classification algorithms will be applied. A benign tumour is one that is not malignant and does not metastasize. Malignant cells divide abnormally and unchecked and have the potential to infiltrate nearby tissues. The survival of many people worldwide will be assuredly boosted by early tumour diagnosis.

2. OBJECTIVE

The main purpose of the project is to reduce the amount of manual work and use more up-to-date technology, such as deep learning, transfer learning, and machine learning techniques, such as image processing, transfer learning, and CNN (convolutional neural network), mobilenet.

Our proposed methodology will help develop better treatments that increase the survival chance and quality of life for patients. We will use deep learning algorithms to identify lung nodules on a given CT scan image and detect lung cancer, which will help in the increase of patient's survival rate.

3. EXISTING SYSTEM

The existing method for predicting lung cancer relies on a number of deep learning and machine learning algorithms. However, these algorithms are not as accurate as they could be. This is because deep learning is better suited for dealing with categorical data, whereas the machine learning algorithms are more effective for dealing with numeric data. As a result, the predictions made using these algorithms are not always as accurate as they could be .

When it comes to machine learning, they will work with the numerical data accurately than the 2 dimensional data. It cannot perform the multi classifications.

Disadvantages:

- Less feature compatibility
- Low accuracy
- Less performance

4. PROPOSED SYSTEM:

To improve the accuracy of the existing system, we proposed a method for building a pre-trained model that is better suited for classifying two-dimensional lung CT scans.

- This approach involves these steps.
- We need to collect a dataset of images of CT scans of people with cancer, as well as people who don't have cancer.
- After collecting the data, we will need to do some pre-processing work to resize and reshape the images into the correct format for training our model.
- We will use the pre-processed training dataset to train our models, which are CNN and CNN with Transfer Learning (MobileNet).
- At the time of training our model, we also perform testing simultaneously. After the model is trained, classification is performed.
- The results of a CT scan can either show that the person has cancer, or that the person does not have cancer.

Algorithms

CNN algorithm:

The Convolutional Neural Network is a deep learning network architecture that is commonly used for image recognition. It has been particularly effective for pattern recognition and is often used to classify CT scan images as either positive or negative.

MobileNet:

CNNs are a type of computer vision technology that can identify objects in photos or videos. However, in order to get even better results, modern CNNs get deeper and more complex.

MobileNet is a CNN architecture that is used in real-world applications. The MobileNet variant uses depth-wise separable convolutions in place of the standard convolutions used in earlier architectures, making the models lighter.

Advantages

- As we used pre trained model ,it provides high performance .
- Accurate classification.
- Less complexity.

5. ARCHITECTURE

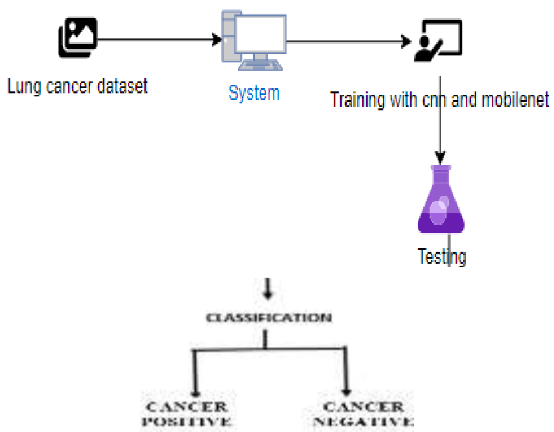

Fig.1 Architecture Diagram

6. SYSTEM REQUIREMENTS

Hardware Requirements

- Processor : I5
- Ram : 8 GB(min)
- Hard Disk :128 GB

Software Requirements

- **Frontend :** HTML,CSS, Java script

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 04, April 2023, pp : 483-488

e-ISSN : 2583-1062

Impact Factor : 5.725

- **Backend**: Python
- **Libraries :** Numpy ,Keras, Pandas, Django, Tensorflow .
- **IDE:** PyCharm, Jupyter notebook

7. Modules

The modules are

- System
- User

8. SYSTEM

1. Data Collection

A dataset is collected from Kaggle which consists of 5000 lung CT Scan images. 2500 of these images are cancer-affected and 2500 are unaffected. For the classification, the dataset is split into a training set (containing the cancer-affected images) and a testing set (containing the unaffected images). The testing set is smaller than the training set, since it is used to test the accuracy of the classifier.

2. Data Pre-Processing

Data pre-processing is going to resize and reshape the images so that we can train our model.

3. Training:

We can use the training data that's been pre-processed before to train our model with a deep learning algorithm, as well as using MobileNet transfer learning methods.

4. Classification :

After the model is trained the classification of the lung CT scan is done which is either positive or negative.

User

1.Upload Image

The user needs to upload an image for which they need to be classified.

2.View Results

The classified image results are shown to the user.

9. UML DIAGRAMS

DFD DIAGRAM

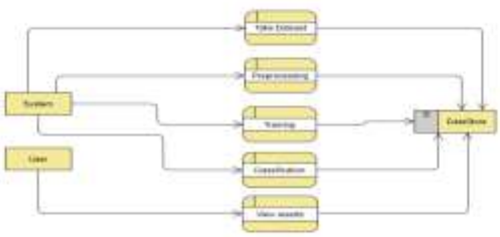

Fig.2 Data Flow Diagram

10. ACTIVITY DIAGRAM

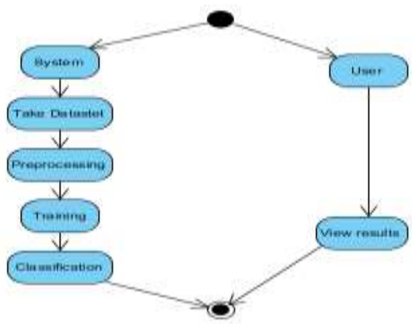

Fig.3 Activity Diagram

ER DIAGRAM

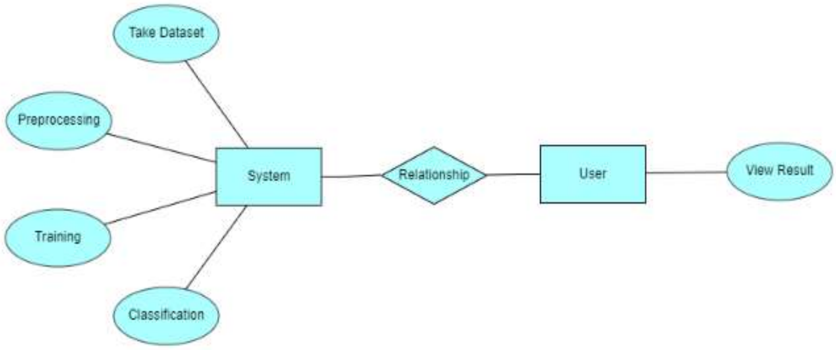

Fig.4 ER Diagram

11. SYSTEM DESIGN

FLOW CHART

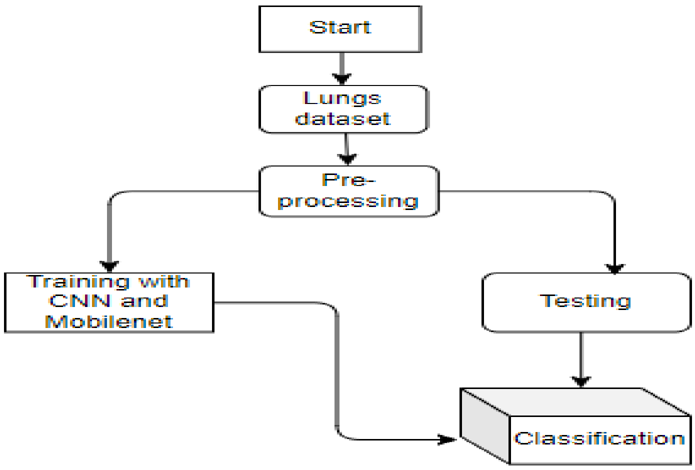

Fig.5 Flow chart

12. MODULE IMPLEMENTION

1. Data Collection Module

We have taken a dataset from the website Kaggle which includes CT scan images of lung cancer patients from both groups - those who have the disease and those who do not. This dataset includes 5000 images..

2. Data Preparation Module

In this module, the collected data is cleaned, preprocessed, and formatted to ensure that it is ready to be used for training and testing the model. This includes tasks such as resizing and reshaping images, splitting data into training and testing sets, converting data to numerical or categorical formats, etc.

3. Model Selection Module

In this module, the appropriate Deep Learning model is chosen for the specific task at hand. We selected CNN and MobileNet models for the classification of lung cancer .

4. Model Training Module

This module is responsible for training the chosen Deep Learning model using the prepared data. The model is optimized and fine-tuned during this process to achieve the best possible performance.The model is starting to learn from the dataset we've given it. We've given it 80% of the data, so it can learn a lot from that. But we also give it a huge amount of data so that the model can learn what kind of patterns are common in the data.

5. Model Evaluation Module

After the model has been trained, it is evaluated to assess its accuracy and effectiveness. The model's performance is tested on previously unseen data to check if it is able to make accurate predictions.

6. Parameter Tuning Module

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
www.ijprems.com
editor@ijprems.com
Vol. 03, Issue 04, April 2023, pp : 483-488
e-ISSN : 2583-1062
Impact Factor : 5.725

In this module, the parameters of the model are adjusted and optimized to improve its performance. The values of these parameters are determined through experimentation and fine-tuning.

7. Prediction Module

The model has been trained and optimized to make accurate predictions on new data. This module uses the model to make predictions in real-time, so we can predict who may have lung cancer.

13. RESULT

The uploaded image is classified.

Fig 6 Image classification of lung cancer negative CT scan image

Fig 7 Image classification of lung cancer positive CT scan image

14. CONCLUSION

In this project we have successfully classified the images of Identification of lungs and liver disease, or its healthy using the deep learning and machine learning. Here, we have considered the dataset of lungs images which will be of different types (healthy or disease) and trained using Mobile Net, CNN along with some transfer learning method. After the training we have tested by uploading the image and classified it.

15. FUTURE SCOPE

This can be used to easily classify the types of different lung cancer, which can tend to be easy to predict and can take the initial medications, precautions, and take measures.

16. REFERENCES

[1] D. S. Kermany et al., ''Identifying medical diagnoses and treatable diseases by image-based deep learning,'' Cell, vol. 172, no. 5, pp. 1122–1131, Feb. 2018, doi: 10.1016/j.cell.2018.02.010.

[2] [R. Kohavi, ''A study of cross-validation and bootstrap for accuracy estimation and model selection,'' in Proc. 14th Int. Joint Conf. Artif. Intell. vol. 2, Aug. 1995, pp. 1137–1143.

[3] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, ''Deep feature extraction and classification of hyperspectral images based on convolutional neural networks,'' IEEE Trans. Geosci. Remote Sens., vol. 5

[4] A. L. Maas, A. Y. Hannun, and A. Y. Ng, ''Rectifier nonlinearities improve neural network acoustic models,'' presented at the Workshop Deep Learn. Audio, Speech Lang. Process., Int. Conf. Mach. Learn. (ICML), Atlanta, GA, USA, vol. 30, Jun. 2013.

[5] B. Udugama, P. Kadhiresan, H. N. Kozlowski, A. Malekjahani, M. Osborne, V. Y. C. Li, H. Chen, S. Mubareka, J. B. Gubbay, and W. C. W. Chan, ''Diagnosing COVID-19: The disease and tools for detection,'' ACS Nano, vol. 14, no. 4, pp. 3822–3835, Mar. 2020, doi: 10. 1021/acsnano.0c02624.

[6] American College of Radiology (ACR). ACR Recommendations for the Use of Chest Radiography and Computed Tomography (CT) for Suspected COVID-19 Infection. Accessed: Jul. 15, 2020. [Online]. Available:

[7] T. Ai, Z. Yang, H. Hou, C. Zhan, C. Chen, W. Lv, Q. Tao, Z. Sun, and L. Xia, ''Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: A report of 1014 cases,'' Radiology, vol. 296, no. 2, pp. E32–E40, Aug. 2020, doi: 10.1148/radiol. 2020200642.

[8] A. Dangis, C. Gieraerts, Y. D. Bruecker, L. Janssen, H. Valgaeren, D. Obbels, M. Gillis, M. V. Ranst, J. Frans, A. Demeyere, and R. Symons, ''Accuracy and reproducibility of low-dose submillisievert chest CT for the

diagnosis of COVID-19,'' Radiol., Cardiothoracic Imag., vol. 2, no. 2, Apr. 2020, Art. no. e200196, doi: 10.1148/ryct. 2020200196.

[9] A. Jacobi, M. Chung, A. Bernheim, and C. Eber, ''Portable chest X-ray in coronavirus disease-19 (COVID-19): A pictorial review,'' Clin. Imag., vol. 64, pp. 35–42, Aug. 2020, doi: 10.1016/j.clinimag.2020.04.001.

[10] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, ''A survey of the recent architectures of deep convolutional neural networks,'' Artif. Intell. Rev., early access, Apr. 21, 2020, doi: 10.1007/s10462-020-09825-6.