2581-4575

International Journal For Recent Developments in Science & Technology

Crossref

A Peer Reviewed Research Journal

LUNG CANCER DISEASE PREDICTION WITH CT SCAN AND HISTOPATHOLOGICAL IMAGES FEATURE ANALYSIS USING DEEP LEARNING TECHNIQUES

[1]R SHESHAN, [2]M SANDEEP, [3]BODDUPALLY VINOD KUMAR, [4]VUCHIDI YASHASWINI

[123]ASSISTANCT PROFESSOR, BRILLIANT INSTITUTE OF ENGINEERING & TECHNOLOGY, ABDULLAPURMET(V&M) RANGA REDDY DIST-501505

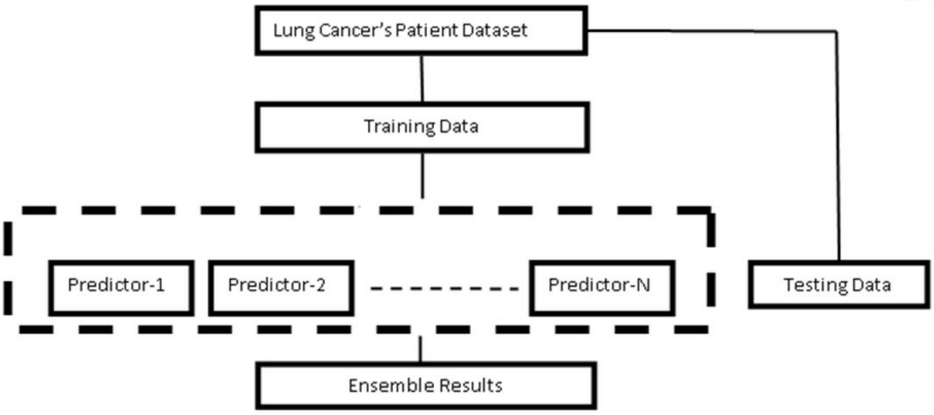
[4]UG SCHOLAR, DEPARTMENT OF CSE, BRILLIANT INSTITUTE OF ENGINEERING & TECHNOLOGY, ABDULLAPURMET(V&M) RANGA REDDY DIST-501505

ABSTRACT

Research and Development on cancer detection is more on imaging than textual data. With the help of documented symptoms in the form of text and Machine Learning (ML) techniques, it is possible to predict the lung cancerstages effectively. This paper conjectures the oeuvre modelwhich is efficient in predicting the stages of lung carcinoma by applying the concepts of ML algorithms. The proposed model is combination of K-Nearest Neighbours, Decision Tree and Neural Networks modelsalong with bagging ensemble method for enhancing the accuracy of the overall prediction. The predictedresults of the suggested model are showing better accuracy compared to individual algorithms.

I.INTRODUCTION

The American Cancer Society, one in four deaths is due to cancer in general with overall survival ratio of 10-15%. World Health Organisation [WHO] says that cancer is a leading cause of death in France and is responsible for 1,50,000 deaths every year.Lung cancer is the most recurrent ones causing high ephemerality rates due to huge pollution and smoking habits. Though prostate and breast cancer occur in male and female, mortality rates caused by lung cancer is higher.There are many prevention and treatment methods like chemotherapy, radiotherapy and surgeries to remove the tumour. Most of the patients across the world are diagnosed at the advanced stage. It has become difficult for the doctors to diagnose in early stage as the symptoms are not much noticeable. Using the initial documented signs as important factors and machine learning techniques, it is possible to predict various stages to certain extent. Machine learning - a field of artificial intelligence, works on programmed algorithms by using their past learnt experienceto draw new conclusions with better accuracy. There are many learning techniques such as classification, regression, association etc. that can be applied in the applications as per the need to meet the best prediction accuracy as close as possible to the human predictions. The choices of algorithms are based on the type of data being operated. Lot of inbuilt libraries are supported in the ML tools and scripting languages for realizing the proposals.



system architecture

II.EXISTING SYSTEMS

Existing systems for predicting lung cancer stages typically rely on conventional methods such as imaging-based assessments and statistical models. Traditional approaches often use imaging techniques like Computed Tomography (CT) scans to visualize tumor characteristics and staging is based on expert radiologists' interpretation of these images. Statistical models, such as logistic regression or decision trees, are also employed to correlate clinical data, such as tumor size, lymph node involvement, and metastasis, with cancer stages. While these methods can provide valuable insights, they have several limitations. First, the accuracy of imaging-based assessments heavily depends on the expertise and experience of radiologists,

which can lead to variability in staging results. Additionally, conventional statistical models may not fully capture the complex and non-linear relationships between various factors influencing cancer stages. These systems are often limited by their inability to integrate and analyze large volumes of heterogeneous data effectively, which can result in suboptimal predictive performance and reduced precision in staging predictions.

III.PROPOSED SYSTEM

The proposed system introduces a novel approach by leveraging deep learning techniques to enhance the prediction of lung cancer stages. This system utilizes advanced neural networks, such as Convolutional Neural Networks (CNNs) and Transformer-based models, to analyze and interpret medical imaging data more effectively. CNNs are employed to automatically extract and learn features from CT scans, capturing complex patterns and abnormalities associated with different cancer stages. Additionally, Transformer models are used to integrate and analyze diverse data sources, including imaging, genetic information, and clinical records, providing a more comprehensive understanding of the disease. One of the key advantages of this approach is its ability to improve prediction accuracy by learning intricate patterns and correlations that are not easily discernible through traditional methods. The deep learning models can handle large-scale data and adapt to new information, offering more precise and consistent staging predictions. Furthermore, the proposed system reduces the reliance on expert interpretation, minimizing human error and variability. This integration of advanced deep learning techniques enhances the overall diagnostic capability and supports more informed treatment planning and patient management.

IV.METHODOLOGY

1. Data Collection and Preprocessing: The initial step involves collecting a comprehensive dataset for lung cancer stage prediction. This dataset typically includes medical imaging data, such as CT scans, along with associated clinical and demographic information, such as tumor size, lymph node involvement, patient age, and smoking history. Publicly available datasets like the LIDC-IDRI or private medical datasets can be utilized for this purpose. Preprocessing of the collected data is crucial for ensuring the quality and consistency of input to the deep learning models. For CT scans, preprocessing steps include resampling to standardize image resolutions, normalization to adjust pixel intensity values, and augmentation techniques such as rotation and flipping to increase data diversity. Clinical data is cleaned and transformed to handle missing values and standardize formats. Data is then divided into training, validation, and test subsets to facilitate model development and evaluation.

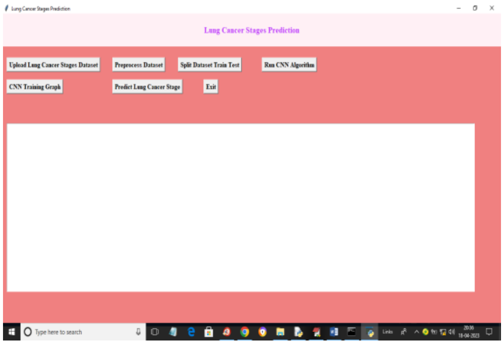
2.Feature Extraction and Model Design : In this phase, features relevant to lung cancer staging are extracted from the preprocessed CT scans. Convolutional Neural Networks (CNNs) are employed for feature extraction, as they are adept at identifying spatial hierarchies and patterns in imaging data. The CNN architecture typically includes multiple convolutional layers followed by pooling layers to reduce dimensionality and capture essential features from the scans. For enhanced predictive capability, Transfer Learning can be applied, using pre-trained models on similar tasks (e.g., ImageNet or other medical imaging datasets) to leverage learned features and fine-tune the model for the specific task of lung cancer staging. Additionally, Transformer-based models or Multi-View Learning techniques may be used to integrate and analyze additional data sources, such as genetic information and clinical records, which contribute to a more comprehensive understanding of the disease.

3. Model Training and Optimization: The deep learning models are trained using the training subset of the dataset. During this process, the models learn to correlate features extracted from CT scans and other data with the known cancer stages. Training involves optimizing model parameters through backpropagation and using optimization algorithms such as Adam or RMSprop. The loss function, often categorical cross-entropy or mean squared error, is minimized to improve the model's accuracy in predicting cancer stages. Regularization techniques, including dropout and weight decay, are applied to prevent overfitting and enhance generalization. The validation subset is used to fine-tune hyperparameters and monitor the model's performance to ensure it performs well on unseen data.

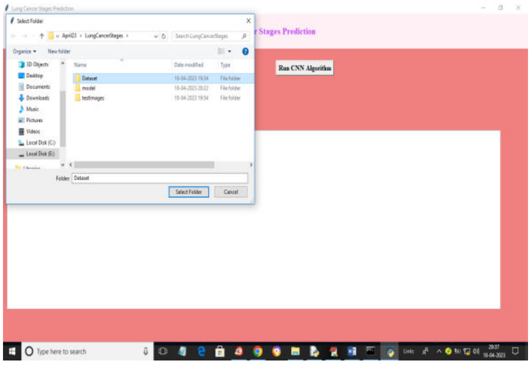
4. Model Evaluation and Testing : After training, the models are evaluated using the test subset to assess their performance in predicting lung cancer stages. Evaluation metrics such as accuracy, precision, recall, F1-score, and Area Under the Receiver Operating Characteristic Curve (ROC AUC) are calculated to gauge the effectiveness of the model. A confusion matrix is generated to visualize the performance across different cancer stages. Comparative analysis is

conducted to benchmark the proposed deep learning models against traditional staging methods, such as radiologist interpretations and statistical models, highlighting improvements in predictive accuracy and reliability. Sensitivity analysis is performed to test the model's robustness under various scenarios and data conditions.

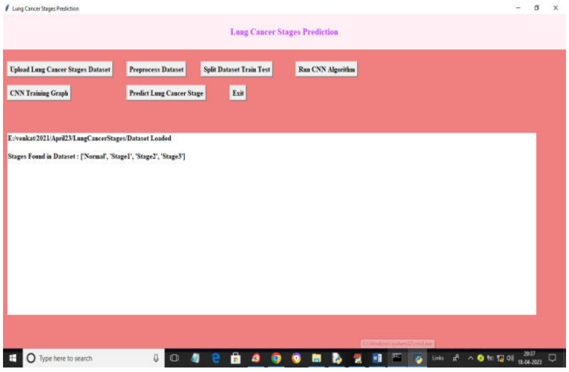
5. Deployment and Integration: The final phase involves deploying the trained deep learning models into a clinical or research setting for practical use. The deployment system integrates the model with existing medical imaging systems to enable real-time staging predictions based on new CT scans. A user-friendly interface is developed to present predictions and associated confidence scores to healthcare professionals. Continuous monitoring and periodic updates are performed to maintain model performance as new data becomes available and as the model encounters evolving patterns in lung cancer staging. Feedback from medical practitioners is incorporated to refine the model and ensure its alignment with clinical needs and standards. To run project double click on run.bat file to get below screen


Lung Cancer Stages Prediction
Lung Cancer Stages Prediction
Upload Lung Cancer Stages Dataset
Preprocess Dataset
Split Dataset Train Test
Run CNN Algorithm
CNN Training Graph
Predict Lung Cancer Stage
Exit
Type here to search

In above screen click on 'Upload Lung Cancer Stages Dataset' button to upload dataset to application and get below output

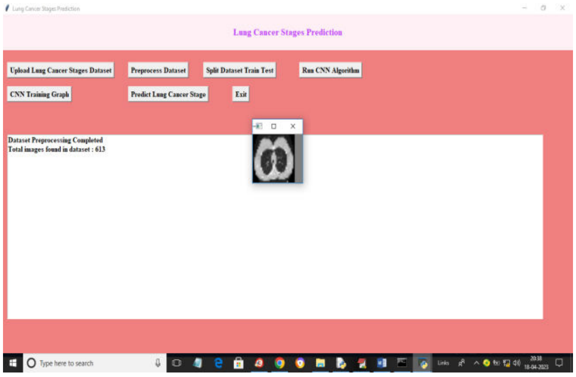


In above screen selecting and uploading dataset folder to application and then click on 'Select Folder' button to load dataset and get below output


Lung Cancer Stages Prediction
Lung Cancer Stages Prediction
Upload Lung Cancer Stages Dataset
Preprocess Dataset
Split Dataset Train Test
Run CNN Algorithm
CNN Training Graph
Predict Lung Cancer Stage
Exit
Type here to search

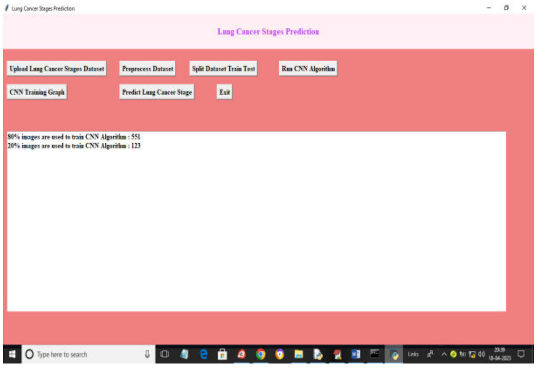
```
E:/venkat/2021/April23/LungCancerStages/Dataset Loaded

Stages Found in Dataset : ['Normal', 'Stage1', 'Stage2', 'Stage3']
```

In above screen dataset loaded and displaying stages found in dataset and now click on 'Preprocess Dataset' button to normalize, shuffle and resize images and get below output


Lung Cancer Stages Prediction
Lung Cancer Stages Prediction
Upload Lung Cancer Stages Dataset
Preprocess Dataset
Split Dataset Train Test
Run CNN Algorithm
CNN Training Graph
Predict Lung Cancer Stage
Exit
Type here to search

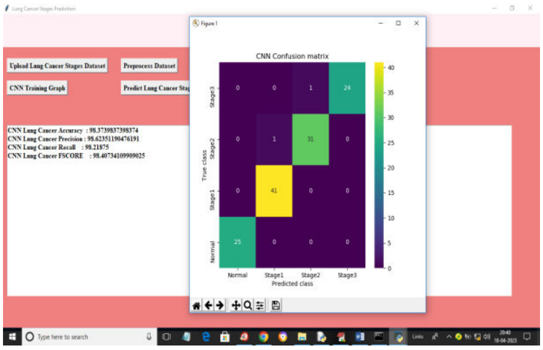
```
Dataset Preprocessing Completed
Total images found in dataset : 613
```

In above screen dataset contains 613 images and we can see processed images and now click on 'Split Dataset Train Test' button to split dataset into train and test and get below output


Lung Cancer Stages Prediction
Lung Cancer Stages Prediction
Upload Lung Cancer Stages Dataset
Preprocess Dataset
Split Dataset Train Test
Run CNN Algorithm
CNN Training Graph
Predict Lung Cancer Stage
Exit
Type here to search

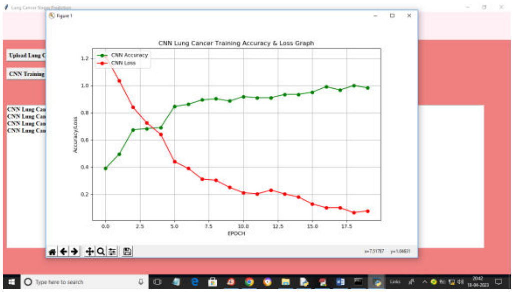
```
80% images are used to train CNN Algorithm : 551
20% images are used to train CNN Algorithm : 123
```

In above screen application using 80% (551) images for training and 20% (123) images for testing. Now click on 'Run CNN Algorithm' button to train CNN and get below output

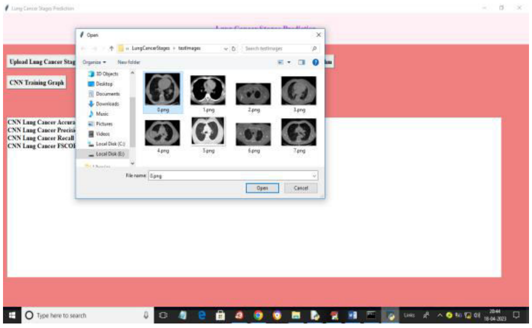


```
CNN Lung Cancer Accuracy  : 98.3739837398374
CNN Lung Cancer Precision : 98.62351190476191
CNN Lung Cancer Recall    : 98.21875
CNN Lung Cancer FSCORE    : 98.40734109909025
```

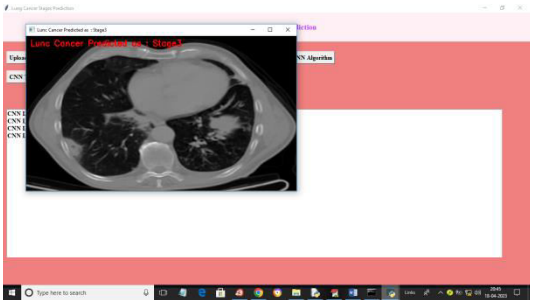
In above screen with CNN we go 98% accuracy and we can see other metrics such as precision, recall and FSCORE. In above confusion matrix graph x-axis represents Predicted Labels and y-axis represents True Labels. All different colour boxes represents correct prediction count and all blue boxes contains incorrect prediction count which is only 2. Now close above graph and then click on 'CNN Training Graph' button to get below graph



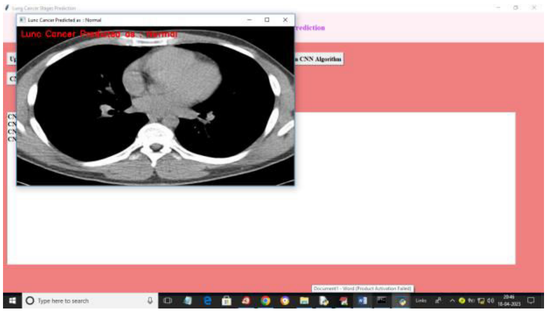
In above graph x-axis represents training epoch and y-axis represents accuracy and loss. Red line represents loss and green line represents accuracy and we can see with each increasing epoch accuracy got increase and loss got decrease. Now close above graph and then click on 'Predict Lung Cancer Stage' button to upload test image and get below output



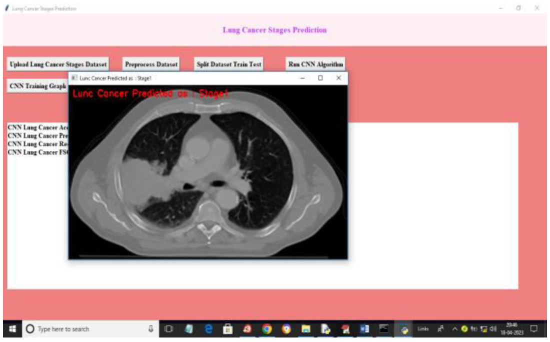
In above screen selecting and uploading '0.png' and then click on 'Open' button to get below output


Lunc Cancer Predicted as : Stage3
Lunc Cancer Predicted as : Stage3
N Algorithm
Type here to search

In above screen in red colour text we can see cancer in image predicted as 'Stage 3'. Similarly you can upload and test other images


Lunc Cancer Predicted as : Normal
Lunc Cancer Predicted as : Normal
rediction
n CNN Algorithm
Type here to search

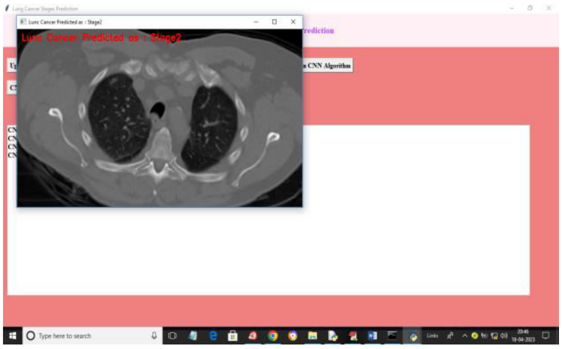
In above screen detected as Normal



Lunc Cancer Predicted as : Stage1

CNN Lung Cancer Ac
CNN Lung Cancer Pre
CNN Lung Cancer Re
CNN Lung Cancer FS

In above screen cancer detected as Stage



Lunc Cancer Predicted as : Stage2

In above screen cancer stage detected as

V.CONCLUSION:

Machine learning algorithms are being extensively used in lot of applications these days.The proposed model uses combination of K-Nearest Neighbour, Decision Tree and Neural Network models with bagging ensemble method topredict the stages of Lung cancer on textual data, enhancing the accuracy of the overall prediction. Conclusions are drawn by comparing the models with bagging and without bagging.It is observed that bootstrap aggregating technique enhances the performance of the individual models with the accuracy scores 0.97 (Decision Tree), 0.94 (K-NN) and 0.96 (Neural Networks). The accuracy score of the integrated model is noticed as 0.98.Integrated model enhances the accuracy by 3.33%.Theproposed model can be used in future for predicting the other chronic diseases in the healthcare and other related domains.Further, the model can be tuned to work with clinical observations along with additional symptoms if recorded during diagnosis phase.

VI.REFERENCES:

[1] R. Zemouri , N. Omri , C. Devalland , L. Arnould , B. Morello , N. Zerhouni , F. Fnaiech, " Breast cancer diagnosis based on joint variable selectionand Constructive Deep Neural Network", IEEE 4th Middle East Conference on Biomedical Engineering (MECBME) 2018.

[2] R. Kaviarasi, A. Valarmathi, "Recognition and Anticipation of Cancer and Non Cancer Prophecy using Data Mining Approach", IEEE International Conference on Emerging Trends in Engineering, Technology and Science(ICETETS) 2016.

[3] Muhammad Imran Faisal , Saba Bashir , ZainSikandar Khan , Farhan Hassan Khan, "An Evaluation of Machine Learning Classifiers and Ensembles for Early Stage Prediction of Lung Cancer", IEEE 3 rd International Conference on Emerging Trends in Engineering Sciences and Technology (ICEEST) 2018.

[4] Yu, Z., Chen, X. Z., Cui, L. H., Si, H. Z., Lu, H. J., & Liu, S. H.,"Prediction of lung cancer based on serum biomarkers by gene expression programming methods", Asian Pacific Journal of Cancer Prevention (APJCP) 2014.

[5] Jennifer Cabrera , AbigaileDionisio, Geoffrey Solano, " Lung cancer classification tool using microarray data and support vector machines", IEEE6th International Conference on Information, Intelligence, Systems and Applications (IISA) 2015.

[6] Ching-Hsien Hsu, GunasekaranManogaran, ParthasarathyPanchatcharam, Vivekanandan S., “A New Approach For Prediction of Lung Carcinoma Using Back Propogation Neural Network with Decision Tree Classifiers”, IEEE 8th International Symposium on Cloud and Service Computing (SC2), 2018

[7] ZhuqingCaia, ZhuliangYua, HaiyuZhoub, ZhenghuiGua, “The Early Stage Lung Cancer Prognosis Prediction Model based on Support Vector Machine”, IEEE 23rd International Conference on Digital Signal Processing (DSP) 2018.

[8] ShutingShen, Ziqiang Fan, Qi Guo, “Design and Application of Tumour prediction model based on statistical method", ISSN:2469-9322.