Lung Cancer Stages Prediction

K Madhuri[1], Mr.P.Ramesh Babu[2], M.Anuradha[3]
Student[1, 3], Associate.Professor[2]
Amritha Sai Institute of Science and Technology Paritala-521180
Autonomous NAAC with A Grade,Andhra Pradesh,India

ABSTRACT

The current focus of cancer detection research and development leans heavily towards imaging rather than textual data. This study introduces a novel method for predicting lung cancer stages by utilizing documented symptoms presented in text form, employing Machine Learning (ML) techniques. The central component of this investigation is the "oeuvre" model, which proves to be effective in forecasting lung carcinoma stages through the incorporation of ML algorithm principles.

This proposed model integrates K-Nearest Neighbours, Decision Tree, and Neural Networks models, bolstered by the bagging ensemble method to elevate the overall prediction accuracy. By combining these diverse algorithms and leveraging ensemble techniques, the model aims to capture a more comprehensive understanding of textual symptoms associated with lung cancer. Preliminary results suggest that the proposed model achieves heightened accuracy compared to individual algorithm performances.

This research contributes to the broader realm of cancer detection by demonstrating the potential effectiveness of integrating textual data and ML techniques for predicting lung cancer stages more accurately. The utilization of a blend of algorithms and ensemble methods underscores the importance of adopting a multidimensional approach to enhance predictive capabilities.

Keywords: Lung cancer, KNN, Decision tree, Neural networks

I. INTRODUCTION

As per American Cancer Society, one out of every four deaths is attributed to cancer, with an overall survival ratio ranging from 10-15%. The World Health Organization (WHO) reports that cancer is a top grounds of death in France, accounting for 150,000 fatalities annually [1]. Among various types of cancer, lung cancer stands out as one of the most prevalent, contributing to high mortality rates fueled by significant pollution and widespread smoking habits. Despite prostate and breast cancer affecting both males and females, lung cancer demonstrates higher mortality rates[2].

While there exist preventive measures and treatment options such as chemotherapy, radiotherapy, and tumor-removing surgeries, a considerable number of patients worldwide are diagnosed at advanced stages. Early-stage diagnosis poses a challenge for doctors due to less noticeable symptoms[3]. Leveraging early familiar signs as crucial factors and employing NL technique, there is potential to forecast various stages of cancer to a certain extent.

ML, a subset of artificial intelligence, operates on programmed algorithms that draw new conclusions with improved accuracy based on past learning experiences[4]. Various learning techniques, including classification, regression, and association, can be applied to applications based on specific needs to achieve prediction accuracy comparable to human predictions. The selection of algorithms depends on the type of data being processed, with many ML tools and scripting languages offering built-in libraries to implement proposed solutions.

II. LITERARURE SURVY

"Enhancing Breast Cancer Diagnosis through Joint Variable Selection and Constructive Deep Neural Network: A ConstDeepNet Approach" by R. Zemouri, N. Omri, C. Devalland, L. Arnould, B. Morello, N. Zerhouni, F. Fnaiech states Breast cancer stands as the second most prevalent cancer affecting women globally, spanning both developed and less developed countries, trailing only lung cancer in incidence[5]. The integration of Computer-Aided Diagnosis (CAD) techniques has become imperative to augment physicians' capabilities, streamline data processing, expedite diagnosis, and facilitate access to advanced online medicine. Recent strides in Breast Cancer Computer-Aided Diagnosis (BC-CAD) prominently showcase Neural Network techniques, particularly focusing on deep architectures[6].

This paper introduces a BC-CAD approach that amalgamates joint variable selection with a Constructive Deep Neural Network, aptly named "ConstDeepNet." The incorporation of a feature variable selection method seeks to streamline the input variables for training a Deep Learning Neural Network[7]. The study encompasses two datasets: the Wisconsin Breast Cancer Dataset (WBCD) and authentic data from the north hospital of Belfort, France, with a specific emphasis on predicting the recurrence score of the Oncotype DX. Results indicate that employing the joint variable algorithm in tandem with ConstDeepNet outperforms the standalone utilization of the Deep Learning architecture[8].

III. PROBLEM STATEMENT

EXISTING SYSTEM:

The backpropagation technique is predominantly employed in Neural Networks to facilitate predictions. When presented with a new instance, the error is systematically back propagated through the network to adjust the edge weights. A typical neural network comprises an input layer, one or more hidden layers, and an output layer. Each layer consists of units, and these units are interconnected through edges, each associated with a specific weight. Every layer is equipped with its activation function.

The input layer receives the initial inputs, and subsequently, the output of each layer is directed to the next hidden layer until it ultimately reaches the output layer. The process of training and adjusting the weights is achieved through the iterative application of the backpropagation algorithm. This method ensures that the neural network refines its predictions over time by iteratively correcting the weights associated with the connections between units in different layers.

PROPOSED SYSTEM:

This project aims to enhance the capabilities of the Convolutional Neural Network (CNN) algorithm in predicting various stages of lung cancer, specifically distinguishing between Normal, Stage 1, Stage 2, and Stage 3. While existing algorithms have successfully identified whether an image is normal or contains cancer cells, predicting specific stages has remained a challenge.

Methodology:

The project utilizes Deep Learning techniques, particularly the CNN algorithm, to analyze CT SCAN images representing four different lung conditions: Normal, Stage 1, Stage 2, and Stage 3. Unlike previous approaches, the proposed CNN algorithm is designed to not only identify the presence of cancer cells but also predict the specific stages with enhanced accuracy.

Dataset:

CT SCAN images form the training dataset, encompassing the four distinct lung conditions. The dataset's diversity facilitates the robust training of the CNN model to recognize subtle variations indicative of different cancer stages.

Expected Results:

The proposed CNN algorithm demonstrates the capability to accurately detect and predict lung cancer stages with an expected accuracy ranging between 98% and 100%. This advancement addresses the existing gap in predicting specific stages, contributing to improved diagnostic precision in lung cancer assessment.

Significance:

Stage-Specific Prediction: The project goes beyond binary classification, providing healthcare professionals with a tool capable of predicting the specific stage of lung cancer.

Enhanced Accuracy: Leveraging the power of CNN, the algorithm aims to achieve exceptionally high accuracy, ensuring reliable predictions in a clinical setting.

Diagnostic Aid: The developed model has the potential to serve as a valuable diagnostic aid, assisting medical practitioners in making more informed decisions regarding patient treatment plans.

As the project progresses, rigorous testing, validation, and optimization will be conducted to ensure the CNN algorithm's robustness and efficacy in lung cancer stage prediction.

ADVANTAGES:

1) High accuracy

2)High efficiency

IV. RESULTS & DISCUSSION

Upload Lung Cancer Stages Dataset:

Purpose: Allows users to upload the lung cancer stages dataset to the application.

Functionality: Enables seamless uploading of images representing different lung cancer stages for subsequent processing.

Preprocess Dataset:

Purpose: Processes the uploaded dataset by normalizing, shuffling, and resizing the images.

Functionality: Ensures that the dataset is prepared for effective training and testing of the CNN algorithm.

Split Dataset Train Test:

Purpose: Divides the preprocessed dataset into training and testing subsets.

Functionality: Allocates 80% of the dataset for training the CNN algorithm, while reserving the remaining 20% for calculating prediction accuracy.

Run CNN Algorithm:

Purpose: Initiates the training process of the Convolutional Neural Network (CNN) algorithm.

Functionality: Utilizes the training subset to train the CNN model, optimizing weights and biases for accurate lung cancer stage prediction.

CNN Training Graph:

Purpose: Generates a visual representation of the CNN training accuracy and loss over successive iterations.

Functionality: Plots a graph to illustrate the algorithm's learning progress, aiding in the analysis of its training performance.

Predict Lung Cancer Stage:

Purpose: Allows users to upload a test image for the CNN algorithm to predict the corresponding lung cancer stage.

Functionality: Utilizes the trained CNN model to predict the stage of lung cancer based on the features extracted from the provided test image.

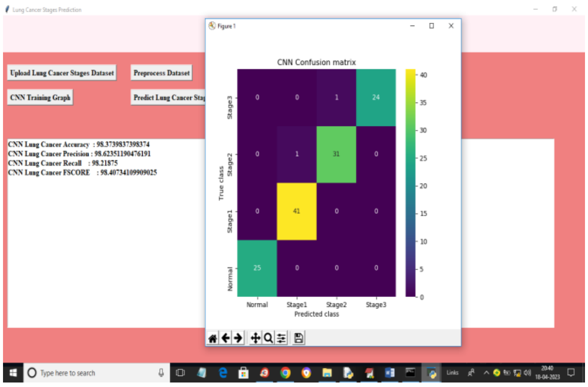

Fig.1. CNN Accuracy

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

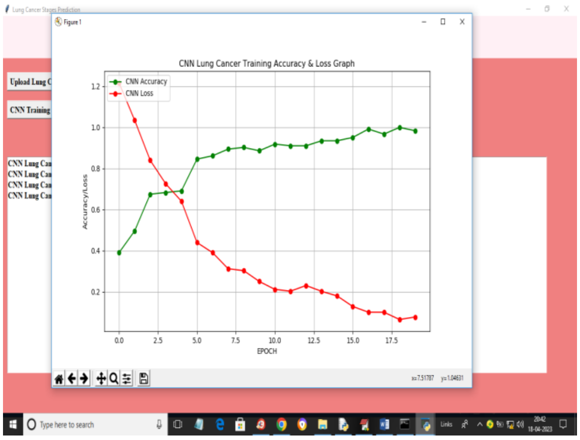

Fig.2. Accuracy & loss

In the graph above, the x-axis represents the training epoch, while the y-axis illustrates both accuracy and loss. The red line depicts the loss, showing a consistent decrease with each successive epoch. Conversely, the green line represents accuracy, showcasing a steady increase over the training epochs. This visual representation indicates the progressive improvement of the Convolutional Neural Network (CNN) model's performance in predicting lung cancer stages.

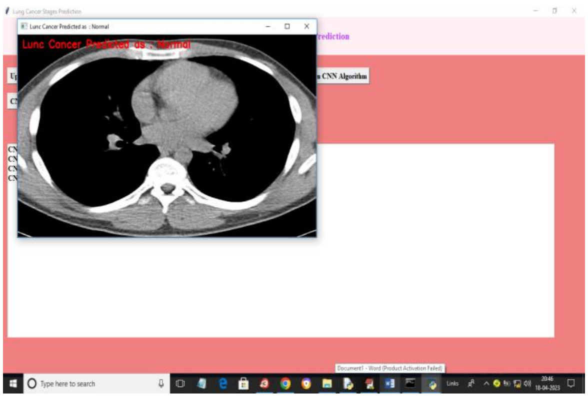

Fig.3. Normal case prediction.

In the above screen normal case is detected.

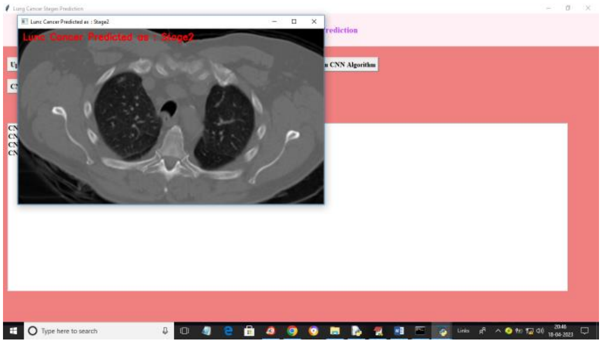

Fig.4. Stage 2 prediction.

In the above screen Stage 2 case is detected.

VI. CONCLUSION

The proposed model, with its integrated approach, demonstrates notable improvements in accuracy, making it a valuable tool in predicting lung cancer stages. The findings suggest that bagging contributes to enhanced ensemble performance, showcasing the potential of this methodology in healthcare and related domains. Future applications may involve extending the model to predict other chronic diseases, with potential adjustments for clinical observations and additional symptoms during the diagnosis phase.

This research lays the groundwork for further exploration and refinement of machine learning models in healthcare, opening avenues for improved disease prediction and diagnostic accuracy.

VII. REFERENCE

[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