Lung Cancer Prediction Model using Resnet50

DR.P.Maragathavalli
Information Technology
Puducherry Technological University
Puducherry,India
marapriya@ptuniv.edu.in

Sanam Reddy Sai Venkata Reddy
Information Technology
Puducherry Technological University
Puducherry,India
Sanamsai2000@pec.edu

Swathi Raj.A.B
Information Technology
Puducherry Technological University
Puducherry,India
swathi04001@pec.edu

Nivedha.C
Information Technology
Puducherry Technological University
Puducherry,India
nive28rockz@pec.edu

***Abstract*— Lung cancer is known as lung carcinoma. It is a disease which is malignant tumor leading to the uncontrolled cell growth in the lung tissue. Lung cancer is caused generally by smoking and the use of tobacco products. Lung Cancer disease is one of the most prominent cause of death in all over world. Early detection of this disease can assist medical care unit as well as physicians to provide counter measures to the patients. Using the pre-trained model is a highly effective approach, ResNet-50 (Residual Network)is a pretrained Deep Learning model for image classification of the *Convolutional Neural Network(CNN, or ConvNet)* .ResNet-50 is 50 layers deep and is trained on a million images of 1000 categories from the ImageNet database.**

Keywords*—*lung cancer, Resnet50, CT images,CNN.

I. INTRODUCTION

Lung cancer is a serious disease which is the major cause of cancer deaths in people worldwide. Timely detection and screening play leading role in prevention of lung cancer. This paper focuses on predicting patients with lung cancer severity at an early stage so that counter measures can be suggested by the physicians. Prediction at an early stage will assist health care systems to handle this disease carefully. Handling the consequence with care may help medical experts to take informed decision and act accordingly Prediction is ubiquitous across the spectrum of cancer care from screening to hospice. Indeed, oncology is often primarily a prediction problem; many of the early stage cancers cause no symptoms, and treatment is recommended because of a prediction that tumor progression would ultimately threaten a patient's quality of life or survival. Recent years have seen attempts to formalize risk prediction in cancer care. In place of qualitative and implicit prediction algorithms, such as cancer stage, researchers have developed statistical prediction tools that provide a quantitative estimate of the probability of a specific event for an individual patient.

.

II. MOTIVATION

Prediction models generally have greater accuracy than reliance on stage or risk groupings, can incorporate novel predictors such as genomic data, and can be used more rationally to make treatment decisions.Resnet50 model was proposed to solve the issue of diminishing gradient. The idea is to skip the connection and pass the residual to the next layer so that the model can continue to train. With Resnet models, CNN models can go deeper and deeper. The best result obtained via Resnet 50 is to re-train nearly 40% of all the parameters.

III. LITERATURE SURVEY

Ai-Min Yang et al discovered that with the development of computer technology, the research of clinical diagnosis and treatment data can be further developed. In recent years, several complex models represented by neural networks, support vector machines and combinatorial lifting algorithms have been widely used. Researchers can find sophisticated methods to predict the risk of disease. Support vector regression machine (SVR) is a branch of support vector machine in the field of regression, twin support regression vector machine (TSVR) is a larger model for regression model improvement, and it greatly reduces the computational complexity of the model by changing the co straints.

Majd A.M.Alhaj et al found that Predicting the survivability of a disease is one of the most interesting purposes of developing a medical data mining applications. This paper applies two classification models (Rule Induction and Random Forest) on the Gaza Strip 2011 cancer patient's dataset, to predict the survivability of cancer patients. The experiments were conducted on the dataset using RapidMiner tool which is used to build the classification models and to measure the performance of them in terms of time consumption and model accuracy. They found that the two algorithms have a convergent accuracy while random forest was less time consuming than rule induction. The rule induction algorithm has the accuracy of 73.63% while Random Forest scores 74.6% accuracy.

Nishtha Hooda et al found that Cancer of the female breast is one of the leading types of cancers worldwide. This paper presents a case study of Malwa Belt in India that has witnessed the proliferation in the overall mortality rate due to breast cancer. The paper researches mortality aspect of the disease and its association with the various risk parameters including demographic characteristics, percentage of pesticides residue present in the water and soil, life style of the women in the affected area, water intake, and the amount of pesticide exposure to the patient. The levels of organochlorine pesticides like DDT and its metabolites and isomers of HCH in blood, tumor and surrounding adipose are estimated. Additionally, an extent of exposure of the subjects to environmental pollutants like heavy metals (Lead, Copper, Iron, Zinc, Calcium, Selenium, and Chromium etc.)are also examined. For the obtained experimental data, an efficient ensemble machine learning based framework called Bagoost is proposed to predict the risk of breast cancer in Malwa women. The performance of the proposed machine learning model results in an accuracy of 98.21 percent, when empirically tested using K-fold cross validation over the real time data of malignant and benign cases and is established to be efficacious than the existing approaches.

[1] discussed that Live wire with Active Appearance model (AAM) strategy is called Oriented Active Appearance Model (OAAM). The Geodesic Graph-cut calculation creates much better division results than some other completely programmed strategies distinguished in writing in the expressions of exactness and period preparing. This strategy besides viably consolidates the Dynamic Appearance Model, Live Wire and Graph Cut tips to abuse their integral focal points. Although the accuracy of predictive models using the decision tree method has no significant advantage, the tree models reveal the most important predictors among genomic information (e.g. KRAS, EGFR, TP53), clinical status (e.g. TNM stage and radiotherapy) and demographics (e.g. age and gender) and how they influence the prediction of recurrence and survivability for both early stage LUAD and LUSC. The machine learning models have the potential to help clinicians to make personalised decisions on aspects such as follow-up timeline and to assist with personalised planning of future social care needs.

IV LIMITATION IN THE EXISTING SYSTEM

The existing paper which follows D-TSVR algorithm has an accuracy of 94% for lung cancer. Medical datasets are often not balanced in their class labels. Most of the existing classification methods tend to perform poorly on dataset which is extremely imbalanced. An increasing number of applications deployed over the cloud operate on datasets which is large and complex that it becomes difficult to gather, store, analyse and visualize. So there arise a scalability issue. The algorithm used can be changed to further increase the accuracy of the model prediction .Accuracy, precision and F1 score can be maximised.

V PROPOSED SYSTEM

A. Input

Dataset is collected from various people. It contains CT image scans of lung from patients. CT images help us to clearly segment and study the image with fine detailing.

B. Proposed system

Firstly, the libraries are imported and the dataset is loaded into the system. Dataset is then cleaned and pre-processed. TensorFlow is used. It is an end-to-end open source platform for machine learning. Keras is also used as it is a high-level neural network library that runs on top of TensorFlow. [4] discussed that Tumor segmentation required also the identical automatic initialization as regarding the liver. This phase was applied only in order to liver volume, obtained following automatic delineation of lean meats surface: this latter, used to original dataset quantity, was used as a new mask in order to be able to prevent processing overloads and even avoid errors related to be able to arsenic intoxication surrounding tissues delivering similar gray scale droit. [5] discussed about detection of leukaemia using a small picture handling method that distinguishes between red blood cells and young white cells. Visual examination of minuscule photos by looking at alterations such as surface, calculation, shading, and measurable research of photographs is now the only recognisable proof of blood trouble. One of the leading causes of death in humans is leukaemia. [10] discussed about diabetic retinopathy from retinal pictures utilizing cooperation and information on state of the art sign dealing with and picture preparing. The Pre-Processing stage remedies the lopsided lighting in fundus pictures and furthermore kills the fight in the picture. Although the Disease Classifier step was used to identify arising wounds and other data, the Division stage divides the image into two distinct classes. [11] discussed that Liver tumor division in restorative pictures has been generally considered as of late, of which the Level set models show an uncommon potential with the advantage of overall optima and functional effectiveness.

C. Output

The performance of the model is evaluated based on precision, recall rate, F1score and support. Confusion matrix for the test data is done. Confusion matrix is a performance measurement for machine learning classification problem where output can be two or more classes. It is a table with 4 different combinations of predicted and actual values. The patient is able to identify whether he/she has cancer or not using this prediction model.

VI DESIGN DIAGRAM

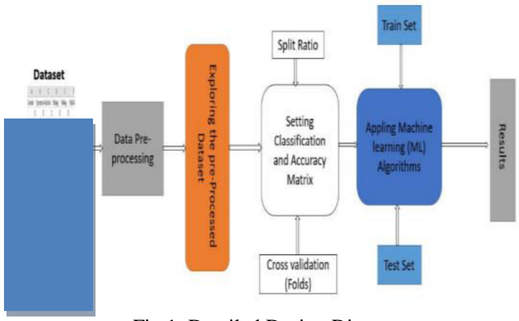

Fig.1 .Detailed Design Diagram.

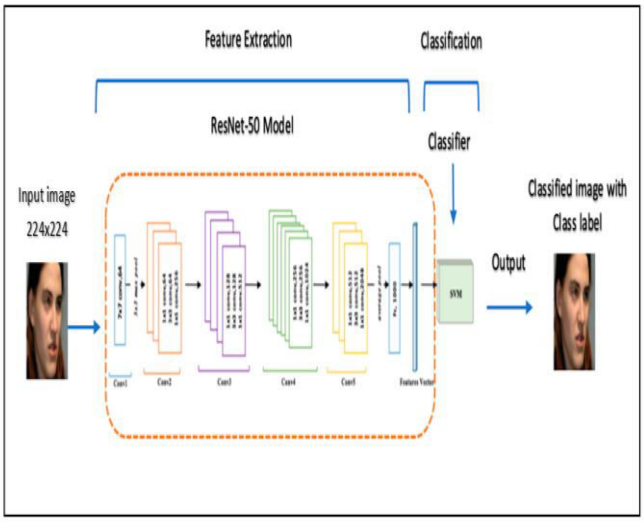

Fig.2. Resnet50 Architecture

VII RESULT ANALYSIS AND DISCUSSION

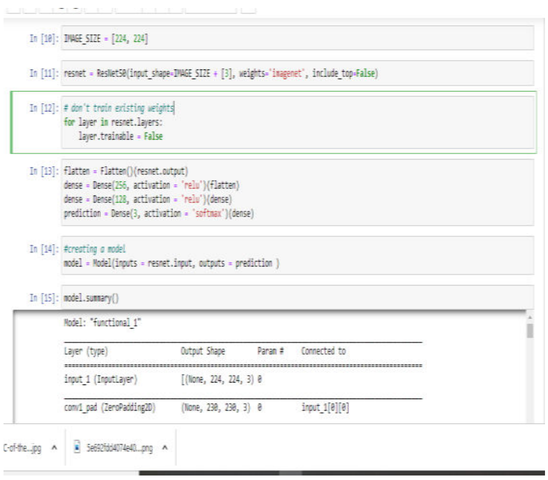

FIG.3. RESNET50 ALGORITHM

Resnet50 algorithm is applied to predict the l stage of the lung cancer

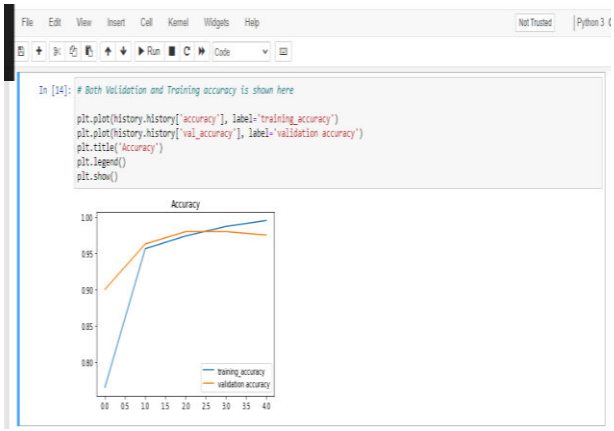

Fig.4.Accuracy Graph

Accuracy graph is plotted between testing and training accuracy.

VIII CONCLUSION

Lung cancer is the fourth most common diagnosed cancer and the second common cause of cancerrelated death worldwide. The prediction of lung cancer is developed based on a supervised learning technique of ML and classification algorithms to produce an accurate outcome. We develop a prediction model using Resnet50. Using Resnet50 networks with large number of layers can be trained easily without increasing the training error percentage. ResNet50 uses the bigger channel layers on much lower image resolution. As it is seen from the prognosis detection and relapse prediction results how much the accuracy of classification is strongly increased.

ACKNOWLEDGMENT

We are deeply indebted to Dr. P. Maragathavalli, Assistant Professor, Department of Information Technology, Puducherry Technological University, Puducherry, for her valuable guidance throughout the project work.

REFERENCES

[1] Christo Ananth, Ruchin Jain, "Enhancing Segmentation Approaches from Oaam to Fuzzy K-C-Means", The Journal of Research on the Lepidoptera, Volume 51, No. 2, 2020,(1086-1108)

[2] Majd A.M.Alhaj, Ashraf Y.A.Maghari,"Cancer survivability prediction using random forest and rule induction algorithms",IEEE, 8th International conference on IT, pp.172-192, 2020.

[3] Nishtha Hooda,Ruchika Gupta, Nidhi Rani Gupta," Prediction of Malignant Breast Cancer Cases Using Ensemble Machine Learning: A Case Study of Pesticides Prone Area", IEEE Access Vol,19, no.2,pp.1096 – 1104,2022.

[4] Christo Ananth, Bindhya Thomas, Priyanka Surendran, Dr.A.Anitha, "Enhancing Segmentation Approaches from GC-GGC to Fuzzy K-C-Means", Annals of the Romanian Society for Cell Biology, Volume 25,Issue 4, April 2021,pp. 2829 – 2834.

[5] Christo Ananth, P. Thenmozhi, Stalin Jacob, Dr.A. Anitha, "Leukemia Blood Cancer Detection Using MATLAB", Turkish Journal of Physiotherapy and Rehabilitation, Volume 32,Issue 3, December 2021,pp. 10257-10261

[6] Nadia G. Elseddeq; Sally M. Elghamrawy; Mofreh M. Salem; Ali I. Eldesouky, "A Selected Deep Learning Cancer Prediction Framework", IEEE Access, vol.9,pp. 151476 – 151492,2021.

[7] Nimmagadda Satyanarayana Murthy, Chaitanya Bethala, "Cancer Prediction and Prognosis using Machine Learning and Deep Learning Models", Springer, Journal of Ambient and humanised computing 2021.

[8] Bijaya Hatuwal, Himal Chand Thapa, "Lung Cancer Detection Using Convolutional Neural Network on Histopathological Images", International Journal of Computer Trends and Technology, pp. 68(10):21-24,2020.

[9] DendiGayathri,Reddy, Emmidi,NagaKumar, DesireddyLohithSaiCharan Reddy, Monika P, " Integrated Machine Learning Model for Prediction of Lung Cancer Stages from Textual data using Ensemble Method", IEEE Access, Vol.29, 2020.

[10] Christo Ananth, D.R. Denslin Brabin, Jenifer Darling Rosita, "A Deep Learning Approach To Evaluation Of Augmented Evidence Of Diabetic Retinopathy", Turkish Journal of Physiotherapy and Rehabilitation, Volume 32,Issue 3, December 2021,pp. 11813-11817

[11] Christo Ananth, M Kameswari, Densy John Vadakkan, Dr. Niha.K., "Enhancing Segmentation Approaches from Fuzzy-MPSO Based Liver Tumor Segmentation to Gaussian Mixture Model and Expected Maximization", Journal Of Algebraic Statistics, Volume 13, Issue 2, June 2022,pp. 788-797

Author Biography

Dr. P. Maragathavalli. She received her B.E. degree in CSE from Bharathidasan University, M.Tech. and Ph.D. degree in CSE from Pondicherry University. She joined Pondicherry Engineering College in 2006 and currently working as Assistant Professor in the Department of Information Technology. She has published several research papers in various refereed journals and international conferences. Her area of interest includes Security Testing, Optimization Techniques, Genetic Algorithm, Machine Learning and Information Security. She is a Life member of ISTE.

Swathi Raj.A.B She is pursuing her B.Tech degree in the Department of Information Technology, Puducherry Technological University.

Nivedha.C. She is pursuing her B.Tech degree in the Department of Information Technology, Puducherry Technological University.

Sanam Reddy Sai Venkata Reddy .He is pursuing her B.Tech degree in the Department of Information Technology, Puducherry Technological University.