Industrial Engineering Journal
ISSN: 0970-2555
Volume: 54, Issue 1, January:2025

LUNG CANCER DETECTION USING DEEP LEARNING

Dr. Shahana Tanveer, Associate Professor, Department of CSE, Deccan College of Engineering and Technology, Osmania University, Hyderabad, Telangana :: shahanatanveer@deccancollege.ac.in
Ayesha Ifteqar, PG Scholar, Department of CSE, Deccan College of Engineering and Technology, Osmania University, Hyderabad, Telangana :: ayeshaifteqar.10@gmail.com

***Abstract*—** In the last many years, lung cancer has become a major public health concern. To examine cell breakdown in the lungs in its starting stages, doctors often use imaging modalities such as X-ray chest films, CT scans, MRIs, etc. The timing of diagnosis determines the course of therapy. Artificial intelligence (AI) is a hotspot for developing computational models of human intellect. In this research, we aim to enhance the detection and classification of lung nodules from CT images using a novel deep learning approach. Our study builds upon an extensive review of existing lung cancer detection methods, highlighting their strengths and weaknesses. Lung Cancer is one of the leading lives taking cancer worldwide. Early detection and treatment are crucial for patient recovery. Medical professionals use CT scan images from potentially infected areas of lungs for diagnosis. Most of the time, the diagnosis regarding the types of lung cancer is error-prone and time-consuming. Convolutional Neural networks can identify and classify lung cancer types with greater accuracy in a shorter period, which is crucial for determining patients' right treatment procedure and their survival rate. adenocarcinoma, large cell carcinoma, normal, squamous cell carcinoma images dataset are considered in this research work

.

***Index Terms*— Artificial Intelligence, Deep Learning, Lung Cancer, Machine Learning**

I. INTRODUCTION

Lung cancer is prominent cancer among both men and women, making up almost 25% of all cancer deaths [1]. The primary cause of death from lung cancer, about 80% is from smoking. Lung cancer in non-smokers can be caused by exposure to radon, second-hand smoke, air pollution, or other factors like workplace exposures to asbestos, diesel exhaust, or certain other chemicals lung cancers some people who do not smoke [2]. Various tests like imaging sets (x-ray, CT scan), Sputum cytology, and tissue sampling (biopsy) are carried out to look for cancerous cells and rule out other possible conditions. While performing the biopsy, evaluation of the microscopic histopathology slides by experienced pathologists is indispensable to establishing the diagnosis [3], [4], [5], and defines the types and subtypes of lung cancers [6]. For pathologists and other medical professionals diagnosing lung cancer and the types is a time-consuming process. There is a significant change the cancer types are misdiagnosed, which directs to incorrect treatment and may cost patients' lives. Machine Learning (ML) is a subfield of Artificial Intelligence (AI) that allows machines to learn without explicit programming by exposing them to sets of data allowing them to learn a specific task through experience [7][8]. In previous research papers, most of the authors considered using x-rays, CT scans images with machine learning techniques such as Support Vector Machine (SVM), Random Forest (RF), Bayesian Networks (BN), and Convolutional Neural Network (CNN) for lung cancer detection and recognition purpose. Some papers also considered using histopathological images, but they distinguish between carcinomas and non- carcinomas images and with lower accuracy. This research paper has considered using Convolutional Neural Network (CNN) architecture to classify the benign, Adenocarcinoma, and squamous cell carcinomas. We have not found other papers using the CNN model to classify only the given three different histopathological images and the given model's accuracy.

The treatment of the illness depends on how early the sickness is recognized with the goal that therapy can keep the infection from progressing (in stage) and spreading to different pieces of the body. The infection can be controlled with great clinical consideration and a few treatments including a medical procedure, chemotherapy, and radiographs, for various reasons including the patient's wellbeing and the illness' movement. The survival rate of five years is however just 21%. To overcome the medical challenges, Image Processing and Artificial Intelligence approaches may be used to process medical field data using technology solutions in order to discover and diagnose the disease at an initial stage which will not only help medical practitioners to deliver effective results but also help to save valuable human lives. Machine and deep learning algorithms are critical in training a computer system to become an expert that can assist in making predictions and taking decisions. Machine learning is a subfield of AI that makes it possible for computers to automatically learn new skills by analyzing and interpreting data that has already been collected. As a branch of ML, deep learning enables computers to "learn" from data and form perceptions of the environment according to the rankings of their own ideas. These fields imbibe a computer with intelligence, enabling it to extract patterns based on specific facts and then process them for autonomous reasoning. AI is a prominent area for representing human intelligence in a machine as shown in Fig. 1 AI is a subset of simulated intelligence, while Profound Learning is a subset of AI.

II. LITERATURE REVIEW

Machine learning techniques are used in [2] using K-Nearest Neighbours, Support Vector Machines, Naïve Bayes, Decision Trees, and Artificial Neural Networks. Using principal component analysis (PCA), chest radiograph dimensions are reduced by a factor of 1/8. To compare the efficacy of machine learning methods with and without principal component analysis (PCA), the APR, F-measure, and precision are used. Using an accuracy rate of 93.24%, Decision Tree outperforms all other performance metrics when tested using the actual data.
Methods for segmentation including Region Growing, Marker Controlled Watershed, and Marker Controlled Watershed with Covering have been utilized in [3] Pre-handling. The following kinds of machine learning are used: Support Vector Machine, Neural Network, Naive Bayes classifier, Decision Tree, KNN, Gradient Boosted Tree, and MLP. Compared to previous segmentation approaches, pre-processing marker-controlled watershed with masking segmentation yields more accurate results. An accuracy level of approximately 97% was achieved when image data was segmented using a marker-controlled watershed-based segmentation and a multi-class SVM classifier. Read about SVMs, Random Forests, and Artificial Neural Networks that rely on machine learning in [4]. When comparing algorithms, In order to select the one that provides the most accurate predictions, it is necessary to compute characteristics such as precision, recall, and accuracy. With a precision of 92% for locale based highlights and 96% for surface based highlights, Artificial Neural Networks outperform the other methods. Image categorization, object recognition, and feature extraction are three areas where the authors predict Deep Learning will eventually surpass machine learning. The authors presented a four-step process in [5].

The first step is pre-processing, which involves applying morphological smoothening and median filters.

The GLCM (Dark Level Co-Event Network) strategy is utilized to extricate the elements from the pre-handled picture. In the second and final stages of lung disease identification and separation, classifications based on multilayer perceptrons (MLP), SVM, and KNN are utilized. The last step is to test how well the classifier performed. Authors may save time and memory by using GLCM to derive a matrix that includes just the characteristics that are required. With MLP, the authors attained 98% accuracy, with SVM reaching 70.45% and KNN reaching 99.2%.

The authors W. Ausawalaithong, A. Thirach, S. Marketta, and T. Wilaiprasitporn [9] used deep learning with a transfer learning approach to predict lung cancer from the chest X-ray images obtained from different data sources. Image size of 224X224 with 121-layer Densely Connected Convolutional Network (DenseNet-121) and a single sigmoid node was applied in a fully connected layer. The proposed model achieved 74.43±6.01% mean accuracy, 74.96±9.85% of mean specificity, and 74.68±15.33% mean sensitivity for different image source dataset. T. Atsushi, T. Tetsuya, K. Yuka, and F. Hiroshi [10] applied Deep Convolutional Neural Network (DCNN) on cytological images to automate lung cancer type classification. They considered Small cell carcinoma, Squamous cell carcinoma, Adenocarcinoma images in their dataset. The DCNN architecture of 3 convolution and pooling layers and 2 fully connected layers with dropout 0f 0.5 were used. The model developed was able to achieve the overall accuracy of 71.1%, which is quite low.

W. Rahane, H. Dalvi, Y. Magar, A. Kalane, and S. Mondale [11] proposed using image processing and machine learning (Support Vector Machine) for lung cancer detection on computed tomography (CT) images. Image processing like grayscale conversion, noise reduction, and binarization was carried out. Features like area, perimeter, and eccentricity from the segmented image region of interest were fed to the support vector machine (SVM) model.

M. Šarić, M. Russo, M. Stella, and M. Sikora [12] proposed CNN architectures implementing VGG and ResNet for lung cancer detection using whole side histopathology images, and the output was compared using the receiver operating characteristic (ROC) plot. Patch level accuracy of 0.7541 and 0.7205 was obtained for VGG16 and ResNet50 respectively which is quite low. The authors explained that the given models' low accuracy was due to large pattern diversity through different slides.

The authors S. Sasikala, M. Bharathi, B. R. Sowmiya [13], proposed using CNN on CT scan images to detect and classify lung cancer. They used MATLAB for their work and has two phases in training to extract valuable volumetric features from input data as the first phase and classification as the second phase. Their proposed system could classify the cancerous and non-cancerous cells with 96% accuracy. SRS Chakravarthy, R. Harikumar [14], used Co-Occurrence Matrix (GLCM) and chaotic crow search algorithm (CCSA) for feature selection on computed tomography (CT) and applied probabilistic neural network (PNN) of the classification task. They found that the PNN model build on CCSA features performed better with 90% accuracy.

Industrial Engineering Journal
ISSN: 0970-2555
Volume: 54, Issue 1, January:2025

III. EXISTING METHODS:

Computer-aided diagnosis (CAD) is cutting-edge technology in the field of medicine that interfaces computer science and medicine. CAD systems imitate the skilled human expert to make diagnostic decisions with the help of diagnostic rules.

The performance of CAD systems can improve over time and advanced CAD can infer new knowledge by analyzing the clinical data. To learn such capability the system must have a feedback mechanism where the learning happens by successes and failures. During the last century, there is a dramatic improvement in human expertise and examination tools such as X-ray, MRI, CT, and ultrasound.

Different image processing methods have been innovated for detecting cancer and Implemented as a median-wiener filter in the preprocessing step. To detect whether the nodule is cancerous or not, the classification network has been used, such as Support Vector Machines (SVM.

.

IV. PROPOSED SYSTEM

This project has considered using Convolutional Neural Network (CNN) architecture to classify the adenocarcinoma, large. cell. Carcinoma, normal, squamous. cell. Carcinoma.

We have implemented this project using the CNN VGG 19 model to classify only the given lung cancer images which are of 4 categories images and the given model's accuracy.

In proposed system multiple deep learning algorithms are used efficient net, inception and accuracy, precision, f1 score are calculated and best model is used for prediction.

.

METHODOLOGY:

- **Data COLLECTION:**
  The ct scan images are obtained from Kaggle image dataset . Three classes of adenocarcinoma, large. cell. Carcinoma, normal, squamous. cell. Carcinoma of lungs with 5000 CT scan images in each category, are considered for our work.

- **Pre-processing:**

  Pre-processing is a procedure adopted to enhance the quality of images and increase visualization. In medical imaging, image processing is a crucial phase that helps to improve the images quality. This can be one of the most critical factors in achieving good results and accuracy in next phases of proposed methodology. Lung cancer images may contain a different issue that may lead to poor and low visualization of the image. If the images are poor or of low quality, it may lead to unsatisfactory results. During preprocessing phase..

- **Train-Test Split and Model FITTING:**

  Now, we divide our dataset into training and testing data. Our objective for doing this split is to assess the performance of our model on unseen data and to determine how well our model has generalized on training data. This is followed by a model fitting which is an essential step in the model building process.

- **Model Evaluation and Predictions:**

  This is the final step, in which we assess how well our model has performed on testing data using certain scoring metrics, I have used 'accuracy score' to evaluate my model. First, we create a model instance, this is followed by fitting the training data on the model using a fit method and then we will use the predict method to make predictions on x_test or the testing data, these predictions will be stored in a variable called y_test_hat. For model evaluation, we will feed the y_test and y_test_hat into the accuracy_score function and store it in a variable called test_accuracy, a variable that will hold the testing accuracy of our model. We followed these steps for a variety of classification algorithm models and obtained corresponding test accuracy scores.

- **Prediction show Data poison attack:**

  In final state flask web application is developed in which trained model is used for prediction. User uploads images to website and show segmentation and prediction results.
  .

ARCHITECTURE:

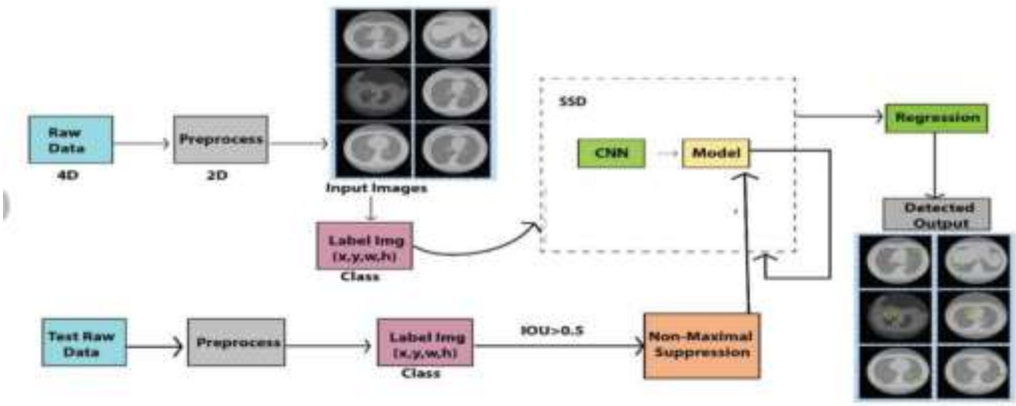


Figure 1. System Architecture

Figure 1 illustrates the framework of the proposed method The demonstration of our lung cancer detection from CT scan images raw data is collected and processed through various steps and prediction of type of cancer is shown in the result.

SVM:
Support Vector Machine (SVM) is a supervised learning model that excels in classification tasks by finding the optimal hyperplane to separate different classes, particularly effective in high-dimensional spaces.
CNN:

Convolution Neural Networks (CNN) mimic the human brain's structure, capable of learning complex patterns and features from data, making them highly effective in tasks requiring deep learning.

FLOW DIAGRAM:

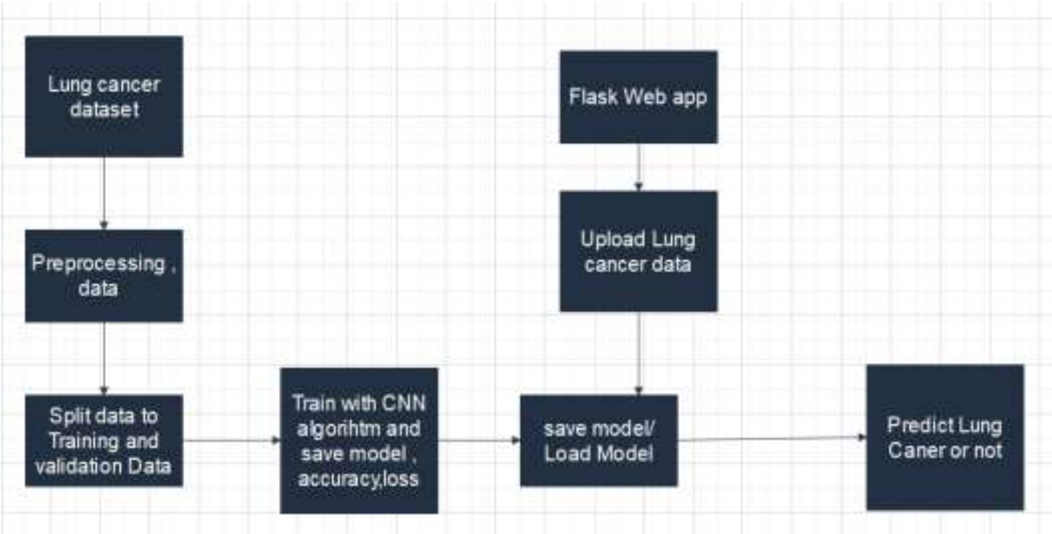


Figure 2. Model Flow Diagram

V. EVALUATION METRICS

Comparative analysis on different algorithms

Algorithm	Accuracy
SVM Algorithm	Accuracy 89 percent
CNN Algorithm	Accuracy 98 percent

COMPARISION GRAPH:

- Comparison of algorithms (CNN sequential model, SVM model)

Industrial Engineering Journal
ISSN: 0970-2555
Volume: 54, Issue 1, January:2025

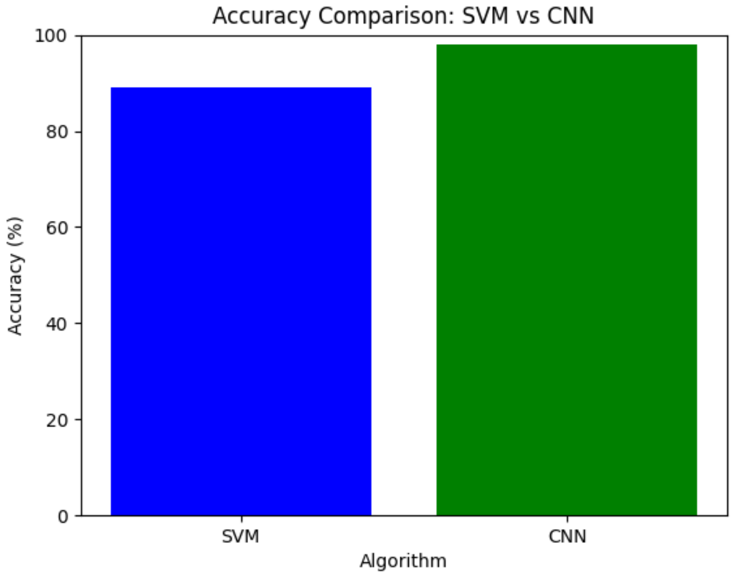


Figure: 3

Figure 3 bar graph shows accuracy comparison of various algorithms in which SVM model shows 89 percent accuracy and CNN model shows 98 percent accuracy.

Accuracy Graphs:

- Algorithm Accuracy and Loss graphs

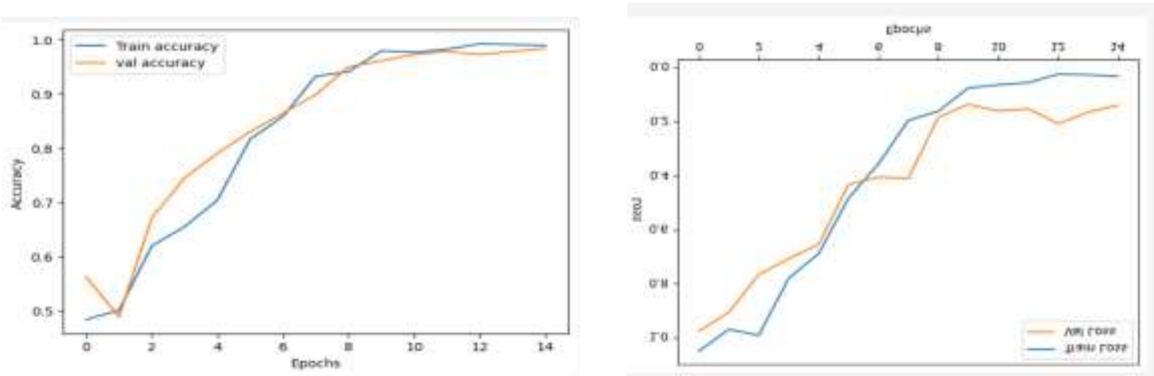
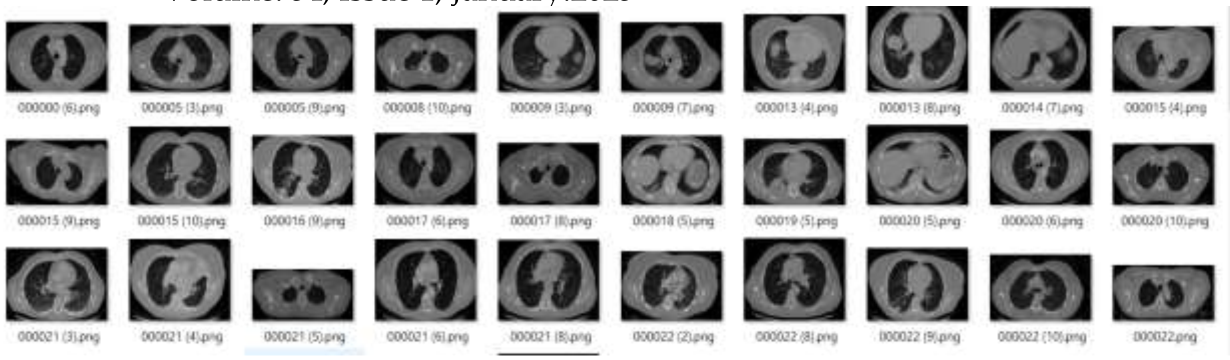


Figure 4

Figure 4 graphs shows accuracy and loss graph for both validation data and training data

RESULTS:

DATASET:

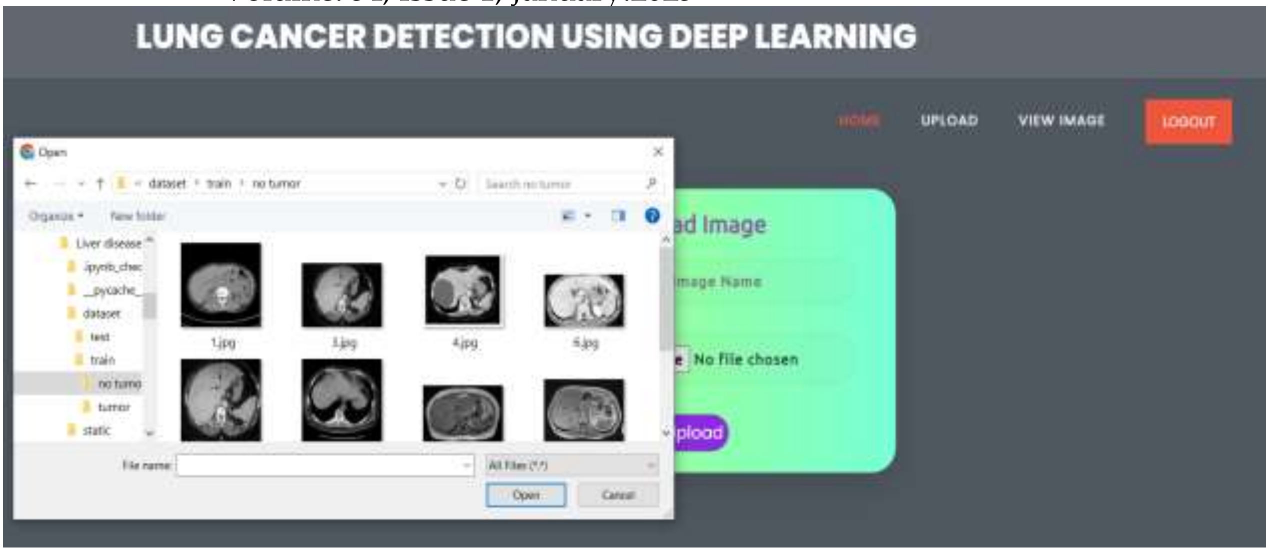


Home Page:

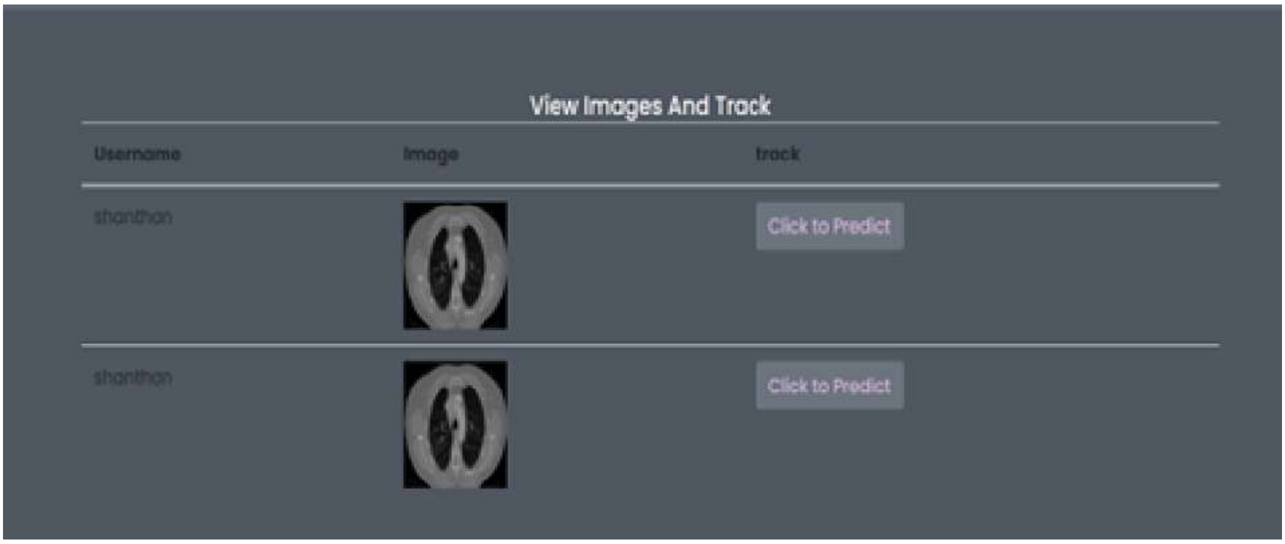
Login Form:

Upload Data:

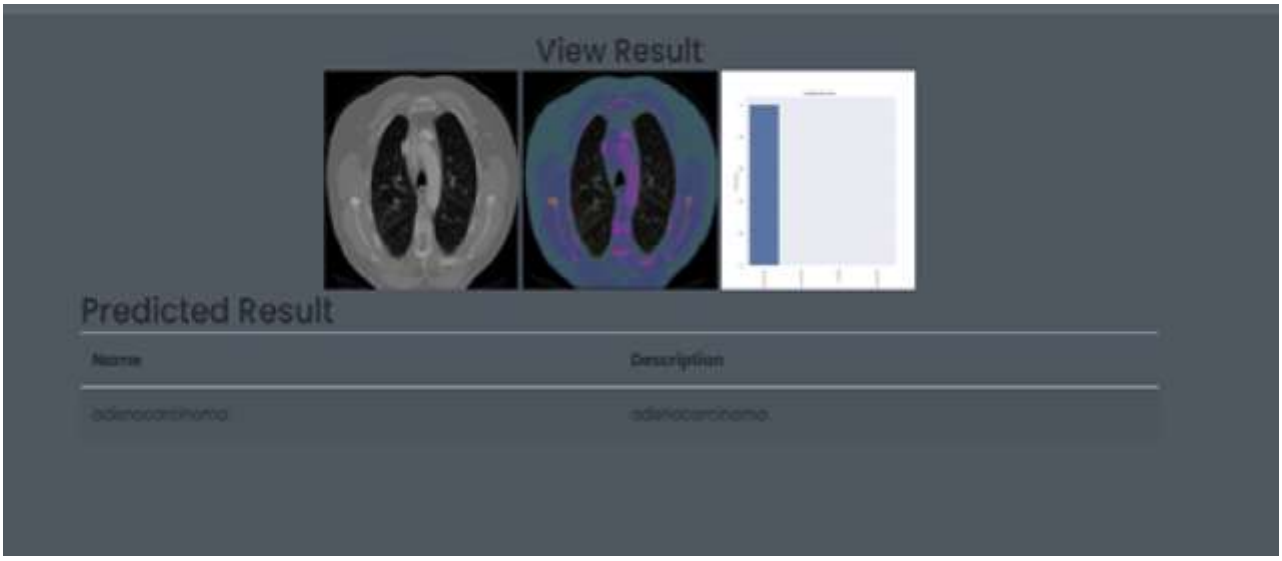
Industrial Engineering Journal
ISSN: 0970-2555
Volume: 54, Issue 1, January:2025



View Image:



Result:



VI. CONCLUSION

In conclusion, our model effectively addresses the problem of detecting lung cancer form CT images and improves accuracy of prediction. The Deep learning algorithm given accuracy of 98 percent and for image segmentation water shed algorithm is used. Using lung CT scan images as input, the proposed model can classify the output as "benign" or "malignant" using binary classification in the LIDC-IDRI dataset. With nearly 50 epochs the model accuracy approaches 98% during the training phase while it is nearly 95% during testing phase. The model loss during testing phase was below 0.3. Thus, the proposed model cannot only predict different type of lung cancers accurately but can also serve as a useful tool for medical image detection.

VII. FUTURE SCOPE

In future work for improving accuracy and segmentation RNN algorithm can be used to find out exact location of cancer form the given image. For training multiple Ensemble algorithms with deep learning can be used to improve accuracy.

VIII. REFERENCES

[1] VikulPawar, "Lung Cancer Detection System Using Image Processing and Machine Learning Techniques", International Journal of Advanced Trends in Computer Science and Engineering, Volume 9, No.4, pp. 5956-5963, 2020.

[2] Sarfaraz Hussein, PujanKandel, Candice W. Bolan, Michael B. Wallace, and UlasBagci, "Lung and Pancreatic Tumor Characterization in the Deep Learning Era: Novel Supervised and Unsupervised Learning Approaches", IEEE Transactions on Medical Imaging, Journal Article, Volume: 38, Issue: 8, pp.1777-1787, 2019.

[3] Dakhaz Mustafa Abdullah, NawzatSadiq Ahmed, "A Review of most Recent Lung Cancer Detection Techniques using Machine Learning", International Journal of Science and Business, Volume 5, issue 3, pp. 159-173,2021.

[4] Nikita Banerjee, Subhalaxmi Das, "Prediction Lung Cancer– In Machine Learning Perspective",International Conference on Computer Science, Engineering andApplications (ICCSEA), IEEE access, 2020.

[5] BinilaMariyamBoban, Rajesh KannanMegalingam, "Lung Diseases Classification based on Machine Learning Algorithms and Performance Evaluation", International Conference on Communication and Signal Processing, IEEE access, 2020.

[6] Surayya Ado Bala, ShriOjha Kant, AdamuGarbaYakasai, "Deep Learning inMedical ImagineandDrugDesign", JournalofHumanPhysiology, Volume02, Issue02,2020.

[7] JahanzaibLatif, Chuangbai Xiao, AzharImran, ShanshanTu, "Medical Imaging using Machine Learning and Deep Learning Algorithms: A Review", International Conference on Computing, Mathematics andEngineering Technologies, iCoMET 2019 IEEE access, 2019.

[8] Qing Zeng Song, Lei Zhao, XingKeLuo, XueChen Dou, "Using Deep Learning for Classification of Lung Nodules on Computed Tomography Images", Journal of Healthcare Engineering, Volume 2017, Article ID 8314740, 2017.

[9] Patrice Monkam, ShouliangQi,HeMa,WeimingGao,Yudong Yao, and Wei Qian," Detection and Classification of Pulmonary Nodules Using Convolutional Neural Networks: A Survey", IEEE Access, Volume 7, pp. 78075-78091, 2019.

[10] ShanchenPang, YaqinZhang, Mao Ding, XunWang, andXianjinXie, "ADeep Model for Lung Cancer Type Identification by Densely Connected Convolutional Networks and Adaptive Boosting" IEEE Access, Volume 8, pp. 4799-4805, 2020.

[11] Imdad Ali , Muhammad Muzammil , IhsanUlHaq ,Amir A. Khaliq and Suheel Abdullah, "Efficient Lung Nodule Classification Using Transferable Texture Convolutional Neural Network", IEEE Access, Volume 8, pp.175859- 175870,2020.

[12] Shanchen Pang, Fan Meng, Xun Wang, Jianmin Wang, Tao Song, Xingguang Wang and Xiaochun Cheng, "VGG16- T: A Novel Deep Convolutional Neural Network with Boosting to Identify Pathological Type of Lung Cancer in Early Stage by CT Images", International Journal of Computational Intelligence Systems, Vol. 13(1), pp. 771–780, 2020.

[13] Xie H, Yang D, Sun N, Chen Z, Zhang Y., "Automated Pulmonary Nodule detection in CT images using Deep Convolutional Neural Networks", Pattern Recognition, Elsevier ,85 pp.109 19, 2019.

[14] Nakrani MG, Sable GS, Shinde UB, "ResNet based Lung Nodules detection from Computed Tomography images", International Journal of Innovative Technology and Exploring Engineering, Volume-9, Issue-4, pp.1711–1714, 2020.

[15] Neal Joshua ES, Bhattacharyya D, Chakkravarthy M, Byun YC, "3D CNN with visual insights for early detection of lung cancer using Gradient-Weighted Class Activation" , Hindawi Journal of Engineering, 10.1155/2021/6695518. Volume 2021, Article ID 6695518, 2021.