International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 07 August 2023 **Revised:** 08 September 2023 **Accepted:** 15 October 2023*

An Innovative Method for Lung Cancer Identification Using Machine Learning Algorithms

Kavita Singh[1] , Usha Chauhan[2] , Lokesh Varshney[3] , A.R. Verma[4]
[1]Department of Electrical, Electronics and Communication Engineering
Galgotias University,
Uttar Pradesh, India
kavitasingh78@gmail.com
Orcid Id: 0000-0002-9728-7460
2Corresponding Author: usha.chauhan@galgotiasuniversity.edu.in
[2]Department of Electrical, Electronics and Communication Engineering
Galgotias University,
Uttar Pradesh, India
Orcid Id: 0000-0003-1359-4082
[3]Department of Electrical, Electronics and Communication Engineering
Galgotias University,
Uttar Pradesh, India
lokesh.varshney@gmail.com
Orcid Id: 0000-0001-8305-1687
[4]Department of Electronics and Communication Engineering
G.B.Pant Institute of Engineering and Technology,
Uttarakhand, India
arverma06ei03@gmail.com
Orcid Id :0000-0003-3139-7103

Abstract— Biological community and the healthcare sector have greatly benefited from technological advancements in biomedical imaging. These advantages include early cancer identification and categorization, prognostication of patients' clinical outcomes following cancer surgery, and prognostication of survival for various cancer types. Medical professionals must spend a lot of time and effort gathering, analyzing, and evaluating enormous amounts of wellness data, such as scan results. Although radiologists spend a lot of time carefully reviewing several scans, tiny nodule diagnosis is incredibly prone to inaccuracy. Low dose computed tomography (LDCT) scans are used to categorize benign (Noncancerous) and malignant (Cancerous) nodules in order to study the issue of lung cancer (LC) diagnosis. Machine learning (ML), Deep learning (DL), and Artificial intelligence (AI) applications aid in the rapid identification of a number of infectious and malignant diseases, including lung cancer, using cutting-edge convolutional neural network (CNN) and Deep CNN architectures, we propose three unique detection models in this study: SEQUENTIAL 1 (Model-1), SEQUENTIAL 2 (Model-2), and transfer learning model Visual Geometry Group, VGG 16 (Model-3). The best accuracy model and methodology that are proposedas an effective and non-invasive diagnostic tool, outperforms other models trained with similar labels using lung CT scans to identify malignant nodules. Using a standard LIDC-IDRI data set that is freely available, the deep learning models are verified. The results of the experiment show a decrease in false positives while an increase in accuracy.

Keywords- Lung Cancer; Image Segmentation; LDCT Screening; Pulmonary nodule; Machine Learning; Deep Learning

I. INTRODUCTION

Accurately diagnosing malignant nodules is of critical importance in improving the prognosis of lung cancer through the use of intelligent data analytics in healthcare, which significantly increases survival rates [1-4]. Access to accurate, complete, and timely relevant lung cancer data is essential for investigating the causes of lung cancer, detecting cancer early, evaluating the efficacy of treatment, identifying the causes of treatment failure, and conducting lung cancer control programs. Medical professionals face challenges in making decisions when presented with large amounts of data points. Intelligent data analysis in healthcare, particularly using AI and ML algorithms, can benefit medical professionals and patients in detecting, classifying, predicting clinical results post-cancer surgery, and predicting survival rates for various cancer types, including lung cancer [5-9]. Lung cancer patients are particularly vulnerable to COVID-19 and its associated consequences due to their compromised immune systems. Diagnosis, therapy, and post-operative care for lung cancer cases are extremely challenging

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 07 August 2023 **Revised:** 08 September 2023 **Accepted:** 15 October 2023*

during an epidemic like COVID-19 [10-14]. Modern AI or ML processes can potentially advance the treatment and care of lung cancer patients during the pandemic by utilizing cutting-edge technologies. By combining experimental and tomographic data through ML techniques, researchers can distinguish between lung abnormalities caused by the Coronavirus and those caused by other factors like immunotherapy and radiation emission. The ISBI 2018 LN Malignancy Prediction Challenge was developed by the quantitative imaging network (NCIQIN) team of the National Cancer Institute (NCI) with the goal of predicting LN Malignancy through two consecutive LDCT screening exams using automated techniques [15-16].

To improve early-stage classification and diagnosis of lung nodules, an effective deep learning-based method is recommended. Medical informatics primarily focuses on analysing large-scale health data systems to enhance clinical decision support systems and assess clinical records for quality standards and usability of medical services. Automatic medical image analysis is crucial in modern medicine as diagnostic procedures relying on image interpretation can be highly subjective. Computer-aided diagnosis (CAD) can provide a neutral assessment of core clinical circumstances, improving patient care when combined with trained clinicians' expertise. To distinguish between noncancerous and cancerous lung nodules of small size, researchers have recommended proteomic classifiers and pulmonary nodule traits. This has led to significant improvements in the size of nodules between 8 and 30 mm, reducing the need for benign nodule biopsies. Various diagnostic techniques are used by medical examiners to correlate with the early detection of malignant nodules, including clinical trials, LDCT screening analysis, Positron Emission Tomography (PET), and tissue removal investigation. Invasive techniques, such as surgery, are frequently used to distinguish between benign and cancerous lung nodules. However, such procedures are risky and may increase patient anxiety [17]. The COVID-19 pandemic has highlighted significant healthcare challenges, including the sharing and connecting of biomedical data from various clinical data sources, increasing the accuracy of clinical data, and reducing structural heterogeneity of diagnostic data to increase research trial sample sizes. Computational modelling can be used to address healthcare needs and requirements, such as building trustworthy ML models for long-term disease prevention, identifying biomarkers, and developing cost-effective medications. Improving patient medical outcomes in lung cancer necessitates early identification and detection. However, detecting malignant nodules at an early stage is challenging due to the time and effort required to screen numerous scans. The development of convolutional neural networks (CNN) and graphic processing units (GPUs) has substantially improved the performance of computer-based decision support systems and image analysis for identifying and classifying lung nodules. In summary, accurate and early diagnosis of lung cancer using intelligent data analytics and AI/ML algorithms can significantly improve patient outcomes and survival rates. The combination of medical expertise and cutting-edge technologies can lead to more effective treatments and care for lung cancer patients, even during challenging times like the COVID-19 pandemic.

II. RELATED WORKS

The authors of the research [18] utilized a 3D CT scan analysis with histogram-weighted sampling to improve performance. They projected a 3D fully connected CNN (FCCNN) to reduce false positives (FPs) in the classification of lung nodules (LNs). Additionally, they combined Gradient Boosting Machine (GBM) with 3D Dual Path Network (DPN) that includes 3D faster Region-based CNN [19] to develop an automated model for identifying and classifying pulmonary lung nodules. The model achieved an error rate of 12.5% and a detection accuracy of 87.5%. Another study [20] implemented a 3D FRCNN model for LN detection, which demonstrated exceptional object detection capabilities using DPN and CNN's advanced version architecture. For group-based lung nodule detection [21], a multi-patch technique using a Frangi filter was developed to enhance performance. Their CAD system achieved a sensitivity of 94% with an FP rate of 15.1% and a sensitivity of 80.06% with an FP rate of 4.4%. In another research [22], a deep fully connected convolutional neural network (DFCCNN) was presented for the identification and classification of lung nodules based on computed tomography imaging. The model used medical IoT data to classify nodules as benign or malignant and further divided them into four sub-classes. The use of 3D Deep CNN [23] with multi-scale prediction methods for segmented image LN identification showed promising results, outperforming 2D CNNs due to its ability to capture more features. The concept of patient-specific healthcare gained interest due to its potential for improved patient care and reduced healthcare costs. Technologies for combining genetic, medical, and imaging data, as well as algorithms for pattern identification, have been developed for disease prediction and treatment [24]. Transfer learning (TL) offers solutions for classification, regression, and clustering problems, and pre-trained models like VGG16 and DCNN are utilized for image classification [25]. The 3D-CNN analysis by the authors helped reduce the risk of false positive results in automated lung nodule diagnosis [26-27]. CNNs have proven vital in digital image processing, and connectivity designs like ResNet, DenseNet, and DPN are showing superior performance among deep CNN topologies [28]. Other mathematical models, such as Mix Network (MixNet) [29] and Adaptive Hierarchical Heuristic Mathematical Model (AHHMM) [30], have also been

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 07 August 2023 **Revised:** 08 September 2023 **Accepted:** 15 October 2023*

successfully used for lung nodule identification and classification. Non-Small Cell Lung Cancer (NSCLC) is classified into four main levels based on primary tumor location, size, number of lymph nodes, and distant metastases [31-32]. Soft computing techniques are required for accurate and automated cancer staging using CT and PET-CT scans to determine prognosis and treatment options.

III. LUNG NODULE CLASSIFICATION USING SOFT COMPUTING ALGORITHMS

The early detection of nodules can improve a patient's chance of survival, and knowing the cancer stages can facilitate appropriate care. However, due to the large size of CT images, radiologists find it challenging to diagnose images accurately, especially when nodules are smaller than anticipated. Factors such as density, size, position, and shape impact nodule visibility, with lesions under 5 mm being difficult to anticipate, potentially leading to missed lesions in LDCT scans. Previous studies focused on larger nodules, identified by a small number of skilled medical professionals experienced in interpreting screening tests. Less experienced radiologists often overlook nodules smaller than 6 mm, leading to a high false positive rate and inaccurate data for lung cancer prognosis. To address this challenge, the watershed method is applied to partition LDCT scan images after pre-processing. A precise lung cancer classification model based on the VGG 16 NET24 transfer learning model is constructed, incorporating the watershed algorithm for cancer cell removal and masking. The integration of trained and optimized models creates the final lung cancer classification model, requiring a systematic technique and framework for accurate identification. After the radiomics technique, which relies on feature extraction, CNNs serve as a backup for classifying LNs from 2D or 3D images. The segmentation and feature extraction might not always be necessary, the CNN still requires the right approach and a sizable dataset for accurate tumor categorization. Fig. 1. the radiomics technique for LN classification involves five main steps: image acquisition from a specific CT scanner, image preparation to reduce noise, tumor segmentation from the image's ROI using various techniques, extraction of properties like texture and gradient from the segmented image, and classification of tumors as benign or malignant based on these features. In the CNN model, convolutional layers (CL), including Soft-max, pooling, and fully connected layers, process input data in Fig. 2. CLs extract low-level properties, while ensuing layers extract semantic data. The convolution layer generates an output by sliding a kernel over the input, introducing bias, and applying a non-linear activation function like ReLU. The pooling layer reduces image size while retaining vital information. The combination of convolution and pooling layers extracts prominent features from the image, forming a 1D vector sent to a fully linked network with multiple hidden layers and non-linear activation functions to improve performance. In the end, the item's activation function relies on probabilities between 0 and 1, such as sigmoid or soft-max, for accurate results.

Figure 1. Basic Classification Procedure

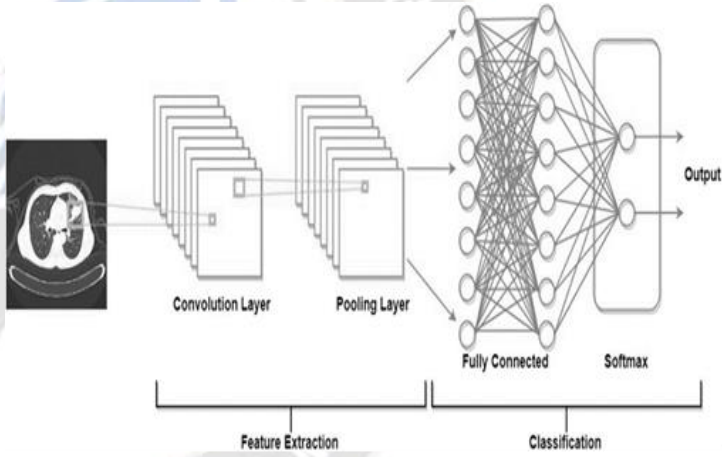


Figure 2. Basic CNN model for lung cancer

A. Data Acquisition

During multiple study phases, heterogeneous CT images of lung nodules were collected and used as the input layer for the dataset. These images underwent filtering based on various criteria. The dataset was generated from LUNA16. To ensure a low likelihood of lung nodules, only very thin segments were retrieved, and scans with segment thickness over 2.5 mm were excluded. Out of 888 CT scans, 36,378 radiologists' annotations were obtained, with images containing missing slices or erroneous segment spacing discarded. The focus was on annotations labeled as "pulmonary nodules under 3 mm," which were considered crucial for technical analysis in predicting lung cancer, as both cancerous and non-cancerous nodules under 3 mm were included in LUNA16. Annotations of nodules observed and noted by different readers were combined when

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 07 August 2023 **Revised:** 08 September 2023 **Accepted:** 15 October 2023*

they had similar qualities, except for the convolution of their radii. This resulted in sets of 2290, 1602, 1186, and 777 nodules, each annotated by 1, 2, 3, or 4 radiologists, respectively. Visualization of the dataset by the training model played a significant role in better understanding the dataset. To address the challenge of viewing low-dose computed tomographic images on a standard computer, the Python package for Digital Imaging and Communications in Medicine (PyDICOM) library was utilized. The metadata records in the CT images provided various details, including case name, case id, case delivery due date, image location, scan number, examiner's name, and more, all available in the picture array. The entire subfolder consisted of 180 two-dimensional picture slices, forming three-dimensional images of the lungs and being assigned an image indexing number. Fig. 3. displayed several slices from a LUNA16 CT scan. 3D medical imaging utilizes imaging modalities in three dimensions to create improved images of the body's interior for scientific study. By combining and integrating all accessible images, a three-dimensional image model, as shown in Fig. 4., was produced within a single subdirectory. This 3D model allowed for a more in-depth study of lung cancer cells and other lung illnesses through imaging of infected lungs.

B. Segmentation Algorithms

Use Image segmentation is a crucial step in medical imaging as it allows for the extraction of regions of interest (ROI) in an image. It finds applications in various clinical tasks, including border identification, cancer detection or division, and mass detection, where body parts and tissues need to be separated based on specific criteria. Fig. 5 lung cancer detection, several segmentation approaches are used, including thresholding, edge-based segmentation, and the Watershed Algorithm. The threshold approach is a simple technique that requires no prior knowledge to calculate threshold levels. It uses histogram peaks of the image for segmentation, but it may have drawbacks due to its reliance on peaks and disregard for spatial details. Edge-based approaches, on the other hand, rely on detecting discontinuities in the image and are beneficial for images with clear object contrast. However, they may not perform well on images with inaccurate or excessive edges. The Watershed Algorithm, based on topological interpretation, is chosen for this study due to its reliable results and ability to continuously discover boundaries at the expense of complex gradient computations. The Watershed technique is particularly useful for segmenting complex images where basic thresholding and contour detection may not yield accurate results. By performing a "watershed" transition on a gray-scale image, different objects in the image can be separated. The technique requires precise background and foreground data, which are used as markers to determine the exact boundaries. This helps identify touching and adjacent objects in the images. The watershed algorithm effectively removes background noise from the image and identifies lung and cancer cells as black pixels. After removing the outer layer using the internal marker and newly constructed outline, the lung filter is generated using bitwise or Numerical Python (NumPy) operations in Fig.6. This combination of watershed methods and the Sobel filter enhances segmentation and removes the outer layer of the lungs. The heart is also removed from CT scan images, and the lung filter is closed off using morphological techniques and gradients. To generate images with related labels, approximately 1002 images with CT scan data from 12 patients, equally composed of cancer and non-cancer patients, are processed using this approach, resulting in better segmented lungs. Image processing techniques and deep learning approaches have been developed to predict cancer malignancy levels, including dimensionality reduction to reduce pixel count while preserving the most important image features. Feature extraction techniques are used, such as horizontal, vertical, and diagonal axes of motion for neural network recognition rate percentage calculation using backward-forward propagation of data.

IV. PROPOSED MODELS

The proposed approach involved the development of CNN-based models, which underwent an initial pre-processing phase using the luna16 dataset for lung segmentation based on low dose computed tomography scans. The subsequent step included lung separation using the watershed algorithm, a semantic segmentation technique that highlights specific lung regions and generates binary masks. Among three distinct CNN models considered, the most optimal candidates were selected through comprehensive assessments across various test datasets and classification tasks. The first model, referred to as SEQUENTIAL 1 (SEQ1), achieved impressive outcomes in number classification by employing a straightforward approach involving convolution, flattening embedding, up-sampling, and dropout in intermediate layers. The second model, SEQUENTIAL 2 (SEQ2), adopted deep convolutional neural network (DCNN) architecture with maximum assembling and fully connected (FC) layers. This model, characterized by a predefined number of layers and components, demonstrated superior performance, consistent with findings from diverse research articles employing different datasets. The third and final model approach leveraged transfer learning from the VGG16 architecture, with slight modifications to the final three fully

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 07 August 2023 **Revised:** 08 September 2023 **Accepted:** 15 October 2023*

connected layers. This model showcased noteworthy capabilities in object categorization. Model training involved binary lung mask segmentation, with 32-image batches processed per iteration and 100 images per epoch for a total of 30 epochs (except for Model-3, which used 500 images per epoch). While Models 1 and 2 were trained on grayscale images with dimensions (512, 512, 1), Model 3 utilized images with a shape parameter of (512, 512, 3). To enhance performance, Model 3 underwent training with augmented parameter sets, incorporating operations such as shear, zoom, horizontal flipping, and shifts to yield improved results. The model incorporated accuracy measures, a learning rate of 0.0001, binary cross-entropy, and an efficient Adam optimizer. In the final stage of binary classification, a single node was chosen in the output layer to effectively distinguish between lung cancer and normal lung cases. The training process was supported by Ker as call backs, facilitating a 50-epoch training session and enabling graphical comparisons to assess model performance. The inclusion of these call backs aided in saving the model with the highest accuracy achieved during training. Notably, accuracy exhibited significant improvement as the number of epochs increased.

V. RESULTS

The results of the conducted experiments are showcased in Figures 7 through Figure 12, alongside accuracy and loss computations over epochs for each model trained on the lung cancer dataset. Figure 13 illustrates a comparison between the accuracy and losses of the suggested models. The call back function is designed to retain data with the highest accuracy for each model. The ensuing observations from this graph are as follows: Model-1 performed the least favourably among all models, succumbing to overtraining, which hindered the attainment of reasonable validation accuracy. This was while it displayed the lowest validation loss, as depicted in Figures 8, Figure 10, and Figure 12. Notably, accuracy is of paramount importance, given the classification nature of the task. A clear pattern emerges from Figures 7, 9, and 11, which correspond to the training and validation accuracy of Models 1, 2, and 3, respectively. It is evident that Models 2 and 3 excelled during training, achieving commendable levels of test accuracy and loss. Notably, Model 2 outperformed Model 3 in accuracy while also experiencing a smaller loss. Figure 14 offers insights into the contributions of different models, highlighting the best accuracy model and the proposed methodology as a potent, non-invasive diagnostic tool. This tool showcases superior performance compared to other models trained on similar labels, leveraging lung CT scans for the detection of malignant nodules. The later stages of the investigation encountered several challenges that influenced the outcomes: Dataset Format: The referenced dataset was in the (.dcm) format, incongruent with formats commonly used in image processing, such as (.jpg), among others. In this context, each selected image underwent a two-dimensional NumPy transformation to acquire values (n, m, m, l), without relying on transfer learning models. Handling float64 data, especially with tools like matplotlib, and Open Source Computer Vision Library (cv2), posed significant challenges. Tensor Flow and Ker as don't accept images in the (n, m, m) format, rendering the transformation into two-dimensional grayscale images a complex process. This transformation's limitations could lead to errors or loss of crucial features required for lung nodule categorization. A visual depiction of this pre-processing, illustrated in Figures 15 to 19, reveals the effects of various conversions and image plotting. The changing radiological characteristics of lungs, including shape, area, parameters, and opacity, are evident in these images. Transforming the plotted images to the standard format (512, 512, 3) in int8 or int64 could potentially enhance the model's accuracy, surpassing 95%, while adhering to known linear trends in the generated plots. Resources and Hardware Constraints: The study encountered limitations in terms of available resources and processing hardware. Sourcing the dataset proved challenging and suitable processing hardware was also scarce. The study heavily relied on Google Colab, which imposed a maximum runtime of 12 hours, followed by a waiting period of 9 to 12 hours. These time constraints slowed down progress. Despite having 12GB of processing memory, training small models on over 1200 RGB or grayscale images became intricate when assembling data into a numpy array. Furthermore, the necessity to reduce batch sizes had ramifications beyond the scope of this research topic, impacting the model's performance.

VI. CONCLUSION

The recent upheaval caused by the devastating impact of COVID-19 on our society, there exists an opportunity to harness state-of-the-art AI technology. By applying Machine Learning (ML) and Artificial Intelligence (AI) to the realm of lung imaging, we can enhance the precision of identifying Incidental Pulmonary Nodules (IPNs) within primary healthcare settings, particularly for COVID-19 Lung Cancer (LC) prediction using Chest X-ray (CXR) and CT scans. This advancement has the potential to alter the timing of LC detection, potentially leading to the identification of the disease at earlier stages. The significance of this lies in the fact that many malignant tumors experience a reduction in size when detected in their initial phases, consequently extending the patient's life expectancy. Remarkably, scant literature addresses the identification of nodules smaller than 3 mm.

The utilization of private datasets by certain researchers has introduced considerable complexity in comparing outcomes with results derived from alternative algorithms. This challenge

underscores the need for establishing a universal dataset accessible to all academic circles. In the pursuit of a comprehensive LC diagnosis, effective data analysis unequivocally demonstrates the superiority of our approach over prior models in categorizing Lung Nodules (LNs) as malignant. This includes surpassing the accuracy of Support Vector Machine (SVM) at 93.2 % and Dilated Residual Network at 85.7 %. Our method achieves an impressive classification accuracy of 94.07 %.

To address the limitations of both invasive and non-invasive cancer detection techniques and to differentiate between benign and malignant nodules during the early stages of the disease based on their morphology, advancements in Deep Learning (DL) approaches are imperative. It is advisable for healthcare institutions to embrace interdisciplinary strategies, incorporating intelligent data analytics to tackle diagnostic challenges. This integration of intelligent data analytics within healthcare holds immense potential to revolutionize the domains of treatment, prevention, and patient care for lung cancer. Furthermore, it can facilitate the formulation and implementation of clinical recommendations, cost reduction, enhancement of patient diagnosis and treatment, and the comprehensive assessment of clinical outcomes.

ACKNOWLEDGMENT

The authors are grateful to the School of Electrical, Electronics and Communication, Galgotias University for granting the assets and details utilized in this paper.

REFERENCES

[1] Balagurunathan Y., Beers A., Mcnitt-Gray M., Hadjiiski L., Napel S., Goldgof D., Perez G., Arbelaez P., Mehrtash A., Kapur T., Yang E., Moon J.W., Perez G.B., Delgado-Gonzalo R., Farhangi M.M., Amini A.A., Ni R., Feng X., Bagari A., Vaidhya K., 2021. Lung Nodule Malignancy Prediction in Sequential CT Scans., IEEE Transactions on Medical Imaging.; 40(12), pp.3748–3761.

[2] Kavita Singh, Usha Chauhan & Lokesh Varshney. 2023. Impact of covid-19 in lung cancer detection using image processing techniques, artificial intelligence and machine learning approaches. Smart Science, 11(4), 728-743, DOI: 10.1080/23080477.2023.2246285

[3] Ahmed I, Chehri A, Jeon G, Piccialli F. 2022. Automated Pulmonary Nodule Classification and Detection Using Deep Learning Architectures. IEEE/ACM transactions on computational biology and bioinformatics., 24(4), pp. 2319-2331.

[4] Ravi D, Wong C, Deligianni F, Berthelot M, Andreu-Perez J, Lo B, Yang G. 2017. Deep Learning for Health Informatics. IEEE Journal of Biomedical and Health Informatics.; 21(1), pp.4–21.

[5] Tanveer M, Lin CT, Singh A., 2023, Guest Editorial Advanced Machine Learning Algorithms for Biomedical Data and Imaging—Part II. IEEE Journal of Biomedical and Health Informatics., 27(1), pp. 188–189.

[6] Xie L, Qi J, Pan L, Wali S., 2020. Integrating deep convolutional neural networks with marker-controlled watershed for overlapping nuclei segmentation in histopathology images. Neurocomputing, 376, pp.166–179.

[7] Silvestri G.A, Tanner N.T, Kearney P, Vachani A, Massion P.P, Porter A, Springmeyer S.C, Fang K.C, Midthun D., Mazzone P.J, 2018. PANOPTIC Trial Team. Assessment of Plasma Proteomics Biomarker's Ability to Distinguish Benign From Malignant Lung Nodules: Results of the PANOPTIC (Pulmonary Nodule Plasma Proteomic Classifier) Trial. Chest. 154(3), pp. 491–500.

[8] Zhao J, Ji G, Qiang Y, Han X, Pei B, Shi Z. A New Method of Detecting Pulmonary Nodules with PET/CT Based on an Improved Watershed Algorithm. PLOS ONE., 10(4), doi:10.1371/journal.pone.0123694.

[9] MacMahon H, Naidich D.P, Goo J.M, Lee K.S, Leung A.N.C, Mayo J.R, Mehta A.C, Ohno Y, Powell CA, Prokop M, Rubin G.D, Schaefer-Prokop C.M, Travis WD, Van Schil PE, Bankier AA. 2017. Guidelines for Management of Incidental Pulmonary Nodules Detected on CT Images: From the Fleischner Society. Radiology., 284(1), pp. 228–243

[10] Akay M, Subramaniam S, Brennan C, Bonato P, Waits CMK, Wheeler BC, Fotiadis DI., 2022. Healthcare Innovations to Address the Challenges of the COVID-19 Pandemic. IEEE Journal of Biomedical and Health Informatics. 26(7), pp. 3294–3302.

[11] Pezoulas VC, Kourou KD, Kalatzis F, Exarchos TP, Zampeli E, Gandolfo S, Goules A, Baldini C, Skopouli F, De Vita S, Tzioufas AG, Fotiadis DI. , 2020. Overcoming the Barriers That Obscure the Interlinking and Analysis of Clinical Data Through Harmonization and Incremental Learning. IEEE Open Journal of Engineering in Medicine and Biology. 1, pp. 83–90.

[12] Wormanns D, Ludwig K, Beyer F, Heindel W, Diederich S., 2004. Detection of pulmonary nodules at multirow-detector CT: effectiveness of double reading to improve sensitivity at standard-dose and low-dose chest CT. European Radiology. , 15(1) pp.14–22.

[13] Demir Ö, Yılmaz Ç. A. , 2015. Computer-aided detection of lung nodules using outer surface features. Bio-Medical Materials and Engineering., 26, :pp. 1213–222.

[14] Li Z, Zhang J, Tan T, Teng X, Sun X, Zhao H, Liu L, Xiao Y, Lee B, Li Y, Zhang Q, Sun S, Zheng Y, Yan J, Li N, Hong Y, Ko J, Jung H, Liu Y, Chen Y., 2021. Deep Learning Methods for Lung Cancer Segmentation in Whole-Slide Histopathology Images—The ACDC@LungHP Challenge 2019. IEEE Journal of Biomedical and Health Informatics., 25(2), pp.429–440.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[15] Harsono IW, Liawatimena S, Cenggoro TW., 2022. Lung nodule detection and classification from Thorax CT-scan using RetinaNet with transfer learning. Journal of King Saud University - Computer and Information Sciences., 34(3), pp.567–577.

[16] Al M.B., Brennan P.C, Mello-Thoms C., 2017. A review of lung cancer screening and the role of computer-aided detection. Clinical Radiology., 72(6), pp.433–442.

[17] LI S, LIU D., 2021. Automated classification of solitary pulmonary nodules using convolutional neural network based on transfer learning strategy. Journal of Mechanics in Medicine and Biology., 21(05), doi:10.1142/s0219519421400029.

[18] Setio AAA, Traverso A, De Bel T, Berens MSN, Bogaard C Van den, Cerello P, Chen H, Dou Q, Fantacci ME, Geurts B, Gugten R van der, Heng PA, Jansen B, de Kaste MMJ, Kotov V, Lin JY-H, Manders JTMC, Sóñora-Mengana A, García-Naranjo JC, Papavasileiou E., 2017. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge. Medical image analysis. 42, pp. 1–13.

[19] Zhang H, Zhang W, Wang S, Kuang X, Zhu Y., 2022. Deep 3D multi-scale dual path network for automatic lung nodule classification. International Journal of Biomedical Engineering and Technology., 39(2), pp.149.

[20] Ren S, He K, Girshick R, Sun J., 2017. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence., 39(6), pp.1137–1149.

[21] Jiang H, Ma H, Qian W, Gao M, Li Y, Hongyang Jiang, He Ma, Wei Qian, Mengdi Gao, Yan Li ., 2018. An Automatic Detection System of Lung Nodule Based on Multigroup Patch-Based Deep Learning Network. IEEE journal of biomedical and health informatics., 22(4), pp.1227–1237.

[22] Masood A, Sheng B, Li P, Hou X, Wei X, Qin J, Feng D., 2018. Computer-Assisted Decision Support System in Pulmonary Cancer detection and stage classification on CT images. Journal of Biomedical Informatics., 79, pp.117–128.

[23] Gu Y, Lu X, Yang L, Zhang B, Yu D, Zhao Y, Gao L, Wu L, Zhou T., 2018. Automatic lung nodule detection using a 3D deep convolutional neural network combined with a multi-scale prediction strategy in chest CTs. Computers in Biology and Medicine. vol.103, pp.220–231.

[24] Akay M, Exarchos TP, Fotiadis DI, Nikita KS., 2012. Emerging Technologies for Patient-Specific Healthcare. IEEE Transactions on Information Technology in Biomedicine., 16(2), pp.185–189.

[25] Tammina S., 2019. Transfer learning using VGG-16 with Deep Convolutional Neural Network for Classifying Images. International Journal of Scientific and Research Publications. Vol.(10), pp.9420

[26] Xiao Z, Du N, Geng L, Zhang F, Wu J, Liu Y., 2019. Multi-Scale Heterogeneous 3D CNN for False-Positive Reduction in Pulmonary Nodule Detection, Based on Chest CT Images. Applied Sciences. , 9(16), pp.32-61.

[27] Zhang H, Peng Y, Guo Y., 2022. Pulmonary nodules detection based on multi-scale attention networks. Scientific Reports. , 12(1), pp.14-66.

[28] Suresh S, Mohan S., 2020. ROI-based feature learning for efficient true positive prediction using convolutional neural network for lung cancer diagnosis. Neural Computing and Applications., 32, pp.15989-16009.

[29] Gu H, Gan W, Zhang C, Feng A, Wang H, Huang Y, Chen H, Shao Y, Duan Y, Xu Z., 2021. A 2D–3D hybrid convolutional neural network for lung lobe auto-segmentation on standard slice thickness computed tomography of patients receiving radiotherapy. BioMedical Engineering OnLine., 20(1), pp.94.

[30] Yu H, Zhou Z, Wang Q. 2020. Deep Learning Assisted Predict of Lung Cancer on Computed Tomography Images Using the Adaptive Hierarchical Heuristic Mathematical Model. IEEE Access., 8, pp. 86400–86410.

[31] Detterbeck FC, Postmus PE, Tanoue LT., 2013. The Stage Classification of Lung Cancer. Chest., 143(5), doi:10.1378/chest.12-2354.

[32] Detterbeck FC., 2018. The eighth edition TNM stage classification for lung cancer: What does it mean on main street? The Journal of Thoracic and Cardiovascular Surgery., 155(1), pp.356–359.

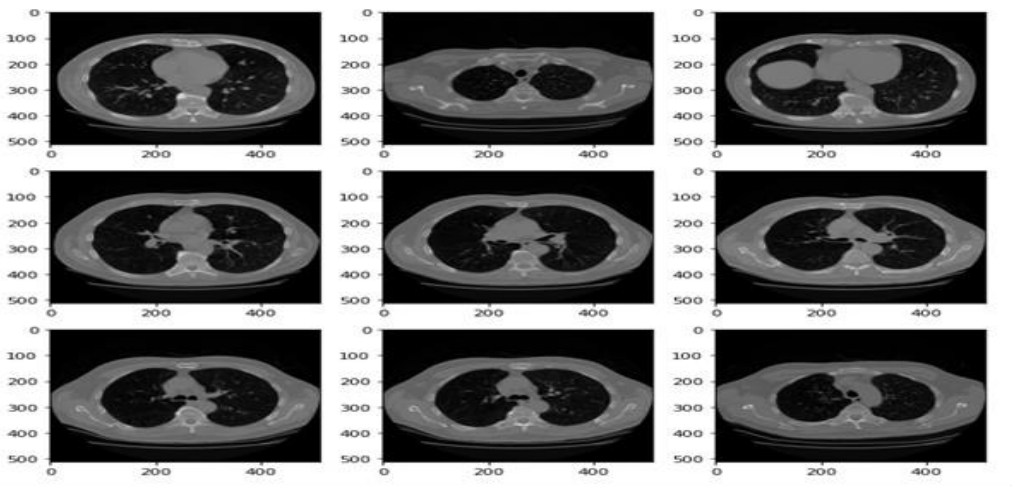


Figure 3. Original DICOM slices from LUNA16 dataset

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *07 August 2023* ***Revised:*** *08 September 2023* ***Accepted:*** *15 October 2023*

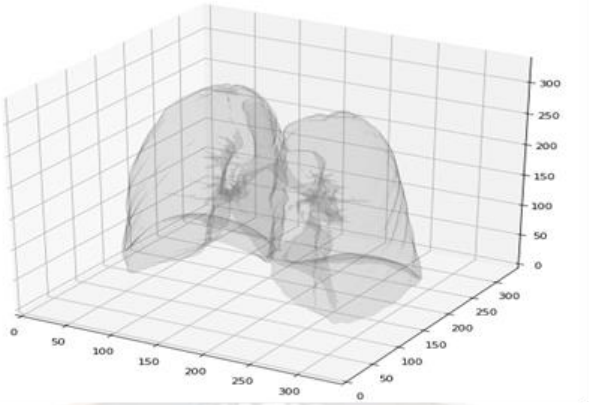


Figure 4. An image of lungs in three-dimension-single patient

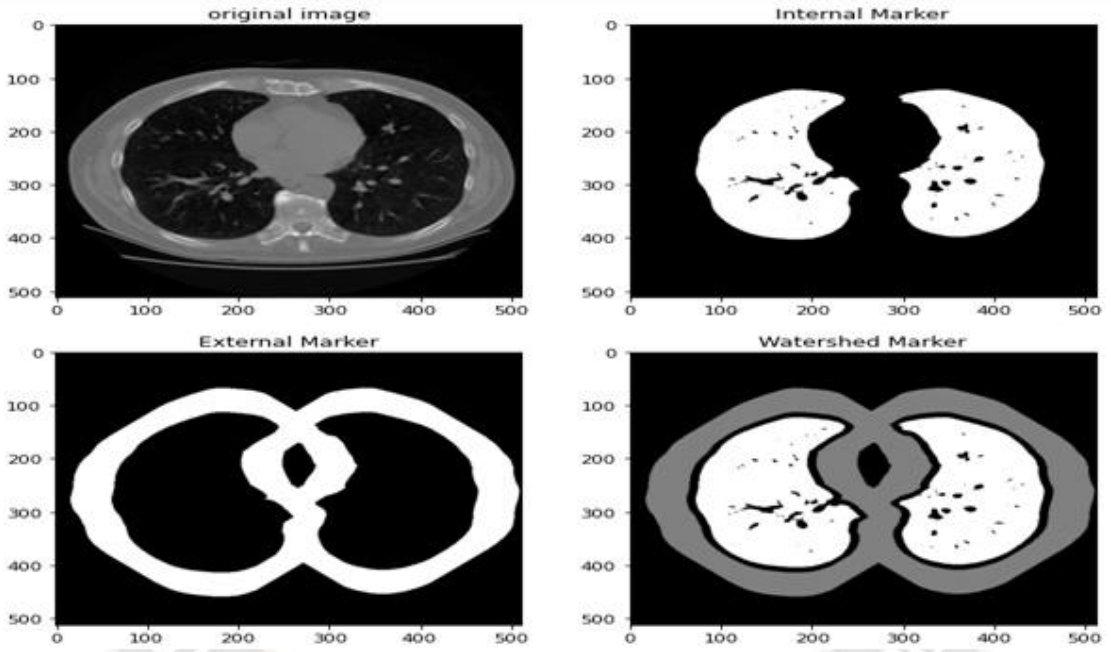
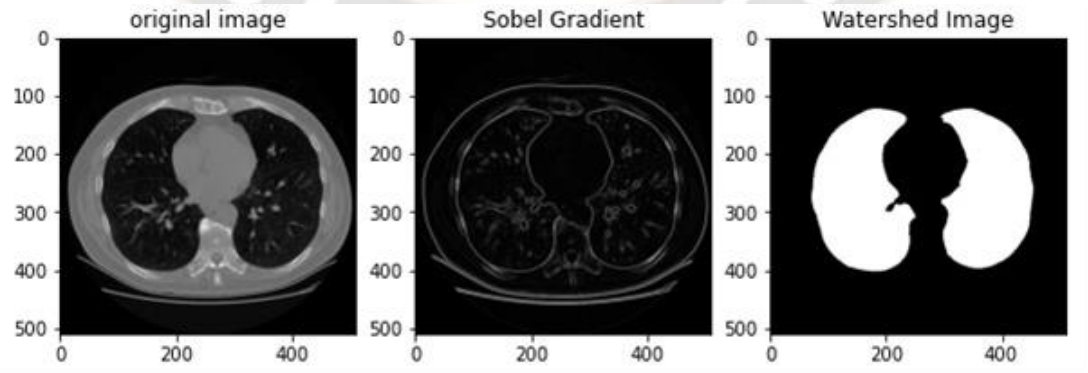


Figure 5. Effect of Water shed markers in CT scan image



International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *07 August 2023* ***Revised:*** *08 September 2023* ***Accepted:*** *15 October 2023*

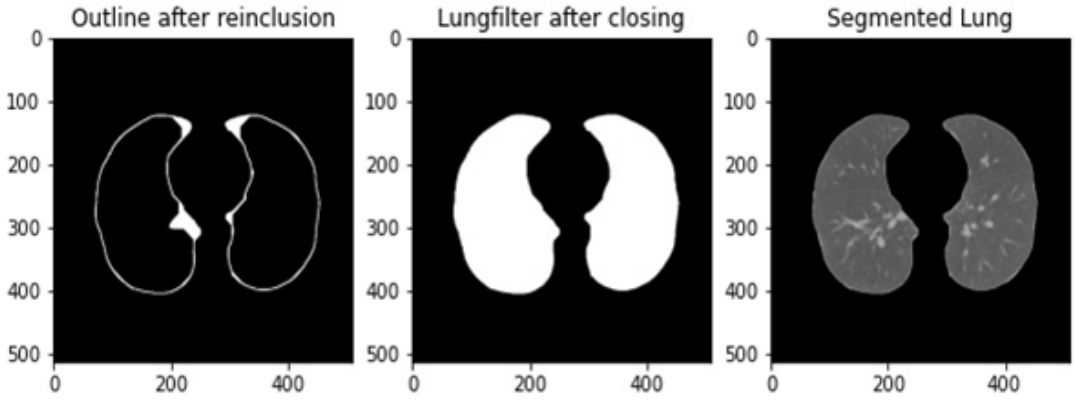


Figure 6. Image segmentation process visualization

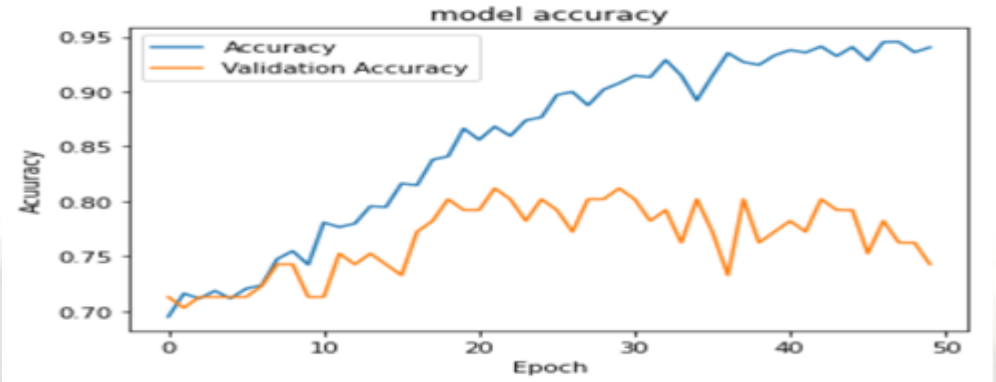


Figure 7. Training and Validation Accuracy of Model-1

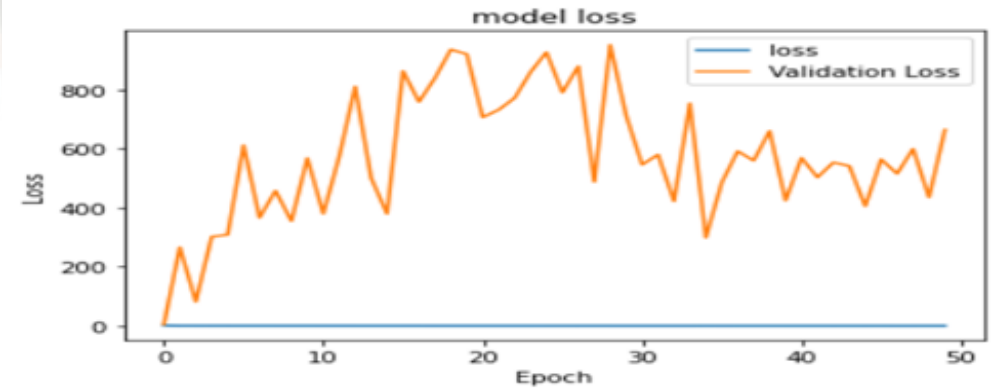


Figure 8. Training and Validation Loss of Model-1

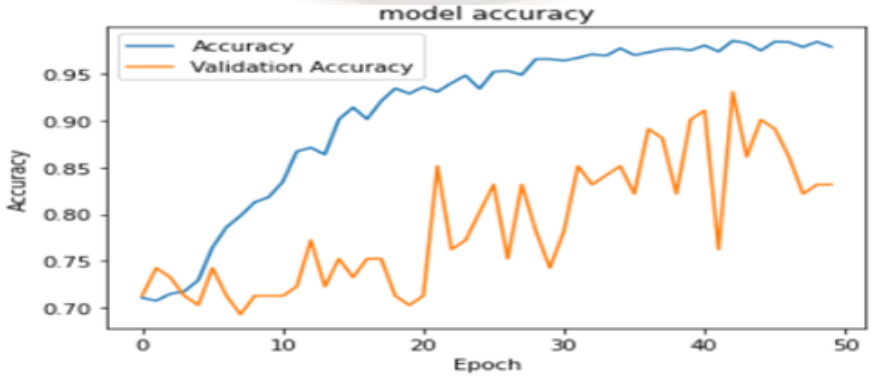


Figure 9. Training and Validation Accuracy of Model-2

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *07 August 2023* ***Revised:*** *08 September 2023* ***Accepted:*** *15 October 2023*

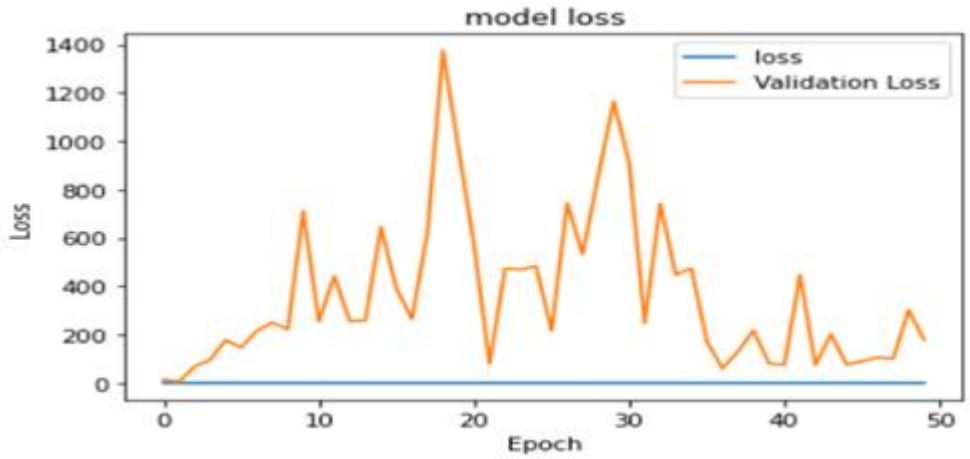


Figure 10. Training and Validation Loss of Model-2

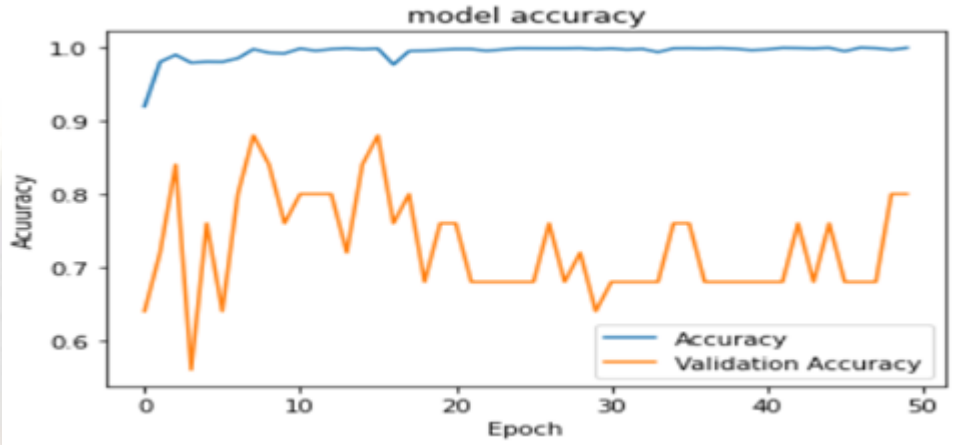


Figure 11. Training and Validation Accuracy of Model-3

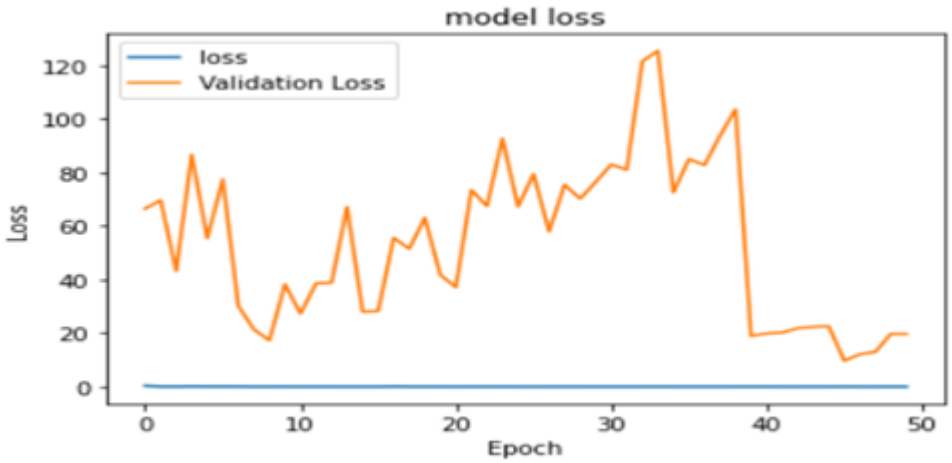


Figure 12. Training and Validation Loss of Model-3

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *07 August 2023* ***Revised:*** *08 September 2023* ***Accepted:*** *15 October 2023*

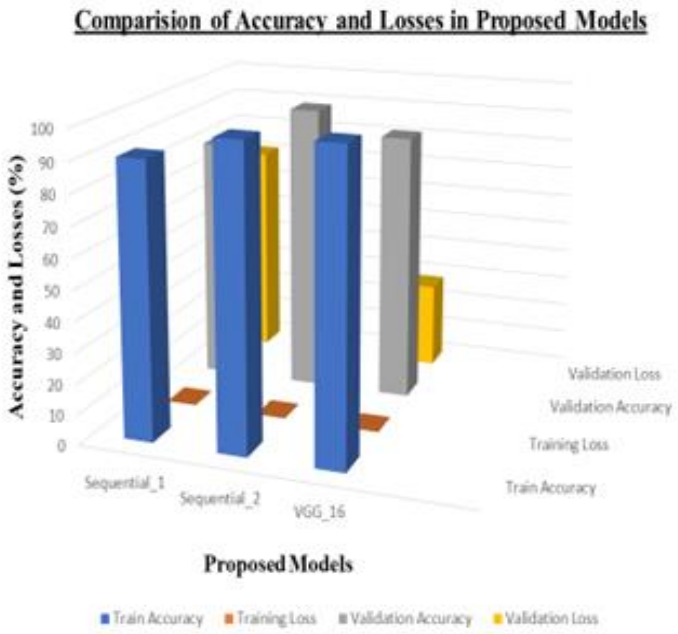


Figure 13. Accuracy and loss comparison of different models

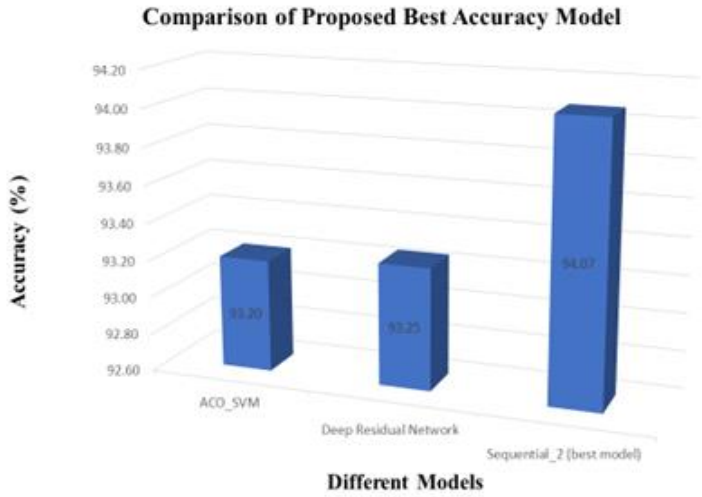


Figure 14. Comparison of SEQUENTIAL2 with preceded models for Accuracy

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 07 August 2023 **Revised:** 08 September 2023 **Accepted:** 15 October 2023*

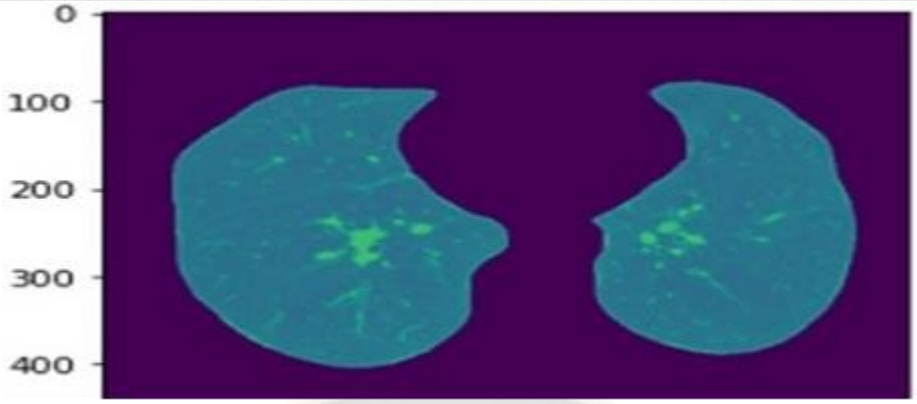


Figure 15. Float64 image of slice from original dataset

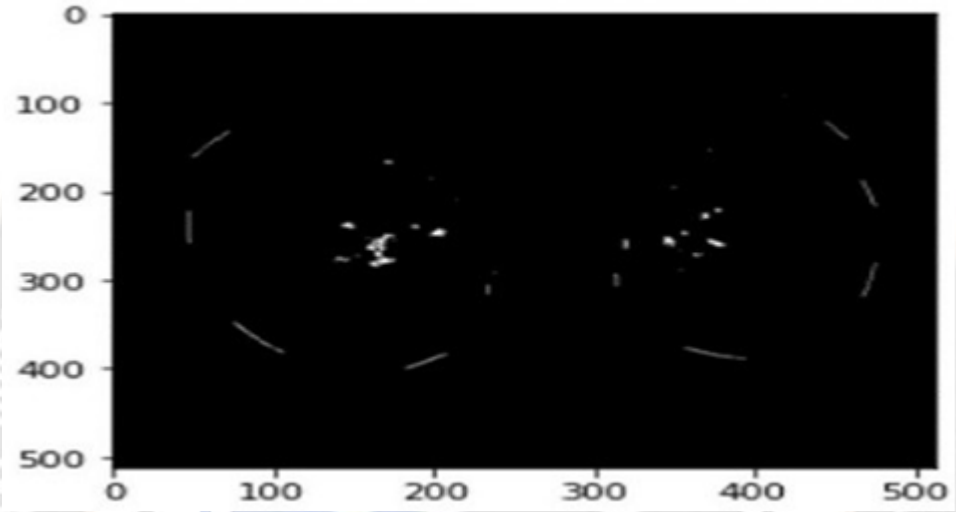


Figure 16. Converted RGB image of a slice using skimage

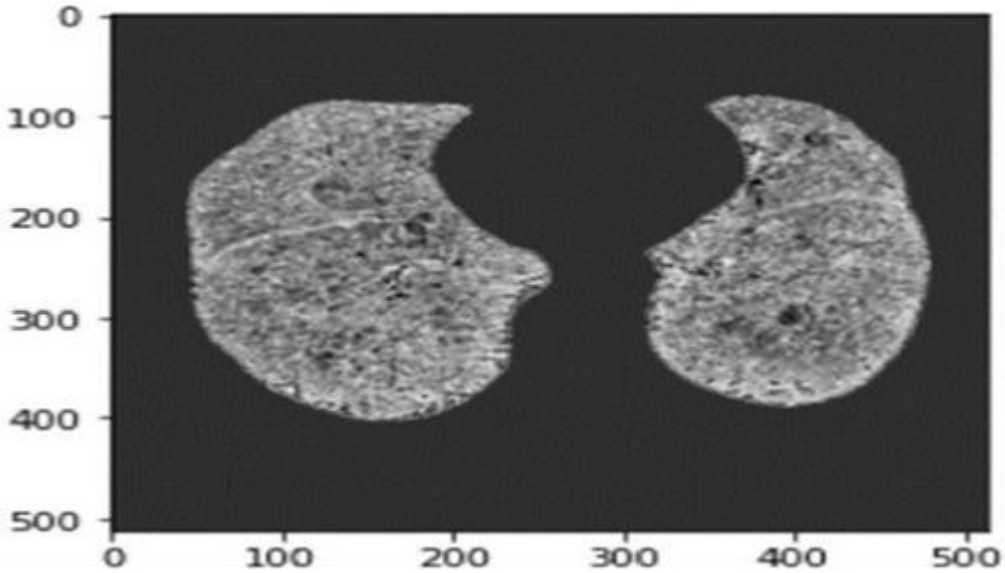


Figure 17. Converted RGB int8 image of a slice using open cv

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *07 August 2023* ***Revised:*** *08 September 2023* ***Accepted:*** *15 October 2023*

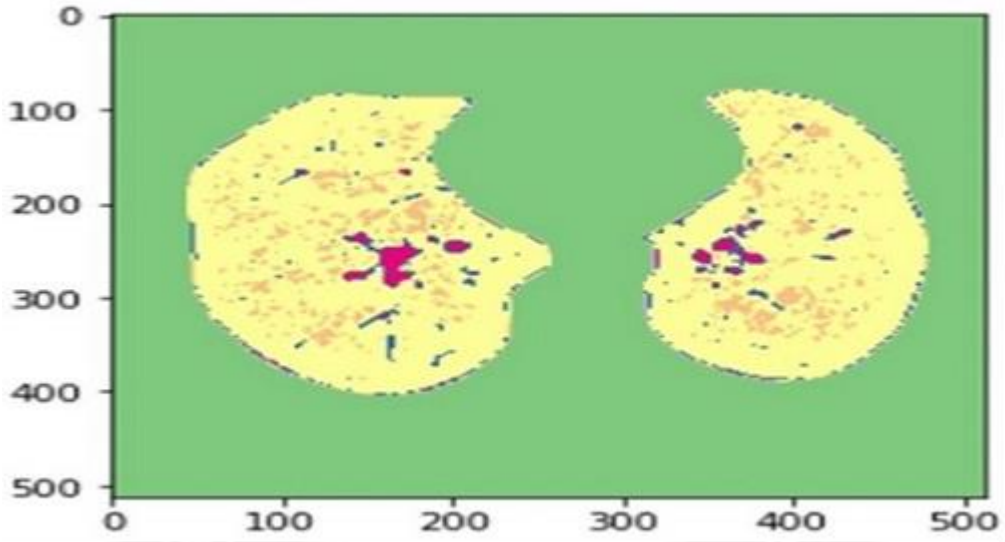


Figure 18. Type of Lung Cancer Image plotted with matplotlib

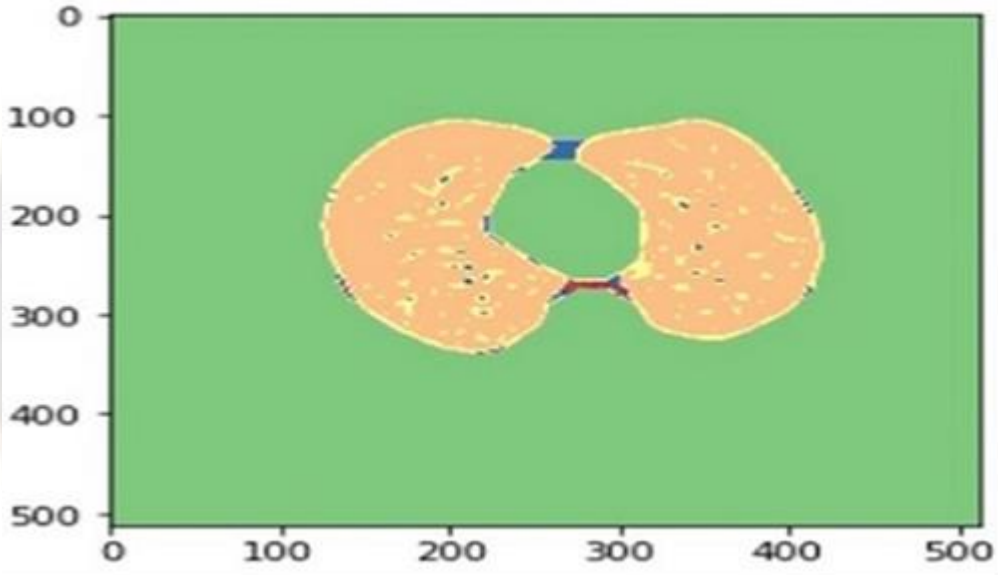


Figure 19. Image of Healthy Lungs plotted with matplotlib