

AI-Powered Segmentation and Validation: A New Era for Lung Cancer Detection in Biomedical

Rajender Kumar¹, Mani Devi², Himanshu Gupta³ and Punit Soni⁴

^{1,3-4}Chitkara University Institute of Engineering and Technology, CSE Department, Chitkara University, Punjab, India
Email: kumar.rajender@chitkara.edu.in, himanshu.gupta@chitkara.edu.in, punit.soni@chitkara.edu.in

²SRM University, Haryana, India
Email: maanisingh38@gmail.com

Abstract— It is still what we see as the number one cause of death world-wide; lung cancer. To improve patient chances of survival what we need is for an early and accurate diagnosis. Biomedical imaging which includes CT scans is what we have for the detection of lung cancer. Still, it is a great amount of time and judgment which goes into very closely looking at these images for signs of lung nodules which are an indicator of lung cancer. The rise of artificial intelligence as a player to really transform this is noteworthy.

This study will contrast the various segmentation methods, identify the best method, and apply it to a hybrid model. Similar to this, segmented regions are used to apply the most effective feature extraction approaches in terms of mean square error and parameter execution time. Using the Feed Forward Back Propagation Neural Network & the SVM classifier (polynomial kernel property) to extract the valid key points from the segmented part classification, it will determine the greatest accuracy with the least amount of hybrid model complexity and execution time.

In order to diagnose lung cancer, this article inquires the decision of AI algorithms on biomedical picture segmentation and validation.

The method of detecting and separating regions of interest (ROIs) within an image automatically is called segmentation. Correctly extracting lung region from its neighboring tissues and further segmenting the potential nodules in the lung parenchyma are important initial steps to diagnose lung cancer.

This limitation of manual analysis potential can be addressed by the AI-related segmentation algorithms, which often provide extreme performance accuracy together with efficiency.

It is necessary to perform validation after segmentation in order to maintain the correctness of the ROIs obtained.

In this paper, we explore the applicability of AI-based validation techniques to assess the robustness and fidelity in the segmentation process for various diseases, ultimately supporting better health management and promoting a healthy life. These strategies could involve training artificial intelligence (AI) models to differentiate between actual lung nodules and other anatomical content or artefacts that might be mistaken for a nodule. Cancer is considered worldwide as one of the major causes of death and morbidity. If current trends continue, up to 2030 an estimated 27 million more cancer cases are expected, according to the International Agency for Research on Cancer (IARC).

Index Terms—SURF, SIFT, PCA, MSER, Machine Learning, CNN and ANN.

I. INTRODUCTION

Around the world, cancer is thought to be a major cause of mortality & morbidity. Before 2030, there are projected to be 27 million additional cases of cancer, according to the report from International Agency for Paper on Cancer (IARC). Over the course of a lifetime, lung cancer is predicted to affect 1 in 18 men and 1 in 46 women. Unchecked cell proliferation in the lung tissues is the cause of lung cancer. Metastasis, also known as infiltration and invasion of adjacent tissue prior to the lungs, may result from this expansion [1]. The type of histology cancer, the extent of dissemination (the stage), and the patient's performance state all influence the prognosis and course of treatment.

Radiation, surgery, and chemotherapy are among the available treatments [2]. After being diagnosed with lung cancer, just 14% of patients survive for five years, although this number varies depending on the patient's overall condition, stage, and other factors [3]. The following are signs of lung cancer detection::

- Hemoptysis (coughing up blood)
- Dyspnea (shortness of breath with activity)

As seen in figure 1 Small-cell lung cancer (SCLC) and non-small-cell lung cancer (NSCLC) comprise the majority of cases. "Bronchogenic Carcinomas," or lung tumours, are classified into two categories [4]:

- Small cell lung cancers (SCLCs)
- Non-small cell lung cancers (NSCLCs)

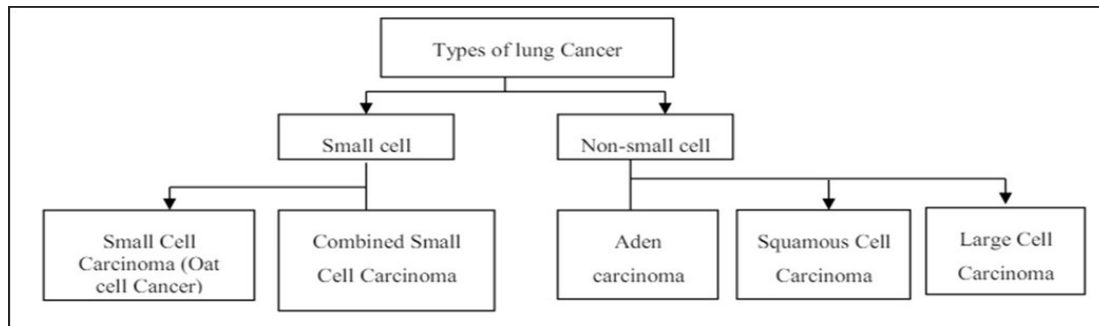


Figure 1. Types of lung cancer

A. Partitionning Lung Cancer and its Methods

Segmentation is the process of pinpointing the precise area in which cancer needs to be found. Pneumonia segmentation with precision is essential and vital. It is important for the proper diagnosis and management of lung cancer. Many circumstances, like Accurate segmentation has a direct impact on the classification of lung nodules and malignancy in CAD for feature extraction [5]. The following describes various segmentation methods (Figure 2) for lung nodules extracted from pictures.

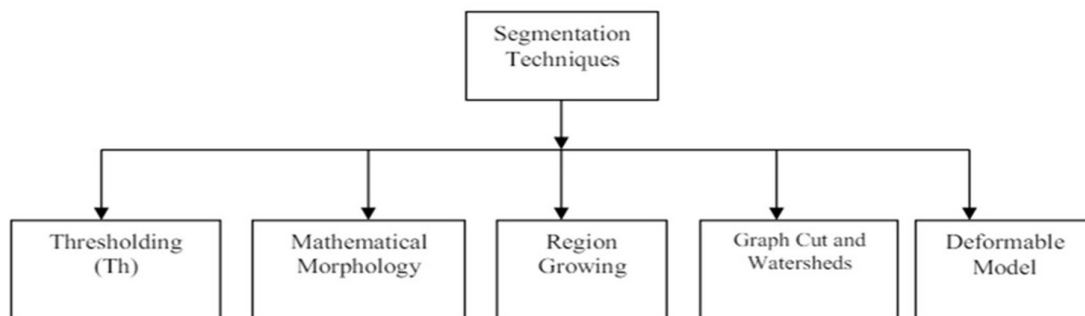


Figure 2. Different segmentation techniques

II. METHODS FOR FEATURE EXTRACTION

Through the computation of certain traits or qualities that are capable of differentiating one input pattern from another, feature extraction seeks to represent real data in a different way.

A. PCA, or principal component analysis

The goal of PCA is to economically describe the data by reducing the observable variables (big dimensionality data space) to autonomous variables (feature space with small intrinsic dimensionality). With basis vectors that track the modes of highest variance in the data, it provides the data in a unique coordinate system. Thus, a unique basis vector is computed for a specific collection of data. The way principle component analysis (PCA) operates is by first creating a new set of linearly uncorrelated variables from the original features.

B. SIFT (Scale-Invariant Feature Transform)

When extracting these traits, there are a lot of things to take into account. The features provided by SIFT pictures properties are for a collection of objects which are independent of many of the complications found in another techniques, like object rotation and scale [6]. SIFT image features enable the identification of items in several photos of same area taken from different places within environment, in addition to allowing object recognition in large images [7]. Additionally, the SIFT function is highly adaptable to the image's "noise."

The SIFT algorithm uses the following four-stage filtering techniques to extract these features:

- Scale-Space Extrema Detection
- Key-point Localisation
- Orientation Assignment
- Key-point Descriptor

C. SURF (Speed Up Robust Features)

It is a descriptor-based limited feature detector. It can be applied to tasks like 3D reconstruction, item detection, and categorization [8]. It is primarily utilised in image processing and is a quick and reliable interest point detector (Figure 4). The algorithm known as SURF. The three essential components for the method are feature extraction, feature matching and feature description.

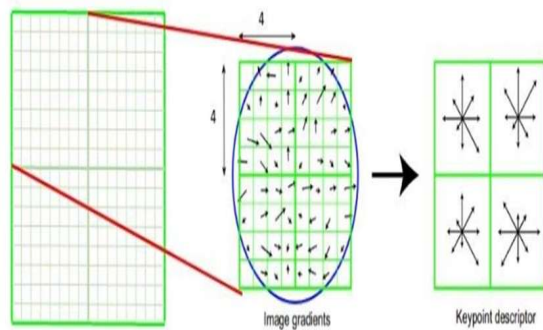


Figure 3. Transform of Scale-Invariant Features

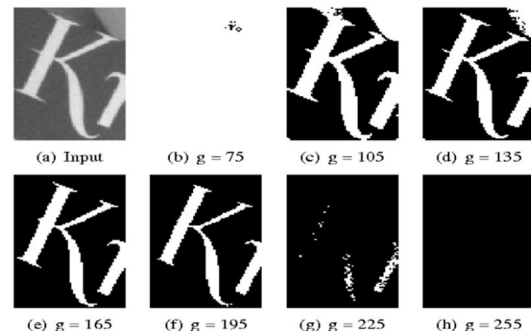


Figure 4. SURF (Speed Up Robust Features)

D. MSER (Maximally Stable Extremal Regions)

A computer vision approach called Maximally-StableExtremal-Regions (MSER) is utilised for segmenting picture regions and blob detection. MSER looks for areas or regions in an image that show consistent colour or intensity throughout a range of scales [9]. These areas are frequently blob-like structures, and MSER is especially helpful in identifying objects or areas of interest with unique textures or shapes. MSER is appropriate for a range of computer vision applications because it is resilient to changes in scale, illumination, and viewpoint.

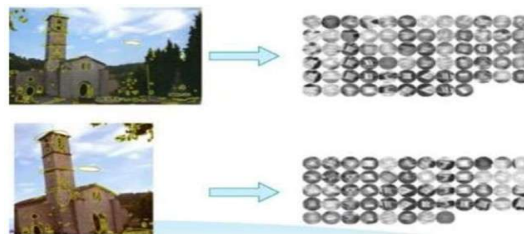


Figure 5. MSER

E. Machine Learning

Machine learning, a branch of artificial intelligence (AI), is concerned with developing models and algorithms that let computers learn from data and make decisions or predictions without the need for explicit programming.

F. Cross Validators (Classifier)

The segmentation approaches may be cross-verified using the classification algorithms. The classification models are used to predict if the processed image is cancerous.

- Support Vector Machines (SVM)
- Linear Discriminate Analysis (LDA)
- Artificial Neural Network (ANN)
- K-Nearest Neighbours (KNN)
- Naïve Bayes
- Convolutional Neural Networks (CNN)

III. CLASSIFICATION OF LUNG CANCER AND SEGMENTATION BY USE OF NEURAL NETWORKS FOR BIOMEDICAL PHOTOS

Every year, lung cancer takes more lives than breast, brain, and prostate cancer combined—especially for individuals between the ages of 45 and 70. It is difficult to detect early since it frequently manifests at an advanced stage. Using feature extraction and optimisation algorithms, Computer-Aided Detection (CAD) facilitates early detection throughout the training and classification stages [10]. After feature extraction using methods like SIFT and SURF, k-Nearest Neighbour (K-NN) classification is performed. With the use of MRI and CT scans, sophisticated techniques like Support Vector Machines (SVM) and Genetic Algorithms (GA) improve feature optimisation and classification accuracy. In crucial point detection, SURF performs better than SIFT thanks to enhanced feature vectors via GA optimisation.

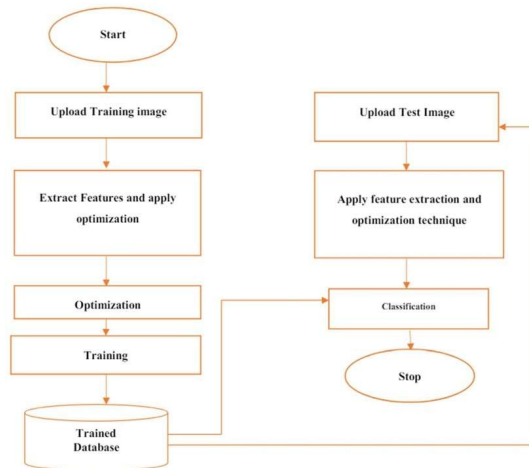


Figure 6. Computer Aided Design Model (CAD) Working

The SVM's Radial Basis Function (RBF) receives the optimised vector in order to classify it. Accuracy weight, True Positive (TP), and False Positive (FP) ratios are used to evaluate performance [11]. Moreover, feature vector optimisation is done with GA. Additionally, work is done on template matching and the suggested K-NN and CCKOO search method for lung cancer categorization. The fitness function that is used seeks to use an ANN to simplify the division process. Kuruvilla's algorithm is then used to compare this paper work.

TABLE I: PCA, SIFT AND SURF FEATURE VECTORS

Image	PCA- χ Principle Eigen Vector)	SIFT- χ Key Points	SURF - χ Key Points
Image 1(255 * 394)	256 * 1	252*252	221*64
Image 2 (194*256)	256 * 1	253*252	166*54
Image 3 (182*278)	256 * 1	252*252	169*32
Image 4 (196*200)	256 * 1	253*252	192*52

TABLE II: MAXIMUM SIGNAL-TO-NOISE RATIO (PSNR) IN THE PRESENCE OF WHITE GAUSSIAN NOISE

PCA— Average PSNR	SIFT— Average PSNR	SURF— Average PSNR
41.36	44.92	45.62

Table 2 illustrates that, out of the three feature extraction strategies, SURF has the highest PSNR. As a result, SURF is ultimately chosen as the main feature holder to be processed further. The given fitness function (Table 3) is used for executing after the retrieved properties set is sent to GA for streamlining [12].

where $f(t)$ is the selection threshold & $f(r)$ is current feature vector. 20–25% fewer SURF functionalities are available after GA is implemented. FFBP NN receives the optimised feature set and uses it for training. A multiclass classifier called FFBP NN uses feed forwarding to meet stopping constraints and backpropagation to confirm the Mean Square [13-14].

TABLE III: USED GENETIC FITNESS FUNCTION

GA Fitness	if $f(r) > f(t)$
	0 otherwise

Paper on machine learning for lung cancer diagnosis is gaining traction, with the goal of increasing survival rates by detecting cancer in nodules that are still in the early stages. There are several imaging modalities used, including CT, MRI, PET, and X-rays. Proposed models by radiologists include cloud-based CNN CAD systems, 3D-CNN, and feature fusion [15].

This paper evaluates current methods and presents a CAD model that uses FFBP-NN, GA, and SURF to identify lung cancer quickly and accurately. SIFT and PCA were used in the initial feature vector testing. GA optimisation improved classification accuracy to 96.5% with 93.8% sensitivity, outperforming earlier techniques, while reducing SURF feature vectors by 21%. To make even more progress, future studies may examine various fitness functions and neuron counts [16-17].

Using a hybrid model that combines ROI identification, segmentation, and feature extraction approaches, this work investigates the classification of lung cancer using a dataset of 800 CT scan pictures. In order to extract features with high accuracy, sensitivity, specificity, and precision, it uses optimisation techniques like PSO and ABC in addition to SIFT and SURF. The use of evolutionary algorithms in feature selection seems promising. The hybrid model has a 93.44% F-measure, 89% recall, and 99.56% accuracy. Results using a polynomial kernel SVM & a neural network with feed forward and reverse propagation were noteworthy. This work provides opportunities for using the same hybrid model to analyse lung cancer in various forms and stages.

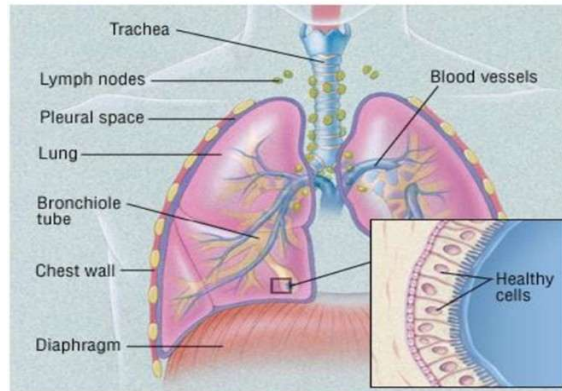


Figure 7. Structure of Human Lung

Techniques used for Detection of Lung Cancer

- Grasshopper Optimization Algorithm (GOA)
- Maximally Stable Extremal Regions (MSER)
- Speed-Up Robust Feature (SURF)
- Convolution Neural Network (CNN)

Using FFBP-NN, GA, and SURF, a hybrid model for lung cancer detection produced 93.8% sensitivity and 96.5% classification accuracy. The use of evolutionary algorithms in feature selection seems promising.

Convolutional neural networks (CNNs) and genetic algorithm optimisation were used in another study to identify lung cancer early. The results showed a high True Positive rate of 99.03% accuracy. CNNbased classifiers improve accuracy and facilitate early illness identification by providing effective feature extraction from CT images.

IV. RESULT ANALYSIS

Figure 8 illustrates the conception and execution of the hybrid model window.

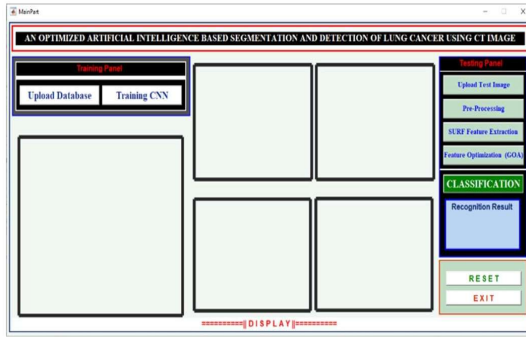


Figure 8. Hybrid Model Window

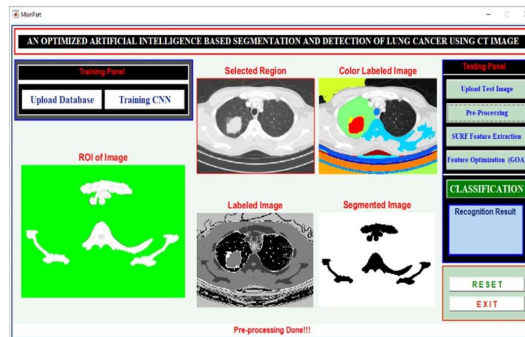


Figure 9. Pre-processing for Test Lung Image Upload

A. Results

SURF, GA, and FFBP-NN provided a computer-aided design hybrid model that allowed for the least mean square error and minimal execution time for the purpose of lung cancer diagnosis. In the early stages, the hybrid model also evaluated PCA and SIFT for feature vectors. Following the addition of white Gaussian noise to the feature vectors, the PSNR for PCA, SIFT, and SURF were 41.45, 44.15, and 45.68, respectively. 76% of the data are utilized in the current study's training, with the remaining material being classified. Using GA, the feature vectors that SURF produced are altered. A 21% reduction in the set of SURF feature vectors is the outcome of GA. Then, FFBP-NN trains on the optimised feature set. Sensitivity and classification accuracy are the characteristics for classification. The suggested algorithm achieves a 96.5% classification accuracy, while the 600 samples' average sensitivity is 93.8%. When compared to the Kuruvilla algorithm that was previously used, it demonstrates an increase of 1-2.5% of each condition. Potential future developments could involve modifications to the current work's fitness function. Future paperers might potentially experiment with different neuron counts for training. When compared to other cutting-edge techniques, the hybrid model produced extremely high levels of accuracy, sensitivity, specificity, and precision in the next phase. New algorithms have demonstrated great assurance in field of property selection recently, owing to their ease of use and potential for worldwide search. A thorough study of the "evaluated features in this paper work" and the "relative review of feature vector" in terms of mean square error and minimum execution time has been provided. Various optimisation strategies (i.e., feature vectors) that are carried out & organized by SVM and obtain a decent accuracy of 99.03% for 800 samples are explained in details.

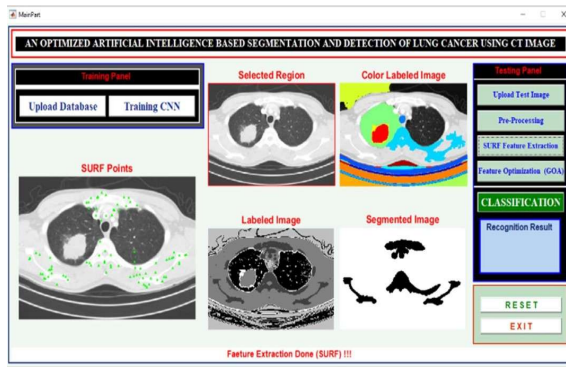


Figure 10. SURF Feature Extraction

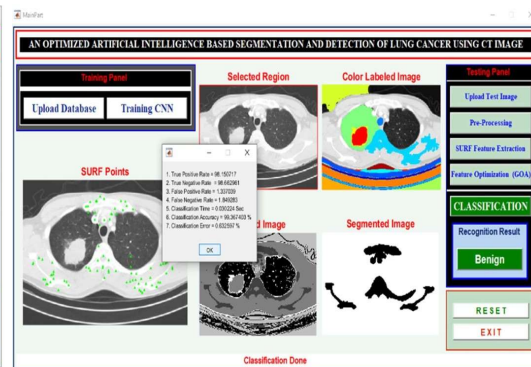


Figure 11. Classification Parameters of Developed Model

TABLE IV: CLASSIFICATION PARAMETERS OF DEVELOPED MODEL

No. of samples	True Positive rate	True Negative rate	False Positive rate	False Negative rate	Classification Time (s)	Classification Accuracy (%)	Classification Error
1	99.1114	97.678	1.3098	1.8845	0.2989	99.6842	0.3158
2	98.6789	97.951	1.312	1.8758	0.2974	98.8765	1.1235
3	98.3456	97.936	1.4678	1.845	0.3768	99.7654	0.2346
4	99.0089	98.178	1.345	1.8678	0.2561	98.9889	1.0111
5	99.10	98.556	1.425	1.8345	0.3011	99.898	0.102
6	98.2345	98.795	1.389	1.8623	0.3057	98.657	1.343
7	98.567	98.981	1.452	1.8734	0.3019	99.2816	0.7184

The current paper has also determined that using Feed Forward Back Propagation Neural Networks with SVM with polynomial kernel property on CT scan pictures generated amazing outcomes in form of parameters like Accuracy, MSE, Precision, Recall, F-Measure, and execution time. The key to saving lives when lung illness is still in its early stages is the bifurcation of lung disease using CTScan pictures. Examination of patterns that can be observed visually on a human lung picture is known as lung cancer detection. Maintaining health and identifying lung diseases are essential objectives for long-term medical paper. Manually observing lung illness is highly complex. Consequently, utilising a human lung CT scan, image processing is employed to identify lung cancer. Several phases, including picture procurement, image preprocessing, segmentation, feature retrieval, & classification, are involved in the image based lung cancer detection process. Accurate feature extraction from CT scans is necessary for the early identification of cancer. In the current study, an SURF descriptor and CNN classifier are used to create an early detection of lung cancer classification system. According to the fitness function assessment criteria, the data value is optimised using the evaluation-based optimisation technique known as the Genetic technique. Convolution Neural Network, a neural network-based classification technique, has been used to extract information from CT images. Three distinct layers serve as the foundation for this classifier: the input layer, the pooling layer, and the convolution layer. The effectiveness of the current study work has been assessed using CT scans in a number of previous studies, which were based on a few metrics including error (%), classification time, true positive, true negative, false positive, false negative, and accuracy (%). The hybrid model's True Positive value is 98.7, while Senthil Kumar et al. (2019) found that it is 95.28. 99% is the average accuracy value. For the detection of lung cancer, hybrid-model combining SURF, GA, & FFBP-NN produced a sensitivity of 93.8% and a classification accuracy of 96.5%. In feature selection, evolutionary algorithms show potential. Another study achieved 99.03% accuracy in early lung cancer diagnosis by using Convolutional Neural Networks (CNN) and Genetic Algorithm optimisation. CNN-based classifiers improve accuracy for early illness detection by providing effective feature extraction from CT images.

REFERENCES

- [1] Dodia, S., Annappa, B., & Padukudru, A. Mahesh. (2022). Recent advancements in deep learning-based lung cancer detection: A systematic review. *Engineering Applications of Artificial Intelligence*, 116*, 105490.
- [2] Nanglia, P., Kumar, S., Mahajan, A. N., Singh, P., & Rathee, D. (2021). A hybrid algorithm for lung cancer classification using SVM and Neural Networks. *ICT Express*, 3*(2), 1-7.
- [3] Nanglia, P., Mahajan, A. N., Rathee, D. S., & Kumar, S. (2020). Lung cancer classification using feed forward back propagation neural network for CT images. *International Journal of Medical Engineering and Informatics*, 12*(5), 447-456.
- [4] Nanglia, P., Mahajan, A. N., Singh, P., & Rathee, D. (2019). Detection & Classification of Lung Cancer at an Early Stage by Applying Feature Extraction Optimization and Neural Network on Hybrid Structure. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 9*(2), 3737-3745.
- [5] Singh, P., Nanglia, P., & Mahajan, A. N. (2019). Improved Lung Cancer Segmentation Using K-Means and Cuckoo Search. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 9*(2), 3746-3758.
- [6] Nanglia, P., Kumar, S., Rathee, D., & Singh, P. (2018). Comparative Investigation of Different Feature Extraction Techniques for Lung Cancer Detection System. In *International Conference on Advanced Informatics for Computing Paper (ICAICR)**. Springer, Himachal Pradesh.
- [7] Shallu, Nanglia, P., Kumar, S., & Luhach, A. K. (2017). Detection and Analysis of Lung Cancer Using Radiomic Approach. In *International Conference on Computational Strategies for Next Generation Technologies**. Springer, Punjab.
- [8] Palani, D., & Venkatalakshmi, K. (2019). An IoT Based Predictive Modelling for Predicting Lung Cancer Using Fuzzy Cluster Based Segmentation and Classification. *Journal of Medical Systems*, 43*(2), 21.

- [9] Lakshmanaprabu, S. K., Mohanty, S. N., Shankar, K., Arunkumar, N., & Ramirez, G. (2019). Optimal deep learning model for classification of lung cancer on CT images. *Future Generation Computer Systems*, 92*, 374-382.
- [10] A. Kumar, R. Gupta, S. Kumar, K. Dutta, and R. Kumar, "Intelligent intrusion detection system using improved Osprey optimization and stacked ensemble learning for IoT-based healthcare systems," *Security and Privacy*, Oct. 2025, doi: 10.1002/spy2.70121.
- [11] P. Soni, X. Cheng, M. Singh, P. Sharma, T. Kumar, M. Paliwal, and R. Kumar, "Multifeature fusion for enhanced content-based image retrieval across diverse data types," *Journal of Electrical and Computer Engineering*, vol. 2025, Art. ID 3889925, 2025, doi: 10.1155/jece/3889925.
- [12] P. Soni, A. Gandhi, S. Mehla, and R. Kumar, "Exploring artificial intelligence transforming image creation based Art.Ai platform and its creative impact," *Journal of Harbin Engineering University*, vol. 44, no. 8, pp. 1526–1538, 2023.
- [13] M. Paliwal, P. Soni, A. Goyal, M. Malik, B. Goyal, and R. Kumar, "Chaotic map-based selective image encryption via diffusion and confusion," *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 27, no. 2-B, pp. 763–773, 2024, doi: 10.47974/JDMSC-1953.
- [14] P. K. Verma and R. Kumar, "A role of information communication technology and its impact on health sector," *International Journal of Advance Research in Computer Science*, vol. 1, no. 2, pp. 122–125, 2010.
- [15] P. Soni and N. Mishra and R. Kumar, "Analysis of Covid-19 pandemic impact," *Turkish Journal of Physiotherapy and Rehabilitation*, vol. 32, no. 3, pp. 9611–9617, 2021.
- [16] D. R. Khanna and S. Kumar and R. Kumar, "Deep learning integrated approach for collision avoidance in Internet of Things based smart vehicular networks," *Journal of Advanced Research in Dynamical and Control Systems*, vol. 10, no. 14 (Special Issue), pp. 1508–1512, 2018.
- [17] D. R. Khanna, P. Verma, and R. Kumar, "Vehicular middleware and heuristic approaches for intelligent transportation systems of smart cities," in *Cognitive Computing for Human-Robot Interaction*, Academic Press, Elsevier, Aug. 2021, pp. 1–18, doi: 10.1016/B978-0-323-85769-7.00019-7.