International Journal of Engineering Technology and Management Science
Website: ijetms.in Issue: 2 Volume No.9 March - April – 2025
DOI:10.46647/ijetms.2025.v09i02.125 ISSN: 2581-4621

Lung Cancer Prediction using CNN

Solimi Raghunat Gowd[1], Bushra Tahseen[2]
[1]*MCA Student, Dr.K.V.Subba Reddy Institute of Technology, Kurnool, Andhra Pradesh, India*
[2]*Associate Professor, Dr.K.V.Subba Reddy Institute of Technology, Kurnool, Andhra Pradesh, India*

Abstract
Lung cancer remains one of the leading causes of cancer-related deaths globally, necessitating the development of advanced diagnostic tools to improve early detection and treatment outcomes. This study proposes a novel approach for lung cancer detection using federated learning, a decentralized machine learning technique that enhances model training while preserving data privacy. Federated learning enables the collaborative training of machine learning models across multiple institutions without the need to centralize sensitive patient data. By leveraging data from diverse sources, including medical imaging and patient records from various hospitals, the proposed system trains a robust deep learning model for detecting lung cancer. This approach not only ensures the privacy and security of patient data but also allows for the incorporation of a wider range of data, leading to improved model generalizability and accuracy. The federated learning framework involves aggregating model updates from participating institutions and combining them to enhance the overall performance of the lung cancer detection model. Preliminary results demonstrate that the federated learning-based model achieves high accuracy and sensitivity in identifying lung cancer from imaging data while maintaining data confidentiality and compliance with privacy regulations.
Keywords: Lung, Cancer, ML, CNN

Introduction
With the rapid increase in population rate, the rate of diseases like cancer, chikungunya, cholera etc., are also increasing. Among all of them, cancer is becoming a common cause of death. Cancer can start almost anywhere in the human body, which is made up of trillions of cells. Normally, human cells grow and divide to form new cells as the body needs them. When cells grow older or become damaged, they die, and new cells take their place. When cancer cells develop, however, this orderly process breaks down. As cells become more and more abnormal, old or damaged cells survive when they should die, and new cells form when they are not needed. These extra cells can divide without stopping and may form growths called tumor. This tumor starts spreading to different of body. Tumors are of two types benign and malignant where benign (non- cancerous) is the mass of cell which lack in ability to spread to other part of the body and malignant (cancerous) is the growth of cell which has ability to spread in other part of body this spreading of infection is called metastasis. There is various type of cancer like Lung cancer, leukemia, and colon cancer etc. The incidence of lung cancer has significantly increased from the early 19th century. There is various cause of lung cancer like smoking, exposure to radon gas, secondhand smoking, and exposure to asbestos etc.

Objectives
1. Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system.
2. It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities.

DOI:10.46647/ijetms.2025.v09i02.125 ISSN: 2581-4621

3. When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as when needed so that the user will not be in maize of instant. Thus the objective of input design is to create an input layout that is easy to follow.

Literature Survey

In the 21st century, cancer is still considered a serious disease as the mortality rates are high. Among all cancer types, lung cancer ranks first regarding morbidity and mortality [1, 2]. There are two main categories of lung cancer: non-small-cell lung cancer (NSCLC) and small cell lung cancer (SCLC). For non-small-cell lung cancer, a subcategorization into lung squamous cell carcinoma (LUSC) and lung adenocarcinoma (LUAD) is further used. These types of cancers account for approximately 85% of lung cancer cases [3]. Compared with the diagnosis of benign and malignant, further fine-grained classification of lung cancers such as LUSC, LUAD, and SCLC is of great significance for the prognosis of lung cancer. Accurately determining the category of lung cancer in the early diagnosis directly influences the effect of the treatment and thus the patients' survival rate [1, 4]. Positron emission tomography (PET) and computed tomography (CT) are both widely used noninvasive diagnostic imaging techniques for clinical diagnosis in general and for the diagnosis of lung cancer in particular [4]. Immunohistochemical evaluation is considered the gold standard for lung cancer classification. However, this procedure requires a tissue biopsy, an invasive procedure with the inherent risk of a delayed diagnosis and thus exacerbation of the patient's pain.

Advances in artificial intelligence research enabled numerous studies on the automatic diagnosis of lung cancer. The use of data in lung cancer-type classification is roughly divided into three categories: CT and PET image data as well as pathological images [5]. The well-known data science community Kaggle provides high-quality CT images for participants with the task to distinguish malignant or benign nodules from pulmonary nodules. Kaggle competitions repeatedly produce excellent deep learning approaches for these tasks [6, 7]. With the progresses in the research of automatic lung cancer diagnosis, studies are no longer limited to the classification of benign and malignant nodules and data sets are no longer limited to CT images [8–12]. Wu et al. [9] use quantitative imaging characteristics such as statistical, histogram-related, morphological, and textural features from PET images to predict the distance metastasis of NSCLC, which shows that quantitative features based on PET images can effectively characterize intratumor heterogeneity and complexity. Two recent publications propose the application of deep learning to pathological images to classify NSCLC and SCLC [10] and to classify transcriptome subtypes of LUAD [11]. The complexity of the clinical diagnosis of lung cancer is also characterized by the wide range of imaging modality, which is employed in the diagnosis [13, 14].

Previous research already proved that deep learning approaches can not only use the feature distribution patterns from different pulmonary imaging modalities but even merging different features to achieve the computer-aided diagnosis. Liang et al. [15] employ multichannel techniques to predict the IDH genotype from PET/CT data using a convolutional neural network (CNN), while other approaches use a parallel CNN architecture to extract several features of different imaging modalities [16, 17]. Compared with the classification of the benign and malignant, the classification of the three types of lung cancer from medical images are more suitable to constitute a finegrained image recognition problem as diverse distributions of features and potential pathological features need to be considered. Because the fine-grained features which need to extract in images, and meanwhile the lesion region is a small part of the whole image, the deep learning framework is susceptible to feature noise. At present, most methods based on various deep learning frameworks have proved to have certain bottleneck in finegrained problems. In order to solve this problem, the previous research mainly implements the attention mechanism from the two dimensions (channel and spatial) of the feature representation. The channel attention mechanism models the relationship between feature channels [18], while the spatial attention mechanism ensures that noise is suppressed by weighting feature representation spatially [19–21]. So far, spatial attention mechanism has been used in medical image processing to enhance extracted features [20, 21]. The channel attention mechanism has been used in the detection and classification of pulmonary disease [22, 23]. The presentation of these attention mechanisms illustrates the source of characteristic noise from different perspectives. There are few related studies on how to use the attention mechanism more effectively on images with different imaging modalities, so the deep learning model based on the multimodality dataset still has problems in finegrained problems.

Existing System

Traditional systems for lung cancer detection primarily rely on centralized machine learning models trained on large datasets of medical imaging, such as chest X-rays and CT scans. These models are typically developed using data collected from various medical institutions, which is aggregated and processed at a central location. Conventional methods involve the use of convolutional neural networks (CNNs) and other deep learning techniques to analyze imaging data for early signs of lung cancer, aiming to enhance diagnostic accuracy and early detection. However, these systems face several challenges, including data privacy concerns, as sensitive patient information must be transferred and stored centrally. Additionally, centralizing data from multiple sources can lead to issues related to data integration, standardization, and the management of large-scale datasets. The effectiveness of these models is often constrained by the availability of high-quality, annotated data and the risk of overfitting to specific datasets. Furthermore, traditional approaches may not fully utilize the diverse data available across various institutions, potentially limiting the generalizability and robustness of the detection models. These limitations underscore the need for innovative solutions that address privacy concerns and leverage distributed data more effectively, paving the way for advanced methodologies such as federated learning.

DRAW BACKS :

- **Communication Overhead**: Federated learning requires frequent communication between the central server and participating institutions to aggregate model updates. This can lead to significant communication overhead and latency, especially when dealing with large models or numerous participants.
- **Data Privacy and Security**: While federated learning aims to preserve data privacy by keeping sensitive data decentralized, it still faces challenges. The transmission of model updates can potentially leak information about the data, necessitating the use of additional privacy-preserving techniques such as differential privacy.
- **Model Aggregation Challenges**: Aggregating updates from models trained on diverse datasets can be complex. Variability in data quality and distribution across institutions may lead to difficulties in combining these updates effectively, potentially impacting the model's overall performance.
- **Inconsistent Data Quality**: The quality and quantity of data available at different institutions can vary significantly. This inconsistency can affect the training process, leading to potential imbalances and reduced model accuracy if certain institutions have less representative data.
- **Computational Resources**: Although federated learning reduces the need to centralize data, it still requires substantial computational resources at each participating site to train local models. This can be a limitation for institutions with limited computational infrastructure.

Proposed System

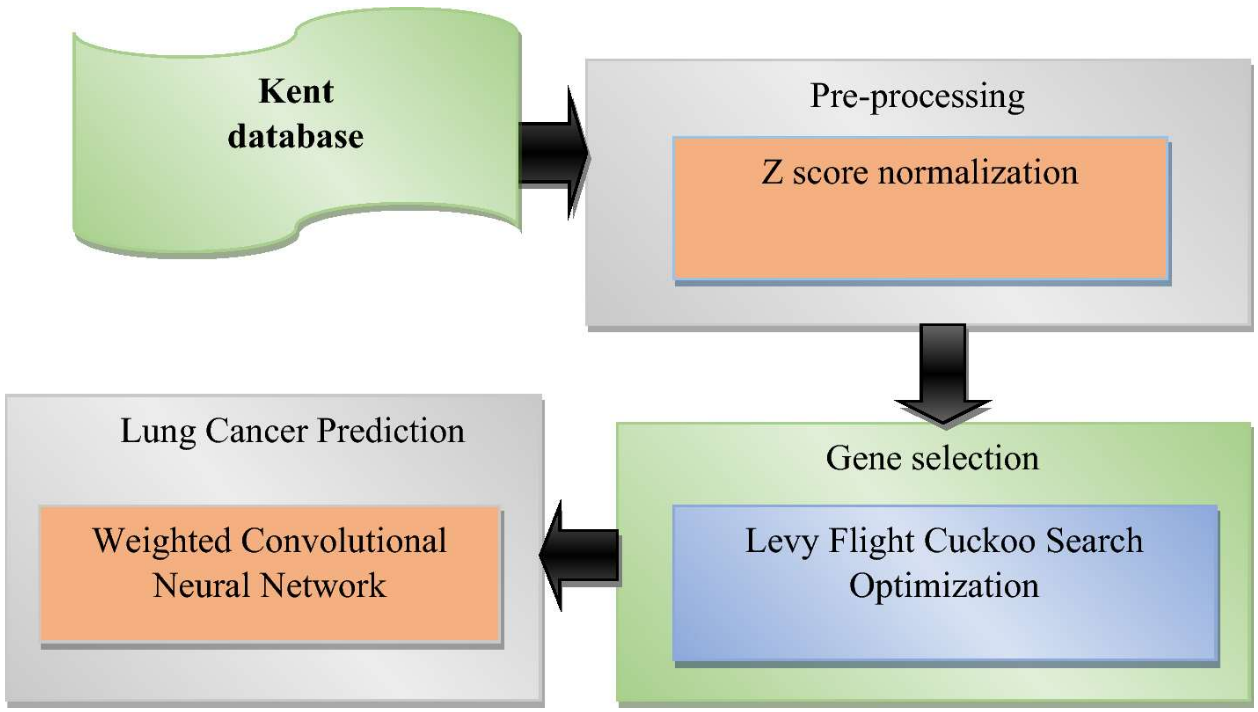
The proposed system for lung cancer detection leverages federated learning to enhance diagnostic accuracy while maintaining stringent data privacy and security. This approach involves a decentralized network where multiple healthcare institutions collaboratively train a shared deep learning model without exchanging sensitive patient data. Each participating institution trains a local model using its own dataset of medical imaging, such as chest X-rays and CT scans. These locally trained models periodically send updates—such as model weights and gradients—to a central server, which aggregates these updates to improve the global model. The federated learning

International Journal of Engineering Technology and Management Science
Website: ijetms.in Issue: 2 Volume No.9 March - April – 2025
DOI:10.46647/ijetms.2025.v09i02.125 ISSN: 2581-4621

framework employs secure communication protocols and advanced privacy-preserving techniques, such as differential privacy and secure multi-party computation, to protect patient information during transmission and aggregation. By incorporating diverse datasets from various institutions, the system enhances the robustness and generalizability of the lung cancer detection model.

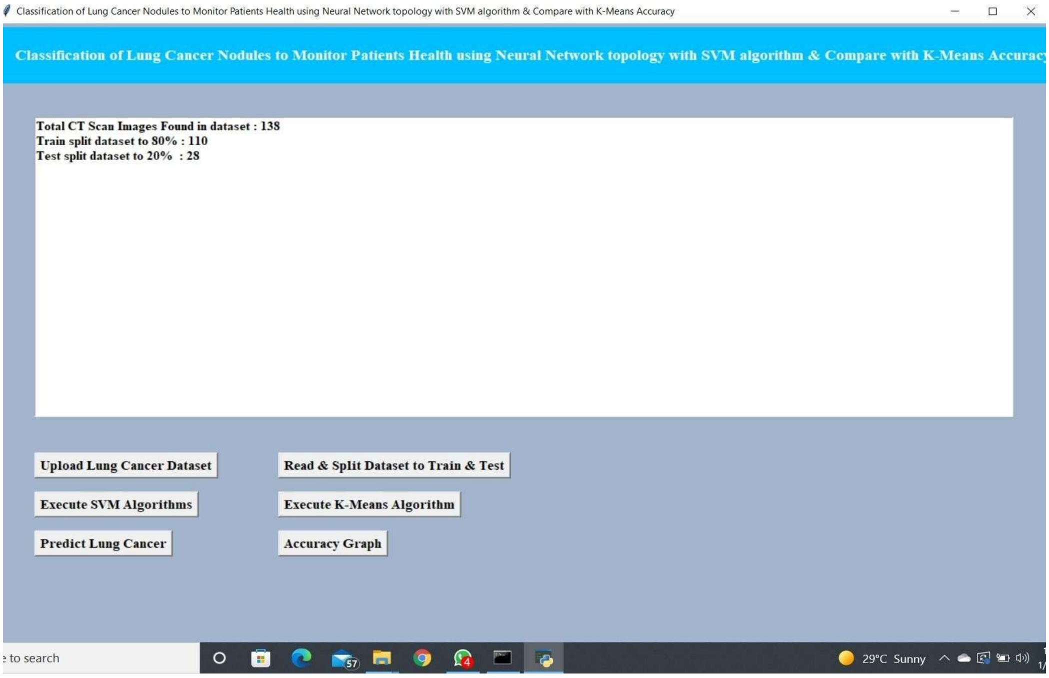
ADVANTAGES :

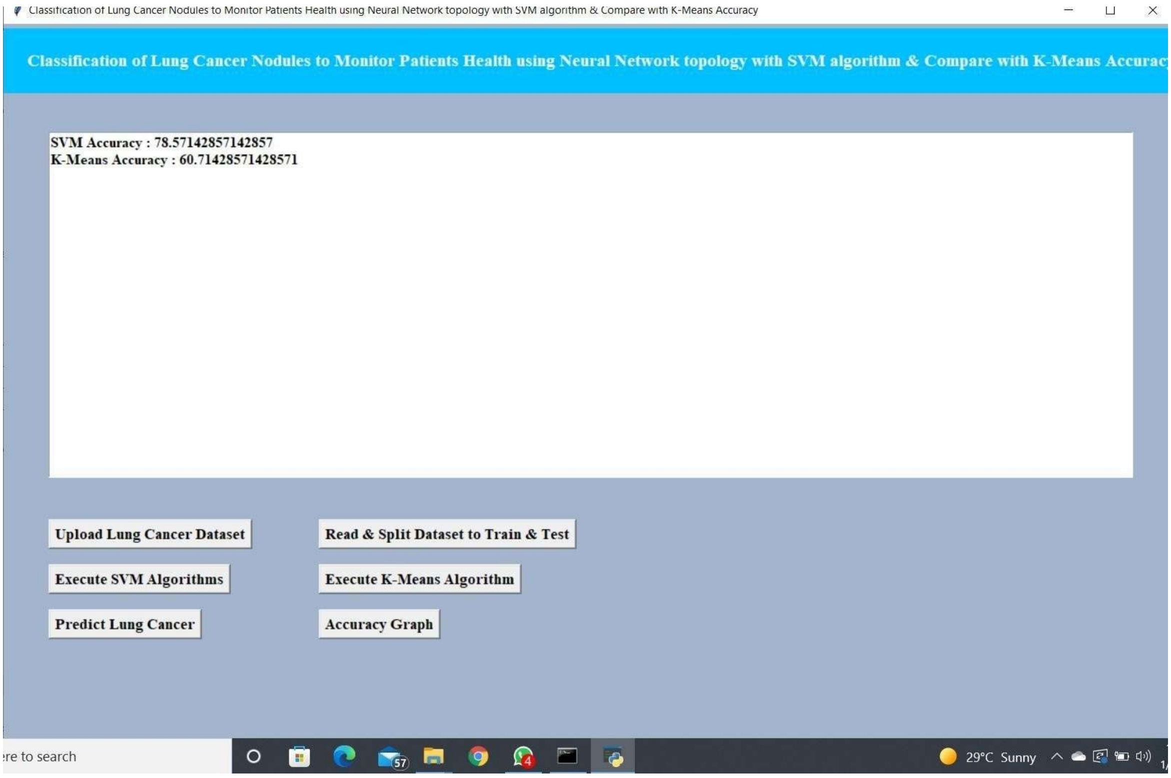
- **Enhanced Data Privacy**: Federated learning allows for the training of models without the need to centralize sensitive patient data. This approach mitigates privacy concerns and adheres to data protection regulations, as individual institutions retain control over their data.
- **Improved Model Generalization**: By incorporating diverse datasets from multiple healthcare institutions, federated learning helps create a more robust and generalized model. This diversity improves the model's ability to detect lung cancer across different populations and imaging conditions.
- **Reduced Data Transfer Risks**: Since only model updates are shared and not raw data, the risk of exposing sensitive patient information during transmission is significantly reduced. This enhances the overall security and confidentiality of the data.
- **Collaboration Across Institutions**: Federated learning fosters collaboration between different healthcare providers, enabling them to contribute to the development of a more comprehensive and effective lung cancer detection model without sharing proprietary or sensitive data.
- **Scalability**: The federated learning approach can scale to include a large number of institutions and datasets, making it possible to leverage a wide range of data sources without the need for a centralized data repository.



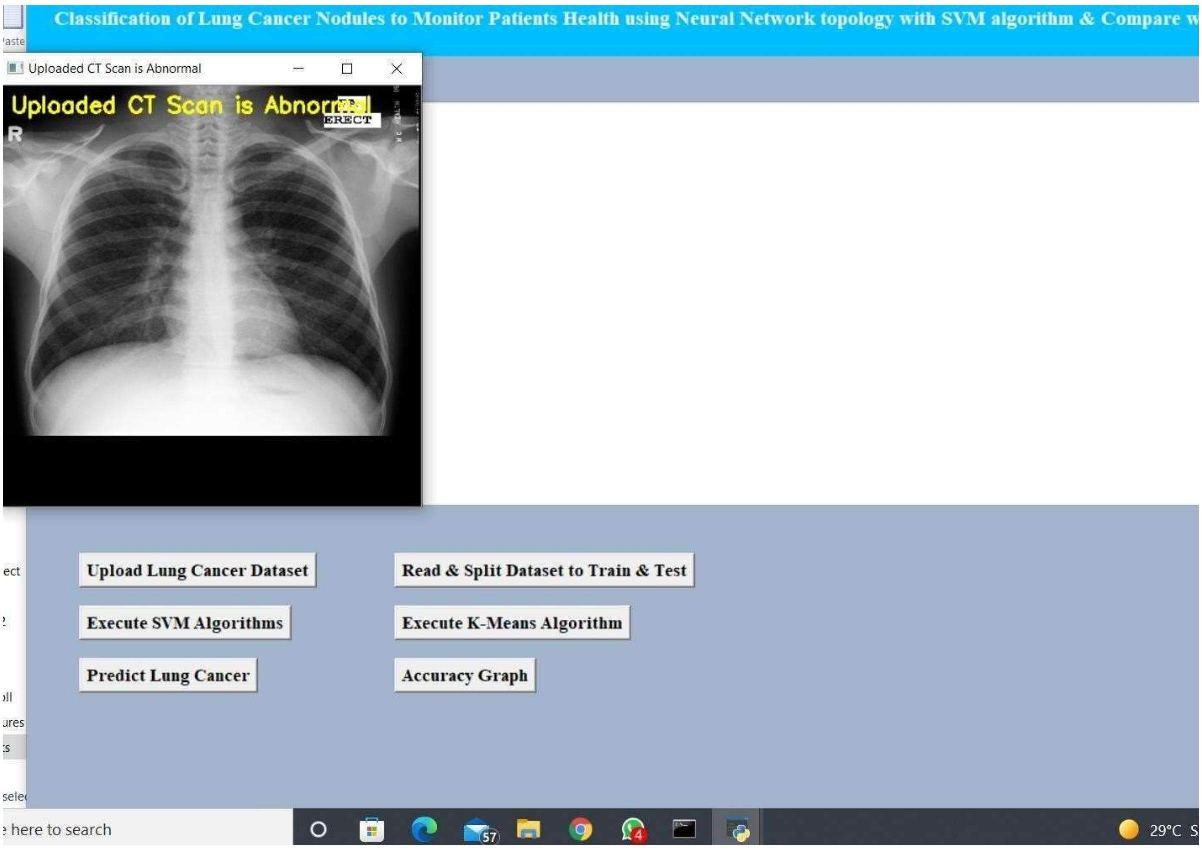
International Journal of Engineering Technology and Management Science
Website: ijetms.in Issue: 2 Volume No.9 March - April – 2025
DOI:10.46647/ijetms.2025.v09i02.125 ISSN: 2581-4621

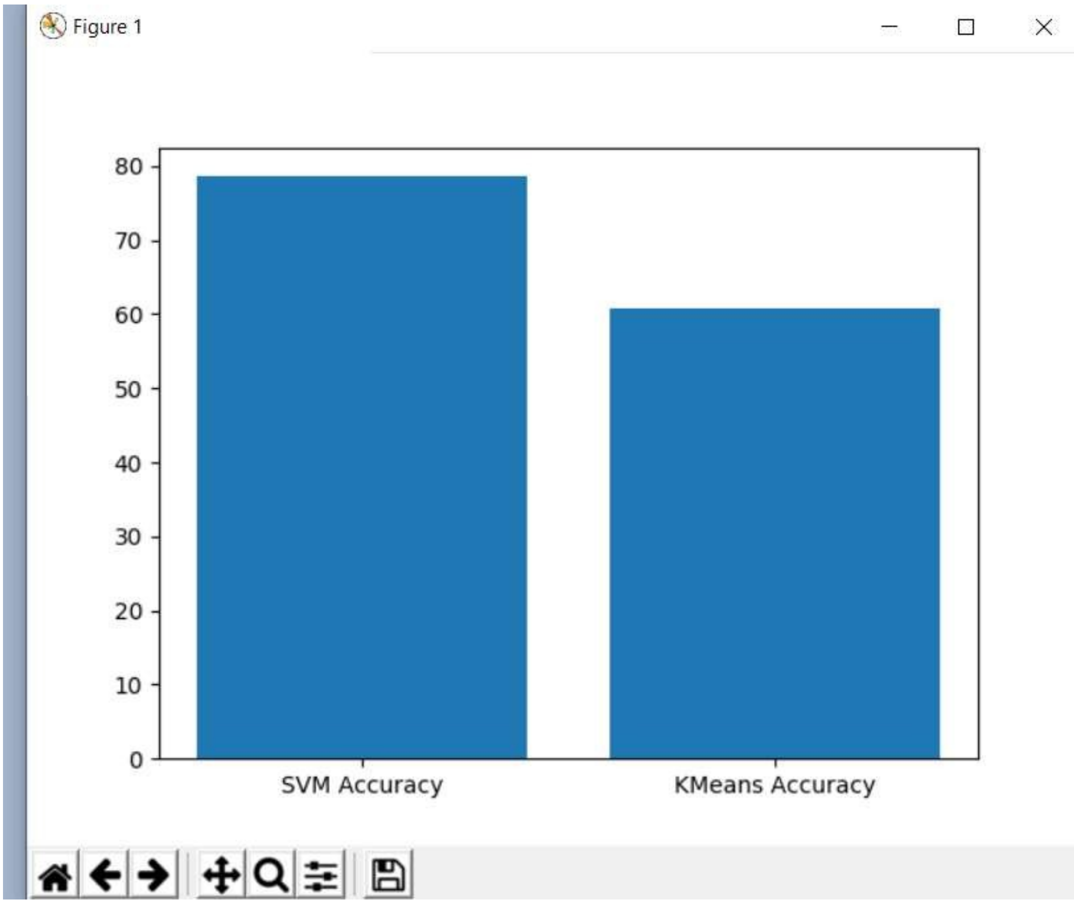
Results





International Journal of Engineering Technology and Management Science
Website: ijetms.in Issue: 2 Volume No.9 March - April – 2025
DOI:10.46647/ijetms.2025.v09i02.125 ISSN: 2581-4621





DOI:10.46647/ijetms.2025.v09i02.125 ISSN: 2581-4621

Conclusion

The proposed model shows the overview of prediction of lung cancer at an early stage. After prediction of the tumour begins malignant or benign, we generate a confusion matrix for each machine learning technique and based on the confusion matrix we calculate accuracy, Recall, precision and F1 score. From the result we can say that our proposed model can distinguish between benign and malignant, and it can be seen that artificial neural network is providing more accuracy in both texture and region based, as well as from the recall value we can say that it has correctly indentified maximum number of malignant tumour In near future deep learning shall outperform machine learning in the field of image classification, object recognition and feature extraction. CNN networks are well known for its features in providing accuracy with higher number of hidden layers in it.

References

[1] Krishnaiah, V., G. Narsimha, and Dr N. Subhash Chandra. "Diagnosis of lung cancer prediction system using data mining classification techniques." *International Journal of Computer Science and Information Technologies* 4.1 (2013): 39-45.

[2] Zhang, Junjie, et al. "Pulmonary nodule detection in medical images: a survey." *Biomedical Signal Processing and Control* 43 (2018): 138- 147.

[3] Fenwa, Olusayo D., Funmilola A. Ajala, and A. Adigun. "Classification of cancer of the lungs using SVM and ANN." *Int. J. Comput. Technol.* 15.1 (2016): 6418-6426. [4]Daoud, Maisa, and Michael Mayo. "A survey of neural network-based cancer prediction models from microarray data." *Artificial intelligence in medicine* (2019).

[5] Palani, D., and K. Venkatalakshmi. "An IoT based predictive modelling for predicting lung cancer using fuzzy cluster based segmentation and classification." *Journal of medical systems* 43.2 (2019): 21.

[6] Lynch, Chip M., et al. "Prediction of lung cancer patient survival via supervised machine learning classification techniques." *International journal of medical informatics* 108 (2017): 1-8.

[7] Öztürk, Şaban, and Bayram Akdemir. "Application of feature extraction and classification methods for histopathological image using GLCM, LBP, LBGLCM, GLRLM and SFTA." *Procedia computer science* 132 (2018): 40-46.

[8] Jin, Xin-Yu, Yu-Chen Zhang, and Qi-Liang Jin. "Pulmonary nodule detection based on CT images using convolution neural network." *2016 9th International symposium on computational intelligence and design (ISCID)*. Vol. 1. IEEE, 2016.

[9] Sumathipala, Yohan, et al. "Machine learning to predict lung nodule biopsy method using CT image features: A pilot study." *Computerized Medical Imaging and Graphics* 71 (2019): 1-8.

[10]A. Jemal, F. Bray, M. M. Center, J. Ferlay, E. Ward, and D. Forman, "Global cancer statistics," *CA: A Cancer Journal for Clinicians*, vol. 61, no. 2, pp. 69–90, 2013.

R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2018," *CA: A Cancer Journal for Clinicians*, vol. 68, no. 1, pp. 7–30, 2018