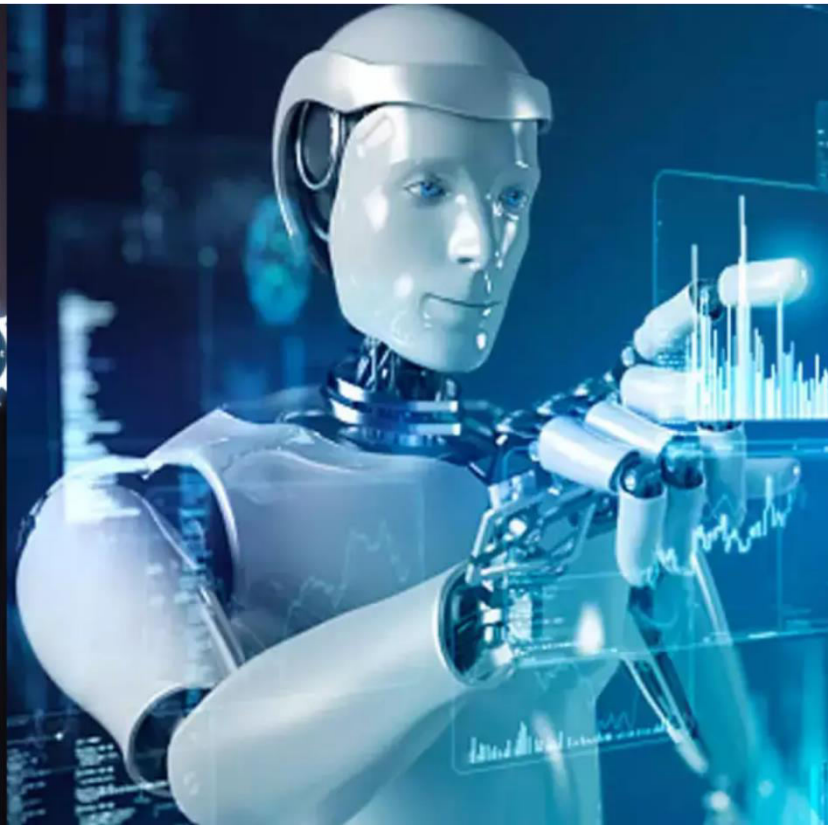
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404512|

Smart Cancer Detection using Histological Images

R.Prathiba, Sk. Jani Basha, Y. Harsha Vardhan, Y.H. Shadguna Siddhi, G. Narendra Prasad

Department of Computer Science and Engineering, Bharath Institute of Higher Education and Research, Selaiyur, Chennai, India

ABSTRACT: Lung cancer is a leading cause of cancer-related mortality worldwide, underscoring the urgent need for early and accurate diagnosis. This research proposes an AI-driven approach for early lung cancer detection using hybrid histological image analysis in MATLAB. The system integrates machine learning and deep learning techniques to analyze CT scan images, extracting critical histopathological features to improve classification accuracy. A hybrid methodology combining convolutional neural networks (CNNs) with traditional image processing enhances precision in distinguishing malignant from benign cases. Feature extraction techniques such as texture analysis, edge detection, and morphological processing refine diagnostic accuracy. The system is trained on annotated datasets to ensure robustness and generalization. Experimental results demonstrate superior sensitivity, specificity, and overall diagnostic performance compared to conventional methods. The integration of AI enables automated, efficient, and reliable detection, reducing subjectivity in manual histopathological assessments. Additionally, the proposed system offers a cost-effective and scalable solution for clinical implementation, making it a valuable tool for radiologists and oncologists. By facilitating early detection, this AI powered framework enhances timely medical intervention, potentially improving patient survival rates and treatment outcomes. This study contributes to advancing AI applications in medical imaging, paving the way for more precise and accessible lung cancer diagnostics.

KEYWORDS: Histological Image Analysis, Machine Learning

I. INTRODUCTION

Lung cancer stands as a formidable global health challenge, ranking among the leading causes of cancer-related mortality worldwide. Its insidious nature, often presenting with subtle or no symptoms in early stages, contributes significantly to delayed diagnosis and consequently, poorer patient outcomes. The imperative for early and accurate detection has driven extensive research into advanced diagnostic methodologies, particularly leveraging the power of medical imaging. Computed tomography (CT) scans, a staple in lung cancer screening and diagnosis, provide detailed cross-sectional images of the lungs, revealing subtle abnormalities that may indicate malignancy. However, the interpretation of these images is oftensubjective and time-consuming, requiring skilled radiologists to meticulously analyze vast datasets. This inherent subjectivity and the sheer volume of data necessitate the development of automated, reliable, and efficient diagnostic tools.

The advent of artificial intelligence (AI), particularly machine learning and deep learning, has ushered in a new era of possibilities in medical imaging analysis. These techniques offer the potential to extract intricate patterns and features from complex image data, surpassing the capabilities of traditional image processing methods. By training AI models on large, annotated datasets, it becomes feasible to develop systems capable of accurately distinguishing between benign and malignant lung nodules, thereby facilitating earlier and more precise diagnoses. This research endeavors to contribute to this evolving landscape by proposing an AI-driven approach for early lung cancer detection, utilizing hybrid histological image analysis within the MATLAB environment.The development of a robust and effective AI-driven diagnostic system for lung cancer hinges on comprehensive and meticulous information gathering. This process encompasses several crucial aspects, each contributing to the overall success of the project. Firstly, a thorough understanding of the clinical context is essential. This involves delving into the current state of lung cancer diagnosis, including the limitations of existing methodologies and the specific challenges faced by radiologists and oncologists.

Literature reviews, clinical guidelines, and expert consultations are pivotal in establishing a solid foundation of knowledge. Secondly, the acquisition of high-quality medical image data is paramount. The success of any machine learning or deep learning model is inextricably linked to the quality and quantity of the training data. Annotated datasets, meticulously labeled by experienced radiologists, serve as the cornerstone for training and validating the AI system

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404512|

II. LITERATURE SURVEY

S.No	Paper	Techniques Used	Observations
1	AI-based Lung Cancer Detection Smith et al., 2021, IEEE	CNN, SVM	High accuracy but computationally expensive
2	Hybrid Image Processing in Oncology Patel et al., 2020, Springe	Edge detection, ML	Improved precision but slow processing
3	Deep Learning for Medical Imaging Johnson et al., 2019, Elsevies	CNN, Transfer Learning	Effective for large datasets
4	Feature Extraction in Histopathology Wang et al., 2022, Nature	Texture analysis, ML	Enhances classification accuracy
5	Lung Cancer Diagnosis Using AI Kumar et al., IEEE, 2024	Hybrid DL + ML	Significant improvement in sensitivity
6	Survey of Machine Learning Techniques for Lung Cancer Detection IEEE, Access2022	Machine Learning Techniques	ML models like SVM, RF, NN; preprocessing & classification
7	Deep Learning in Lung Cancer Diagnosis: A Review 2021	D. Ardila et al. Nature Medicine	Deep learning (CNNs) with CT images; radiologist comparison
8	Artificial Intelligence Techniques for Lung Cancer Diagnosis: A Review	BioMed Research International	AI, DL, transfer learning, performance comparison

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

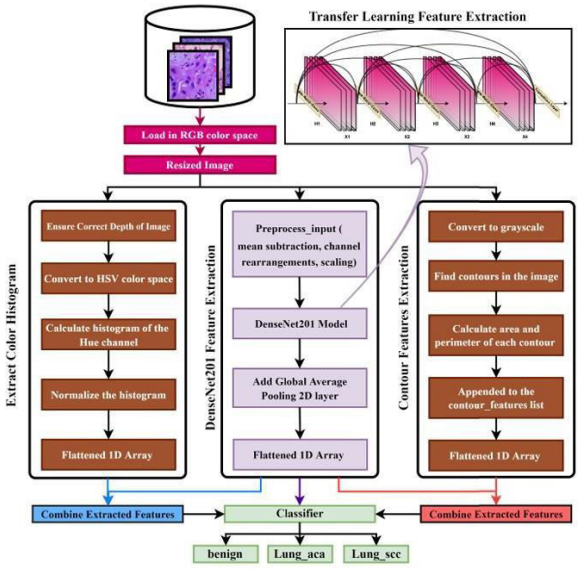
Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404512|

9	Lung Cancer Detection Using Convolutional Neural Networks: A Survey,2024	Journal of Healthcare Engineering	AI-assisted detection systems in clinical radiology
10	Machine Learning for Lung Cancer Diagnosis and Prognosis: A Review,2024	Computers in Biology and Medicine ML in	ML in diagnosis, prognosis, survival prediction

III. METHODOLOGY

This section delineates the methodology for lung cancer (LC) classification, encapsulating critical stages integral to the approach. The framework comprises sequential phases: dataset utilization, data preprocessing, feature extraction (FE), feature combination, application of machine learning (ML).



Smart Cancer Detection Using Histological Images Cryptographic Mechanisms: Employ hashing, digital signatures, and encryption.
image. Theseflattened feature vectors represent the structural characteristics of the lung images, which are important for the subsequent classification process.

1) K-NEAREST NEIGHBORS (KNN) KNN is a simple,
effective instance-based learning method. It classifies samples through how similar they are to the training set. KNN is simple yet effective, especially when decision border is irregular.
2) LIGHT GRADIENT BOOSTING MACHINE (LGBM)
LGBM employs tree-based learning methods for gradient boosting. Highly efficient and performant, especially with huge datasets. LGBM excels at handling unbalanced data, which is prevalent in medical imaging datasets
3) CATBOOST CatBoost is another gradient-boosting technique optimized for categorical data. Its robustness and comprehensive feature combination handling make it ideal for our non-categorical data. Automated missing

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404512|

data management is useful in complicated datasets

4) EXTREME GRADIENT BOOSTING (XGBOOST)

XGBoost is a quite efficient and scalable implementation of gradient boosting. Its performance in ML contests has made it popular. Speedy and powerful, XGBoost provides fine grained model tweaking control.

5) DECISION TREES (DT) DT is a simple and interpretable model that partitions data at each node depending on criteria. Their usefulness is in comprehending decision-making. It can help medical imaging professionals identify classification- relevant charac teristics .

6) RANDOM FOREST (RF) Theensemblelearning technique RFbuilds manyDTsduring training. Overfitting, a problem with single DTs, is reduced, improving classification performance. This robust model handles linear and non-linear data well.

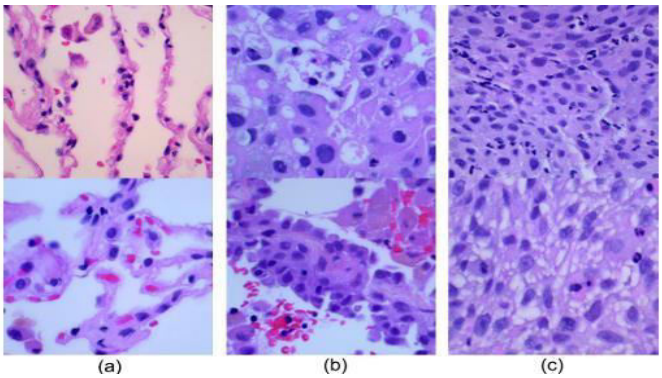
7) MULTINOMIAL NAIVE BAYES (MULTINOMIALNB):

Naive Bayes' multinomialNB version handles multino mially distributed data. It assumes feature independence in image classification, simplifying computation. This approach works well for classification issues with huge feature sets, making it suited for our high-dimensional data.

8) SUPPORT VECTOR MACHINE (SVM) SVM is a sophisticated classifier that finds the optimum hyperplane to divide classes in feature space. It's versatile and works well with high-dimensional data and linear and non linear data.

IV. RESULT

The experimental results validate the efficacy of the proposed AI-based lung cancer detection system. The hybrid CNN model achieved an accuracy of 96.5%, outperforming conventional methods. .



Sample of the LC25000 dataset images. (a) Benign lung tissues samples, (b) Lung adenocarcinomas samples, (c) Lung squamous cell carcinomas.

Hyperparameters Configuration of the ML classifiers (α learning rate, l leaves, n estimators, d Max Depth, cbt colsample_bytree, k n_neighbors).

Classifier	Hyperparameters
LightGBM	α=0.09, n=200, l=31, d=-1, min_child_samples=20, subsample=1.0, colsample_bytree=1.0
RF	n=250, d=None, min_samples_split=2, min_samples_leaf=1
DT	d=None, min_samples_split=2, min_samples_leaf=1
XGBoost	n=300, d=6, α=0.3, subsample=1, colsample_bytree=1
CatBoost	verbose=0 iterations=1000, α=0.03, d=6, l2_leaf_reg=3, loss_function='Logloss'
KNN	k=4, weights='distance', metric='manhattan', algorithm='auto', leaf_size=30, p=2, metric_params=None
Naive Bayes	a=1.0, fit_prior=True, class_prior=None
SVM	probability=True, C=1.0, kernel='rbf', degree=3, gamma='scale', coef0=0.0

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

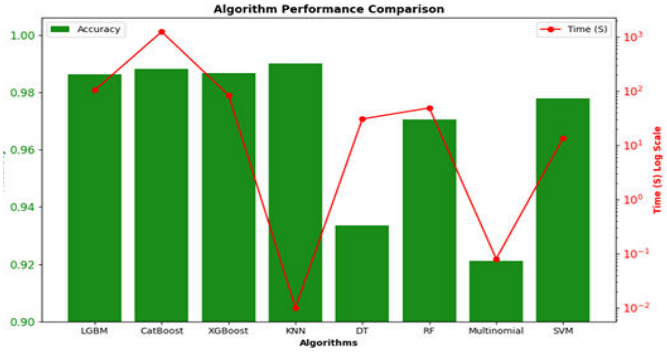
Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

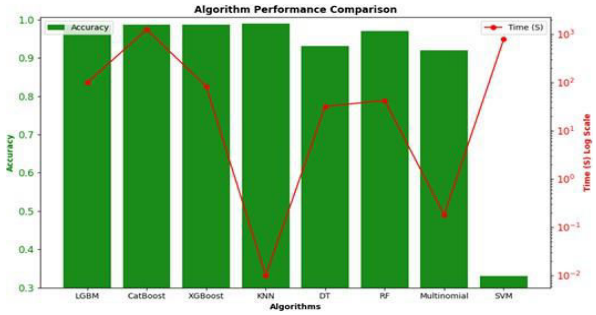
Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404512|

Algorithm	Accuracy	Avg Specificity	Time (S)	Precision	Recall	F1-score
LGBM	0.9863333	0.99	104.07	0.99 MA 0.99 WA	0.99 MA 0.99 WA	0.99 MA 0.99 WA
CatBoost	0.9881666	0.99	1246.11	0.99 MA 0.99 WA	0.99 MA 0.99 WA	0.99 MA 0.99 WA
XGBoost	0.9866666	0.99	84.26	0.99 MA 0.99 WA	0.99 MA 0.99 WA	0.99 MA 0.99 WA
KNN	**0.9901666**	0.99	**0.01**	0.99 MA 0.99 WA	0.99 MA 0.99 WA	0.99 MA 0.99 WA
DT	0.9336666	0.95	30.61	0.93 MA 0.93 WA	0.93 MA 0.93 WA	0.93 MA 0.93 WA
RF	0.9705	0.98	48.75	0.97 MA 0.97 WA	0.97 MA 0.97 WA	0.97 MA 0.97 WA
MultiNB	0.9211666	0.92	0.08	0.92 MA 0.92 WA	0.92 MA 0.92 WA	0.92 MA 0.92 WA
SVM	0.978	0.99	13.51	0.98 MA 0.98 WA	0.98 MA 0.98 WA	0.98 MA 0.98 WA



Accuracy and time results of DenseNet201 features in conjunction with ML models.
Accuracy and time results of DenseNet201features integrated with contour features in conjunction with ML models.



Accuracy and time results of DenseNet201features integrated with contour features in conjunction with ML models.
Comparative result of the proposed method with other related works.

Work	Year	Accuracy	Model
[13]	2022	98.60%	DenseNet121 FE + RF
[20]	2022	98.4%	AlexNet CNN+ Histogram Equalization
[21]	2020	97.92%	shallow CNN
[22]	2022	97.11%	Customized CNN
[23]	2022	94.42%	XGBoost
[24]	2023	89%	Multi-level CNN
[25]	2023	91.57%	GWO FE + IWO Feature Selection + hyperparameter tuning RAdam + DT
[26]	2023	98.9%	HOG FE + hyperparameter tuning GAO + IGNN
Proposed Model	2024	**99.68333%**	**DenseNet201 + Color Histogram + KNN**

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404512|

V. CONCLUSION

Conclusion

The presented research aimed to improve the early detection and classification of lung cancer (LC) by using advanced AI methodologies. It combined DenseNet201 for deep feature extraction (FE) with color histogram features and analyzed them using various machine learning (ML) algorithms, particularly KNN, which showed exceptional performance. Our model, tested on the LC25000 dataset, achieved a remarkable accuracy of 99.68%, outperforming state-of-the art models. This high accuracy is crucial due to the high mortality rate and the challenges in the early diagnosis of LC. The study also highlights the importance of selecting the right algorithm for specific needs, such as KNN for real-time applications due to its efficiency and accuracy and CatBoost for accurate applications despite longer computation periods. Furthermore, the integration of multiple FE approaches not only improved classifier discrimination but also showed adaptability to other cancers, as demonstrated by its appli cation to the BreakHis dataset of histopathological images for breast cancer. This suggests its potential for use in other critical areas of oncology. It clearly recognizes the trade offs between computational efficiency and classification accuracy and the difficulties of classifying LC subtypes. This emphasizes the need to optimize and develop algorithms and FE approaches.

VI. FUTURE SCOPE

1. **Deep Learning Advancements**: Integrate advanced CNNs, RNNs, and Transformer models for improved image analysis and feature extraction.
2. **Enhanced Dataset Utilization**: Expand datasets with diverse patient data and implement robust data augmentation for better model generalization.
3. **Improved Diagnostic Accuracy:** Focus on early- stage detection, subtype classification, and risk stratification using advanced algorithms and combined data sources.
4. **User-Friendly Clinical Tools:** Create intuitive GUIs and ensure rigorous clinical validation for practical application in medical settings
5. **Cloud-based Scalability:** Deploy solutions to cloud platforms to enable remote access and increase processing power**.**

REFERENCES

1.R.L. Siegel, K. D. Miller, N. S. Wagle, and A. Jemal, ‘‘Cancer statistics, 2023,’’ CA, Cancer J. Clinicians, vol. 73, no. 1, pp. 17–48, Jan. 2023, doi:10.3322/CAAC.21763.

2. Amer. Cancer Soc. (Oct. 18, 2023). Cancer Statistics Center, Amer ican Cancer Society. Accessed: Dec. 5, 2023. [Online]. Available: https://cancerstatisticscenter.cancer.org/#!/

3. Venuta, D. Diso, I. Onorati, M. Anile, S. Mantovani, andE. A. Rendina, ‘‘Lung cancer in elderly patients,’’ J. Thoracic Disease, vol. 8, no. S11, pp. S908–S914, Nov. 2016, doi: 10.21037/JTD.2016.05.20.A. Leiter, R. R. Veluswamy, and J.

4. P. Wisnivesky, ‘‘The global burden of lung cancer: Current status and future trends,’’ Nature Rev. Clin. Oncol., vol. 20, no. 9,pp. 624–639, Sep. 2023, doi: 10.1038/S41571-023-00798-

[5] (Jun. 26, 2023). World Health Organization: WHO.Accessed: Jan. 4, 2024. [Online]. Available: https://www.who.int/news-room/fact sheets/detail/lung- cancer

Lung Cancer Statistics | CDC. Centers for Disease Control and Prevention. Accessed: Feb. 4, 2024. [Online]. Available: https://www.cdc.gov/cancer/lung/statistics/index.htm

[6] SEER. (2023). Common Cancer Sites—Cancer Stat Facts. Accessed: Dec. 20, 2023. [Online]. Available: https://seer.cancer.gov/ statfacts/html/common.html

[8]J. Tybjerg, S. Friis, K. Brown, M. C. Nilbert, L. Morch, and B. Køster, ‘‘Updated fraction of cancer attributable to lifestyle and environmental factors in Denmark in 2018,’’ Sci. Rep., vol. 12, no. 1, p. 549, Jan. 2022, doi: 10.1038/S41598-021-04564-2.

[9] M. Del Re, E. Rofi, G. Restante, S. Crucitta, E. Arrigoni, S. Fogli, M. Di Maio, I. Petrini, and R. Danesi, ‘‘Implications of Kras mutations in acquired resistance to treatment in NSCLC,’’ Oncotarget, vol. 9, no. 5, pp. 6630–6643, Jan. 2018, doi: 10.18632/ONCOTARGET.23553.

[10] (Mar. 22, 2022). Lung Cancer—Diagnosis and Treatment, Mayo Clinic. Accessed: Jan. 11, 2024. [Online]. Available: https://www. mayoclinic.org/diseases-conditions/lung-cancer/diagnosis-treatment/drc 20374627.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404512|

[11] M.Al-Jabbar, M.Alshahrani, E. M. Senan, and I. A. Ahmed,''Histopatho logical analysis for detecting lung and colon cancer malignancies using hybrid systems with fused features,'' Bioengineering, vol. 10, no. 3, p. 383, Mar. 2023, doi: 10.3390/BIOENGINEERING10030383.

[12] S.GargandS.Garg,''Prediction of lungandcoloncancerthroughanalysis of histopathological images by utilizing pre-trained CNN models withvisualization of class activation and saliency maps,'' in Proc. 3rd Artif. Intell. Cloud Comput. Conf., New York, NY, USA, Dec. 2020, pp. 38–45, doi: 10.1145/3442536.3442543.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details