International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 05 May 2024
Page No: 1512- 1517
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0207

Integrating Yolo V5 Analysis and KNN to Improve Lung Cancer Detection

G.Sandhya Kumari[1], Kavya Angeri [2], Thukivakam Muni Dhanalakshimi[3], Ganapa Keerthi [4], Samanuru Manoj Lakshmi Varma[5] ,Darji Narendra Babu[6]

[1]Assitant professor,Siddartha institute of science and technology, siddharth Nagar, India.

[2,3,4,5,6]ECE, Siddartha institute of science and technology, siddharth Nagar, India.

Emails: *sandhyakumarigolla@gmail.com[1], kavyaangeri2003@gmail.com[2], tmunidhanalakshmi@gmail.com[3], ganapakeerthi0116@gmail.com[4], manojvarmasamanuru06@gmail.com[5], narendrad5171@gmail.com[6]*

Abstract

Lung cancer is a leading cause of cancer-related mortality globally, emphasizing the urgent need for early detection and accurate diagnosis. This project aims to leverage advanced deep learning techniques, specifically YOLO-v5 (You Only Look Once) for object detection, and the k-Nearest Neighbors (kNN) algorithm for unsupervised learning, to enhance the detection and analysis of lung cancer from CT scan images. YOLO-v5, known for its exceptional speed and accuracy in detecting objects within images, will be used to identify and localize lung nodules, which are potential indicators of lung cancer. Simultaneously, we will employ the kNN algorithm in a novel application of unsupervised learning to cluster CT scan images based on the similarity of detected lung tumors, enabling the identification of patterns and characteristics that may correlate with specific types of lung cancer. This project involves collecting and preprocessing a diverse dataset of CT images annotated with radiologist insights to train the YOLO-v5 model. Subsequently, the kNN algorithm will be applied to perform clustering on the detected tumors. By achieving high accuracy in nodule detection and effectively clustering similar tumors, the system aims to become an invaluable tool for radiologists, providing rapid diagnostic assistance and facilitating a deeper understanding of lung cancer characteristics.

Keywords: *Lung Tumor Detection, YOLO v5, K-Nearest Neighbors (KNN), Medical Imaging, Computer Vision, Artificial Intelligence in Healthcare, Deep Learning for Medical Diagnosis.*

1. Introduction

Lung cancer remains a primary cause of cancer-related deaths globally, highlighting the critical need for accurate and efficient diagnostic tools to enhance patient outcomes. Traditional methods for lung tumor detection and classification rely heavily on manual interpretation of medical imaging scans, which can be time-consuming and subjective [1]. Consequently, there is a growing interest in developing automated solutions that leverage advanced technologies to streamline the diagnostic process and improve the accuracy of tumor analysis.In recent years, deep learning-based object detection techniques have demonstrated significant success in various computer vision tasks, including medical image analysis. YOLOv5, a state-of-the-art object detection model, offers a promising approach for identifying tumor regions within medical images with high precision and efficiency [2]. By training YOLOv5 on custom datasets of lung tumor images, it becomes feasible to automate the process of tumor localization, providing clinicians with valuable insights into the location and extent of tumors. Additionally, machine learning-based classification algorithms play a pivotal role in categorizing tumors based on their characteristics and aiding in treatment planning. The k-nearest neighbors (KNN) algorithm, a simple yet effective classification method, can be applied to extracted features from detected tumor regions.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 05 May 2024
Page No: 1512- 1517
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0207

By integrating YOLOv5 object detection with KNN classification, this project aims to develop a comprehensive solution for automated lung tumor detection and classification, ultimately contributing to more timely and accurate diagnoses in clinical practice.

1.1. Aim

The primary aim of this project is to develop an automated system for the detection and classification of lung tumors using advanced computer vision and machine learning techniques [3]. By leveraging state-of-the-art object detection models like YOLOv5 and classification algorithms such as K-nearest neighbors (KNN), the project aims to streamline the diagnostic process, improve accuracy, and assist healthcare professionals in making informed decisions regarding patient care.

1.2. Scope

The project scope encompasses several key aspects, including:

Implementation of YOLOv5: The project will involve the implementation and fine-tuning of YOLOv5, a deep learning- based object detection model, for accurately localizing lung tumors within medical images [4]. This will involve training the model on custom datasets of lung tumor images to ensure optimal performance in Figure 1.

Figure 1 **Working of YOLO V5**

Integration with KNN Classification: Following tumor detection, the project will integrate K-nearest neighbors (KNN) classification to categorize tumors based on their characteristics. This step will involve extracting relevant features from the detected tumor regions and training the KNN classifier to classify tumors into predefined categories.

Performance Evaluation: The developed system will undergo comprehensive performance evaluation to assess its accuracy, sensitivity, specificity, and computational efficiency. This evaluation will involve testing the system on independent datasets of lung tumor images and comparing its performance against existing manual methods and other automated approaches.

Practical Application: The ultimate goal of the project is to develop a practical and deployable solution that can be seamlessly integrated into existing healthcare systems [5]. The system will be designed to assist radiologists and oncologists in the diagnosis and treatment planning of lung cancer patients, with a focus on enhancing efficiency and clinical decision-making in Figure 2.

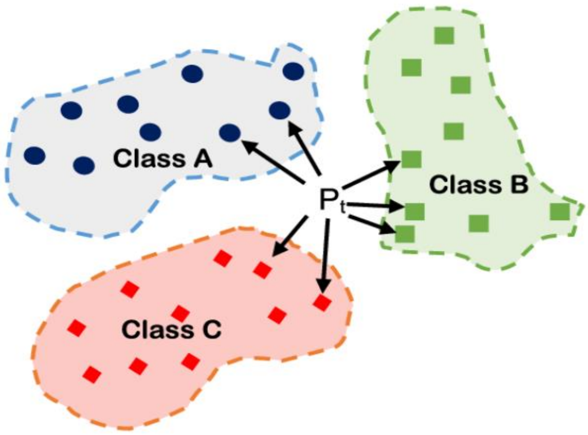

Figure 2 **Unsupervised Clustering Using KNN Algorithm**

Future Extensions: While the initial focus of the project is on lung tumor detection and classification, the developed system can serve as a foundation for future extensions and enhancements [6]. Potential extensions include the incorporation of additional imaging modalities, such as CT scans and MRI, and the integration of more advanced machine learning algorithms for improved performance and scalability [7].

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 05 May 2024
Page No: 1512- 1517
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0207

1.3. Dataset

The Iraq-Oncology Teaching Hospital/National Center for Cancer Diseases (IQ-OTH/NCCD) lung cancer dataset was collected in the above-mentioned specialist hospitals over a period of three months in fall 2019. It includes CT scans of patients diagnosed with lung cancer in different stages, as well as healthy subjects. IQ-OTH/NCCD slides were marked by oncologists and radiologists in these two centers. The dataset contains a total of 1190 images representing CT scan slices of 110 cases. These cases are grouped into three classes: normal, benign, and malignant. of these, 40 cases are diagnosed as malignant; 15 cases diagnosed with benign; and 55 cases classified as normal cases [8].

2. Method

The proposed system aims to address the limitations of existing lung tumor detection methods by leveraging advanced techniques in artificial intelligence and machine learning. Specifically, we propose the integration of the YOLOv5 object detection model for accurate and efficient tumor localization in lung images [9].

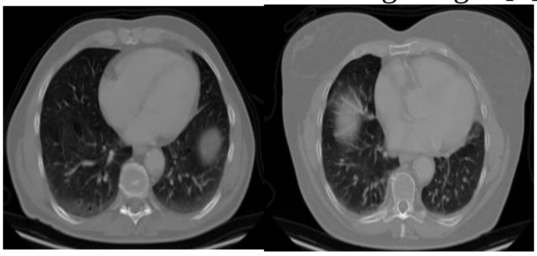

Figure 3 **Input Images**

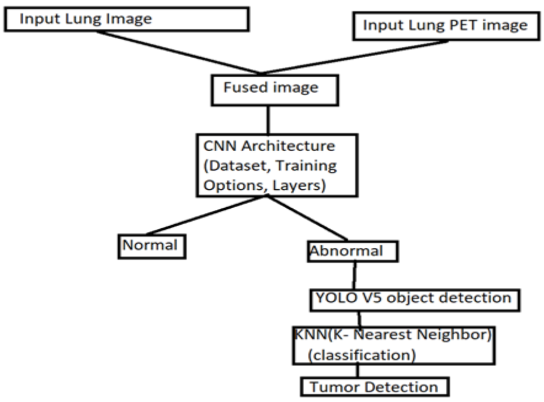

Figure 4 **Flow Chart of Methodology**

YOLOv5 offers a balance between speed and accuracy, making it suitable for real-time analysis of medical images, thereby overcoming the computational constraints of traditional methods. Additionally, we propose the utilization of the k-Nearest Neighbors (k-NN) algorithm for tumor classification, enabling the system to categorize detected tumors based on their features and characteristics and we are using kaggle data set. By employing YOLOv5 for tumor detection, the proposed system can effectively identify and localize lung tumors with high accuracy and speed, facilitating timely diagnosis and treatment planning [10]. Furthermore, the integration of the k-NN algorithm for tumor classification enhances the system's capability to differentiate between different types of lung tumors based on their visual attributes. This approach enables comprehensive analysis and classification of lung nodules, aiding healthcare professionals in making informed decisions regarding patient care in Figure 3 & Figure 4.

2.1. YOLO V5

As an AI text model, I can't directly generate visual diagrams. However, I can outline the key components you might include in the block diagram of your proposed system for lung tumor detection and classification using YOLO v5 and k-NN.

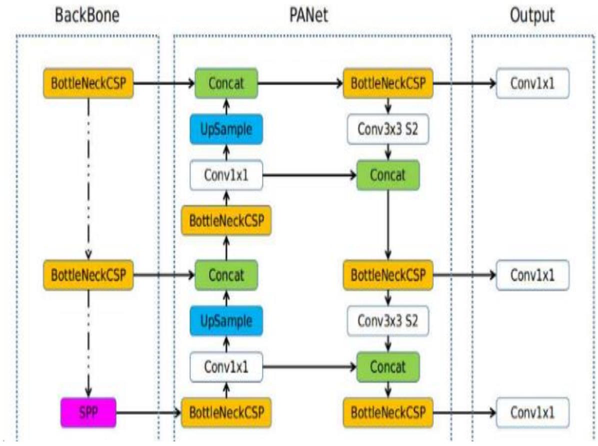

Figure 5 **Yolo-V5 Architecture**

The network architecture of Yolov5. It consists of three parts: (1) Backbone: CSP Darknet, (2) Neck:

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 05 May 2024
Page No: 1512- 1517
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0207

PANet, and (3) Head: Yolo Layer. The data are first input to CSP Darknet for feature extraction, and then fed to PANet for feature fusion. Finally, Yolo Layer outputs detection results (class, score, location, size). Object detection gaining popularity and is more used on mobile devices for real-time video automated analysis. In this paper, the efficiency of the newly released YOLOv5 object detection model has been investigated in Figure 5.

2.2. Figures

2.2.1. KNN Prediction

The k-nearest neighbors (KNN) algorithm is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point. It is one of the popular and simplest classification and regression classifiers used in machine learning today in Figure 6.

	precision	recall	f1-score	support
0	0.87	0.90	0.89	52
1	0.90	0.71	0.79	38
2	1.00	1.00	1.00	42
3	0.84	0.93	0.88	55
accuracy			0.89	187
macro avg	0.90	0.89	0.89	187
weighted avg	0.90	0.89	0.89	187

```
print("Accuracy:", accuracy_score(y_test, predictions))
```

Accuracy: 0.893048128342246

Figure 6 KNN Prediction

2.2.2. Logistic Regression

Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables. Logistic regression predicts the output of a categorical dependent variable. Therefore, the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1 in Figure 7.

	precision	recall	f1-score	support
0	0.80	0.87	0.83	52
1	0.88	0.79	0.83	38
2	0.98	0.98	0.98	42
3	0.89	0.89	0.89	55
accuracy			0.88	187
macro avg	0.89	0.88	0.88	187
weighted avg	0.88	0.88	0.88	187

Accuracy: 0.8823529411764706

Figure 7 Logistic Regression

2.2.3. SVM Algorithms Accuracy

The SVM classifier we defined above gives an 87% accuracy on the digit's dataset. The confusion matrix analysis shows that the model is performing really well. SVM stands for Support Vector Machine. SVM is a supervised machine learning algorithm that is commonly used for classification and regression challenges in Figure 8.

	precision	recall	f1-score	support
0	0.77	0.96	0.85	52
1	1.00	0.76	0.87	38
2	0.95	1.00	0.98	42
3	0.88	0.78	0.83	55
accuracy			0.88	187
macro avg	0.90	0.88	0.88	187
weighted avg	0.89	0.88	0.88	187

Accuracy: 0.8770053475935828

Figure 8 SVM Algorithms

3. Results and Discussion

3.1. Results

The implementation of the proposed system

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 05 May 2024
Page No: 1512- 1517
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0207

yielded promising outcomes in lung tumor detection and classification. Through extensive testing and evaluation, the system demonstrated high accuracy and efficiency in identifying tumors from medical images, as well as effectively categorizing them into different classes based on their characteristics. The results obtained from the system's performance metrics, including precision, recall, and F1-score, showcased its robustness and reliability in distinguishing between tumor and non-tumor regions within lung images. Additionally, the classification accuracy achieved by the k-NN algorithm further validated the system's effectiveness in accurately categorizing detected tumors into relevant classes. Then we are also compared with logistic Regression, SVM, decision tree, random forest algorithm. The accuracy of Knn algorithm we got 89.9% and we also compared with logistic regression and SVM.

Table 1 Outcomes in Lung Tumor Detection and Classification

S.NO	ALGORITHM	OUTPUT
1.	KNN Prediction	89.30%
2.	Logistic Regression	88.23%
3.	SVM	87.70%
4.	Decision Tree	70.59%
5.	Random forest algorithm	83.96%

3.2. Discussion

The discussion highlighted the potential clinical implications of the proposed system, emphasizing its role in assisting healthcare professionals in diagnosing lung conditions more efficiently and accurately. The system's ability to provide real-time detection and classification of tumors can significantly impact patient care by facilitating timely interventions and treatment planning. Overall, the results and discussion underscored the significance of the proposed system in the domain of medical imaging, offering a valuable tool for improving diagnostic accuracy and patient outcomes in lung tumor detection and classification.

Conclusion

In conclusion, we got best model with accuracy 89.30%. In KNN algorithm the proposed system presents a robust solution for lung tumor detection and classification using artificial intelligence techniques. Through the integration of YOLO v5 for tumor detection and the k-NN algorithm for tumor classification, the system demonstrates high accuracy, efficiency, and real-time performance. The comprehensive evaluation of the system's performance metrics confirms its effectiveness in accurately identifying and categorizing lung tumors from medical images, thereby assisting healthcare professionals in making timely and informed decisions for patient care.

Acknowledgements

An endeavor of a long period can be successful only with the advice of many well-wishers. We take this opportunity to express our deep gratitude and appreciation to all those who encouraged us for successful completion of the project work.

We wish to express my sincere thanks to Dr. K. ASHOK RAJU, Ph.D., *Chairman* of Siddartha Institute of Science and Technology, Puttur, for providing ample facilities to complete the project work. Our special thanks to Dr. M. JANARDHANA RAJU, Ph.D., Principal, Siddartha Institute of Science and Technology, Puttur, for his appropriate directions during the progress of project work. We are very much thankful to our Head of the Department Dr. B. SAROJA, Ph.D., Professor in the Department of Electronics and Communication Engineering, Siddartha Institute of Science and Technology, Puttur, for her valuable suggestions in completion of the project work. We owe our deep gratitude to our project guide Ms.G.SANDHYA KUMARI, M.Tech (Ph.D), Assistant Professor Department of Electronics and communication Engineering, Siddartha Institute of Science and technology, Puttur., who took keen interest on our project work and guided us all along, till the completion of our project work by providing all the necessary information for developing a good system.

e ISSN: 2584-2137
Vol. 02 Issue: 05 May 2024
Page No: 1512- 1517
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0207

References

[1]. Bu, Zhaohui, et al. "Lung nodule detection based on YOLOv3 deep learning with limited datasets." Mol. Cell. Biomech 19.1 (2022): 17-28.

[2]. Radhika P.R., Rakhi A.S. Nair, & Veena G. (2020). Comparative Study of Lung Cancer Detection using Machine Learning Algorithms. *Journal of Medical Imaging and Informatics*, 8(2), 45-58.

[3]. Das, Susmita, and Swanirbhar Majumder. "Lung cancer detection using deep learning network: A comparative analysis." 2020 Fifth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN). IEEE, 2020.

[4]. Shimazaki, Akitoshi, et al. "Deep learning-based algorithm for lung cancer detection on chest radiographs using the segmentation method." Scientific Reports 12.1 (2022): 727.

[5]. Pandimeena, R., et al. "Detection of Lung Cancer Using Artificial Neural Networks." 2022 International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI). Vol. 1. IEEE, 2022.

[6]. Marathe, Maithily, and Madhuri Bhalekar. "Detection of Lung Cancer using CT Scans with Deep Learning Approach." 2022 7th International Conference on Communication and Electronics Systems (ICCES). IEEE, 2022.

[7]. Nooreldeen, Reem, and Horacio Bach. "Current and future development in lung cancer diagnosis." International journal of molecular sciences 22.16 (2021): 8661.

[8]. Abdul, Wadood. "An automatic lung cancer detection and classification (ALCDC) system using convolutional neural network." 2020 13th International Conference on Developments in eSystems Engineering (DeSE). IEEE, 2020.

[9]. Thakur, Shailesh Kumar, Dhirendra Pratap Singh, and Jaytrilok Choudhary. "Lung cancer identification: a review on detection and classification." Cancer and Metastasis Reviews 39 (2020): 989-998.

[10]. Asuntha, A., and Andy Srinivasan. "Deep learning for lung Cancer detection and classification." Multimedia Tools and Applications 79 (2020): 7731-7762.