EARLY DETECTION OF LUNG CANCER USING DEEP LEARNING

Mr. SHAIK MAHABOOB BASHA, MALA ARCHANA, LAKAVATH THARUN KUMAR, NEERADI SWAMY RAJ, PANAGANTI SRAVAN KUMAR

Assistant professor

Department of Computer Science and Engineering

Tkr college of Engineering and Technology, Med boli, Meerpet, Balapur, Hyderabad, Telangana 500097, India

ABSTRACT

Lung cancer remains one of the leading causes of mortality worldwide, primarily due to late-stage diagnosis and the complexity involved in detecting early pathological changes. This project presents an automated system for the early detection and classification of lung cancer using deep learning techniques applied to histopathological images. The proposed system utilizes a Convolutional Neural Network (CNN) based on the VGG-19 architecture to classify lung tissue images into three categories: benign, adenocarcinoma, and squamous cell carcinoma. Unlike traditional machine learning approaches such as Support Vector Machines and Random Forests, which rely on handcrafted features and often yield limited accuracy, the CNN model automatically extracts hierarchical features from images, enabling more precise classification. The methodology includes data collection, preprocessing (such as normalization and augmentation), model training, and performance evaluation using metrics like accuracy and confusion matrix analysis. The results demonstrate that the VGG-19 model achieves superior performance in detecting and classifying lung cancer types, with improved accuracy and reduced misclassification rates. Additionally, a user-friendly interface is developed to allow medical professionals to upload images and obtain real-time predictions, thereby assisting in faster and more reliable diagnosis. Overall, the system highlights the potential of deep learning in enhancing early lung cancer detection and supporting clinical decision-making.

Keywords: Lung Cancer Detection, Deep Learning, Convolutional Neural Network (CNN), VGG-19, Histopathological

Images, Medical Image Classification, Computer-Aided Diagnosis (CAD), Artificial Intelligence in Healthcare

INTRODUCTION

Lung cancer is one of the most life-threatening diseases worldwide and remains a leading cause of cancer-related deaths. The primary reason for the high mortality rate is the late-stage detection of the disease, which significantly reduces the chances of effective treatment and patient survival. Early and accurate diagnosis plays a crucial role in improving clinical outcomes and enabling timely medical intervention. However, conventional diagnostic methods, such as manual examination of histopathological images, are time-consuming, require expert pathologists, and are often subject to human error.

With the rapid advancement of artificial intelligence and deep learning techniques, automated medical image analysis has gained significant attention in recent years. Among these techniques, Convolutional Neural Networks (CNNs) have emerged as powerful tools for image classification tasks due to their ability to automatically learn complex features directly from raw data. Unlike traditional machine learning methods that rely on handcrafted features, CNNs can effectively capture subtle variations in texture, shape, and structural patterns within lung tissue images.

Previous studies have applied various machine learning algorithms, such as Support Vector Machines (SVM), Random Forests (RF), and Bayesian Networks, for lung cancer detection. However, these methods often struggle to achieve high accuracy when dealing with complex and high-dimensional medical image data. Additionally, many existing systems focus only on binary classification (cancerous vs. non-cancerous), which limits their practical applicability in clinical diagnosis where identifying specific cancer types is essential for treatment planning.

To address these challenges, this study proposes a deep learning-based approach using the VGG-19 CNN architecture for multi-class classification of lung cancer. The system is designed to classify histopathological images into three categories: benign, adenocarcinoma, and squamous cell carcinoma. The proposed approach includes key stages such as data collection, preprocessing, model training, and performance evaluation. By leveraging the hierarchical feature learning capability of deep neural networks, the model aims to improve diagnostic accuracy and reduce misclassification rates.

Furthermore, the developed system incorporates an automated prediction interface that allows users to upload medical images and receive real-time diagnostic results. This not only reduces the workload on healthcare professionals but also enhances the speed and reliability of diagnosis. Overall, the proposed system demonstrates the potential of deep learning techniques in supporting medical decision-making and improving early detection of lung cancer.

LITERATURE REVIEW

Early research in lung cancer detection primarily relied on classical image processing techniques to identify abnormal patterns in medical images. Studies by Kawahara and Harmane highlighted the importance of early diagnosis and explored traditional approaches for detecting lung abnormalities. These methods focused on basic image enhancement and segmentation techniques but lacked robustness when dealing with complex histopathological data.

Subsequently, preprocessing techniques such as noise removal and filtering were introduced to improve image quality. S. C. Wang emphasized the use of median and Wiener filters to enhance histopathological and radiological images, enabling better feature extraction. While these approaches improved image clarity, they still depended heavily on manual feature engineering.

Rule-based and statistical models were later developed to mimic expert decision-making. Perumal and Velmurugan proposed computer-aided diagnosis (CAD) systems that used predefined rules and statistical analysis for classification. However, these systems were limited by their inability to generalize across diverse datasets and required significant domain expertise.

Machine learning techniques such as Support Vector Machines (SVM), Random Forests (RF), and Bayesian Networks were introduced to address these limitations. R. Yan, F. Ren, and Y. Wang applied SVM for lung nodule classification, while Shalini Wankhade explored RF and Bayesian models for cancer classification. These models improved classification performance compared to rule-based systems but relied on handcrafted features, which restricted scalability and accuracy.

With the advancement of deep learning, Convolutional Neural Networks (CNNs) emerged as a powerful alternative for automated feature extraction. Abdul Rahaman and Wahab Sait demonstrated improved performance using deep learning architectures such as DenseNet-121 combined with preprocessing and augmentation techniques. Similarly, Rabia Javed evaluated deep learning models on PET/CT datasets, highlighting their effectiveness and reliability for real-time clinical applications.

Deep Convolutional Neural Networks (DCNNs) further enhanced feature learning by utilizing multi-layered architectures capable of capturing complex patterns in medical images. Kadam emphasized that DCNNs outperform traditional algorithms across metrics such as accuracy, sensitivity, and specificity. Additionally, Faria and Campelos evaluated lightweight models such as MobileNetV3, which provided high classification performance with reduced computational cost, making them suitable for resource-constrained environments.

Recent advancements include transfer learning and grouped convolution techniques to improve efficiency and reduce computational complexity. Ahmad Chaddad explored these approaches, demonstrating improved classification accuracy and reduced training time.

Despite these advancements, challenges such as dataset imbalance, variability in histopathological images, and the need for multi-class classification remain. Most existing studies focus on binary classification, which limits their applicability in clinical diagnosis. Therefore, there is a need for a robust deep learning-based system capable of accurately classifying multiple types of lung cancer.

EXISTING SYSTEM

Existing systems for lung cancer detection primarily rely on traditional machine learning and statistical techniques for analyzing medical images. These approaches typically involve manual feature extraction, where domain experts identify relevant features such as texture, shape, and intensity from histopathological or radiological images. Commonly used algorithms include Support Vector Machines (SVM), Random Forests (RF), and Bayesian Networks, which are applied to classify images into cancerous and non-cancerous categories. Although these methods have contributed to the development of computer-aided diagnosis (CAD) systems, they suffer from several limitations. First, their dependence on handcrafted features restricts their ability to capture complex patterns present in high-dimensional medical images. As a result, their performance often varies depending on the quality of feature engineering and the expertise of the practitioner.

Second, many existing systems are limited to binary classification, distinguishing only between benign and malignant cases. This limitation reduces their clinical usefulness, as accurate identification of specific cancer types—such as adenocarcinoma and squamous cell carcinoma—is essential for effective treatment planning.

PROPOSED SYSTEM

To overcome the limitations of traditional approaches, this study proposes an automated lung cancer detection and classification system based on deep learning techniques. The system utilizes a Convolutional Neural Network (CNN) architecture, specifically the VGG-19 model, to analyze histopathological images and accurately classify them into three categories: benign, adenocarcinoma, and squamous cell carcinoma.

The proposed system follows a structured workflow consisting of multiple stages, including data collection, preprocessing, model training, prediction, and performance evaluation. Initially, a dataset of histopathological lung images is collected and organized into labeled classes. Preprocessing techniques such as image resizing, normalization, and data augmentation (including rotation and flipping) are applied to improve dataset quality and enhance model generalization.

The core component of the system is the VGG-19 CNN model, which is designed to automatically extract hierarchical features from input images. Unlike traditional machine learning methods that rely on handcrafted features, the deep learning model learns complex patterns directly from the data. This enables the system to capture subtle variations in tissue structure and texture, leading to improved classification accuracy.

During the training phase, the dataset is divided into training and testing sets, allowing the model to learn patterns and evaluate its performance effectively. The trained model is then used to predict the class of new input images. Performance metrics such as accuracy and confusion matrix are used to assess the effectiveness of the model, ensuring reliable classification results.

In addition, a user-friendly interface is developed to facilitate interaction with the system. Users can upload histopathological images through the interface, and the system provides real-time predictions along with classification results. This feature enhances usability and allows medical professionals to quickly obtain diagnostic insights.

Overall, the proposed system provides a robust, efficient, and scalable solution for early lung cancer detection. By leveraging the capabilities of deep learning and automated feature extraction, it significantly improves diagnostic accuracy, reduces dependency on manual analysis, and supports faster clinical decision-making.

SYSTEM ARCHITECTURE

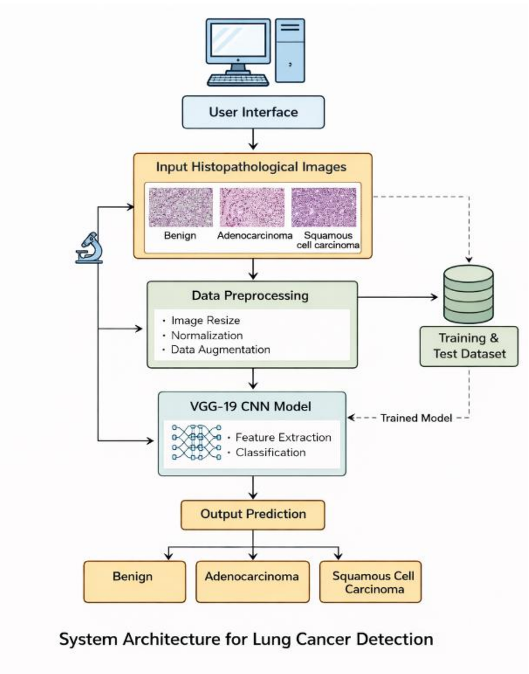

Fig:4.1SystemArchitecture

1. User Interface for image upload and result display
2. Image Input Module to accept medical images
3. Dataset Module with labeled training and testing data
4. Data Preprocessing (resizing, normalization, augmentation)
5. Feature Extraction using CNN layers
6. VGG-19 Model for deep learning-based classification
7. Model Training with training dataset
8. Prediction Module for classifying new images
9. Performance Evaluation using accuracy and confusion matrix
10. Output Module to display final results

RESULTS AND DISCUSSIONS

The proposed lung cancer detection system was evaluated using histopathological image data to assess its effectiveness in classifying different types of lung cancer. The experimental results demonstrate that the deep learning-based approach, particularly the VGG-19 Convolutional Neural Network (CNN), achieves significant improvements in classification performance compared to traditional machine learning models. The system was trained and tested on labeled datasets, and the evaluation was carried out using standard performance metrics such as accuracy and confusion matrix.

The results indicate that the VGG-19 model effectively learns complex patterns and features from histopathological images, enabling accurate differentiation between benign, adenocarcinoma, and squamous cell carcinoma classes. Compared to existing systems that rely on handcrafted features, the proposed model automatically extracts hierarchical features, leading to higher accuracy and reduced misclassification rates. The training and validation curves, as shown in the graphs on page 14, illustrate a steady improvement in accuracy and a decrease in loss over epochs, indicating that the model is well-optimized and does not suffer significantly from overfitting. .

CONCLUSION

This paper presents a deep learning-based approach for the early detection and classification of lung cancer using histopathological images. The proposed system utilizes a Convolutional Neural Network (CNN) based on the VGG-19 architecture to classify lung tissue images into three categories: benign, adenocarcinoma, and squamous cell carcinoma. The system effectively addresses the limitations of traditional machine learning methods by eliminating the need for manual feature extraction and enabling automatic learning of complex patterns from medical images.

The experimental results demonstrate that the proposed model achieves high classification accuracy and improved performance compared to conventional techniques. The use of preprocessing methods such as normalization and data augmentation further enhances the robustness and generalization capability of the model. Additionally, the confusion matrix analysis confirms that the system produces a high number of correct predictions with minimal misclassification.

The integration of a user-friendly interface enables real-time prediction, making the system practical for clinical applications. By reducing diagnosis time and minimizing human error, the proposed approach can assist medical professionals in making faster and more accurate decisions. Overall, the study highlights the effectiveness of deep learning techniques in improving early lung cancer detection and supporting healthcare systems.

References

[1] J. Kawahara and G. Hamarneh, "Multi-resolution-tract CNN with hybrid pretrained and skin-lesion trained layers," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 1, pp. 210–218, Jan. 2017.

[2] S. C. Wang, "Artificial neural network," *Interdisciplinary Computing in Java Programming*, pp. 81–100, 2003.

[3] T. Perumal and T. Velmurugan, "Lung cancer detection using image processing techniques," *International Journal of Engineering Research and Applications*, vol. 6, no. 3, pp. 23–27, 2016.

[4] R. Yan, F. Ren, Y. Wang, et al., "Breast cancer histopathological image classification using SVM," *IEEE Access*, vol. 7, pp. 179665–179675, 2019.

[5] S. Wankhade and P. Kshirsagar, "Lung cancer detection using Random Forest approach," *International Journal of Computer Applications*, vol. 160, no. 3, pp. 25–29, 2017.

[6] A. Rahaman and W. Sait, "Deep learning approach for lung cancer detection using histopathological images," *Procedia Computer Science*, vol. 167, pp. 1240–1249, 2020.

[7] R. Javed, "Lung cancer detection using deep learning and PET/CT imaging," *Journal of Medical Imaging and Health Informatics*, vol. 10, no. 5, pp. 1021–1028, 2020.

[8] S. Kadam, "Deep convolutional neural networks for lung cancer detection," *International Journal of Advanced Research in Computer Science*, vol. 9, no. 2, pp. 45–50, 2018.

[9] K. Liu et al., "Deep learning-based lung cancer classification using ResNet architecture," *IEEE Access*, vol. 6, pp. 39539–39548, 2018.

[10] N. Faria and D. Campelos, "Classification of lung cancer using machine learning algorithms," *Procedia Computer Science*, vol. 121, pp. 454–461, 2017.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.

[12] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[13] G. Litjens et al., "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[14] O. Russakovsky et al., "ImageNet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.

[15] H. Greenspan, B. van Ginneken, and R. M. Summers, "Guest editorial: Deep learning in medical imaging," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1153–1159, May 2016.