ISSN: 2755-0141

Japan Journal of Clinical & Medical Research

Research Article

Open Access

Enhanced Detection and Prediction of Lung Cancer using CNN and RNN Techniques on Text Dataset

Bodicherla Siva Sankar[1] and Asadi Srinivasulu[2*]

[1]Assistant Professor, Department of IT, Institute of Aeronautical Engineering, Dundigal, Hyderabad, Telangana, Pincode: 500043, India

[2]Visiting Academic, Cooperative Research Centre for Contamination Assessment and Remediation of the Environment (crcCARE), Global Centre for Environmental Remediation/College of Engineering, Science & Environment, ATC Building, The University of Newcastle, Callaghan NSW 2308, Australia

ABSTRACT

Lung cancer remains one of the leading causes of cancer-related mortality globally, highlighting the urgent need for precise and reliable detection methods. This research introduces a novel approach to improving lung cancer detection and prediction by combining Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) on an extensive text dataset. Our methodology tackles several common issues in current deep learning techniques used for lung cancer detection. A major problem is the dependence on small sample sizes, which often results in overfitting and limited generalization. To address this, we applied extensive data augmentation and transfer learning techniques, improving the model's ability to perform well on new, unseen data. Additionally, we reduced the over-reliance on previous studies by developing innovative model architectures that merge CNNs' ability to extract spatial features with RNNs' capability to capture temporal dependencies, creating a more robust predictive framework. We also emphasized the importance of validating our model on independent datasets by rigorously testing it on a variety of external datasets, ensuring its robustness and generalizability. To address the resource-intensive nature of training deep learning models, we utilized advanced computational resources to optimize both model training and deployment efficiency. Another challenge is the potential bias in training data, which can lead to skewed predictions. We minimized these biases by carefully selecting and preprocessing our dataset, ensuring our model's applicability across different populations. Moreover, incorporating longitudinal data allowed our model to better understand disease progression, enhancing long-term outcome predictions. Our system also considers practical aspects for clinical integration, ensuring that the models can be easily adopted in healthcare settings. By maintaining high-quality, standardized imaging and expanding our dataset to include more diverse samples, we increased the model's robustness and reliability. In conclusion, our integration of CNN and RNN techniques, along with advanced data handling strategies, offers a comprehensive solution for lung cancer detection and prediction. This approach addresses critical limitations of existing methods and paves the way for more effective clinical applications.

***Corresponding author**

Asadi Srinivasulu, Visiting Academic, Cooperative Research Centre for Contamination Assessment and Remediation of the Environment (crcCARE), Global Centre for Environmental Remediation/College of Engineering, Science & Environment, ATC Building, The University of Newcastle, Callaghan NSW 2308, Australia.

Received: June 07, 2024; **Accepted:** June 14, 2024; **Published:** June 22, 2024

Keywords: Lung Cancer Detection, Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Deep Learning, Data Augmentation, Transfer Learning, Longitudinal Data, Clinical Integration

Introduction

Lung cancer continues to be a major public health concern and is one of the leading causes of cancer-related deaths globally [1]. The necessity for accurate and reliable detection methods is paramount, given the severe morbidity and mortality associated with late-stage lung cancer diagnoses. The Lung Cancer Patient Dataset offers a comprehensive overview of patient profiles, encompassing critical factors such as age, gender, smoking history, and family history of cancer [2]. This extensive dataset is crucial for understanding patient susceptibility to lung cancer and serves as the basis for developing advanced detection and prediction models. Analyzing this data helps to identify patterns and risk factors, which in turn aids in creating more effective screening and early detection strategies [1].

In recent advancements, deep learning techniques have shown significant potential in improving lung cancer detection and prediction. This research presents a novel approach that integrates Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) utilizing a comprehensive text dataset [3]. Traditional methods often struggle with challenges such as small sample sizes, which lead to overfitting and limited generalization. By employing data augmentation and transfer learning techniques, our methodology enhances the model's ability to generalize from training data to new, unseen data, thereby mitigating the common issue of overfitting [4]. Additionally, our approach combines the spatial feature extraction capabilities of CNNs with the temporal dependency capturing abilities of RNNs, resulting in a more robust and reliable predictive framework [1].

Furthermore, this study highlights the importance of validating models on independent datasets to ensure their robustness and generalizability. Advanced computational resources are employed to optimize both the training and deployment of models, addressing the resource-intensive nature of deep learning. Bias in training data, which can lead to skewed predictions, is reduced through careful dataset selection and preprocessing [5]. The inclusion of longitudinal data allows for a better understanding of disease progression, improving long-term outcome predictions. Practical considerations for clinical integration are also included, ensuring that the developed models can be seamlessly adopted in healthcare settings [6]. By maintaining high-quality, standardized imaging and expanding the dataset to encompass diverse samples, the robustness and reliability of the model are significantly enhanced [7]. This integration of CNN and RNN techniques, combined with advanced data handling strategies, provides a comprehensive solution for lung cancer detection and prediction, paving the way for more effective clinical applications [1].

Research Methodology

The methodology for enhancing lung cancer detection and prediction involved integrating Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) using a comprehensive text dataset [8]. The initial step was to compile and preprocess a detailed dataset, capturing essential patient information such as age, gender, smoking history, family history of cancer, symptoms, and diagnosis results. This dataset formed the basis for training and validating our models. To mitigate the common issue of small sample sizes, which can lead to overfitting and limited generalization, we applied extensive data augmentation techniques [9]. These techniques included generating synthetic samples and employing transfer learning with pre-trained models to enhance the learning process. This approach enabled us to create a more diverse and robust dataset, improving the model's ability to generalize to new, unseen data [1].

To create a robust predictive framework, we developed innovative model architectures that combined the strengths of CNNs and RNNs. CNNs are particularly effective at extracting spatial features from medical images, while RNNs are adept at capturing temporal dependencies in sequential data. By integrating these two neural network techniques, we constructed a hybrid model capable of handling both spatial and temporal aspects of the lung cancer dataset [10]. This integration allowed for more accurate detection and prediction of lung cancer. Additionally, we reduced reliance on previous studies by introducing novel elements to our model architecture, ensuring that our approach made significant new contributions to the field [1].

The validation process was thorough, involving testing the model on a variety of independent datasets to ensure robustness and generalizability. To address the resource-intensive nature of training deep learning models, we utilized advanced computational resources, optimizing both training and deployment efficiency [11]. Potential biases in the training data, which could lead to skewed predictions, were minimized through careful dataset selection and preprocessing. We also incorporated longitudinal data, tracking patients over time to improve the model's ability to predict disease progression and long-term outcomes [12]. Practical considerations for clinical integration were included, ensuring that the models could be seamlessly adopted in healthcare settings. By maintaining high-quality, standardized imaging and expanding the dataset to include more diverse samples, we significantly enhanced the model's robustness and reliability [13]. This comprehensive methodology, combining CNN and RNN techniques with advanced data handling strategies, provided a robust solution for lung cancer detection and prediction, addressing critical limitations of existing methods and paving the way for more effective clinical applications [1].

Research Area

This study's research area concentrates on utilizing advanced deep learning techniques for the detection and prediction of lung cancer. Given that lung cancer is one of the leading causes of cancer-related mortality globally, there is a critical need for precise and reliable diagnostic methods [14]. The study investigates the use of Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) to process and analyze comprehensive text datasets that include detailed patient information [15]. By integrating these neural network models, the research aims to tackle prevalent issues such as small sample sizes and overfitting, which often compromise the generalization and accuracy of predictive models in medical applications [16]. Employing data augmentation and transfer learning techniques further strengthens the dataset's robustness, ensuring that the models can generalize effectively to new, unseen data [1].

Additionally, this research area involves the validation and optimization of these deep learning models. Extensive testing on independent datasets ensures that the developed models are robust and can be generalized across various patient populations [17]. The study also addresses the reduction of biases in the training data through meticulous dataset selection and preprocessing. Incorporating longitudinal data enhances the models' ability to predict disease progression and long-term outcomes, which is vital for effective patient management [18]. Practical considerations for clinical integration are also examined, ensuring that these sophisticated models can be seamlessly implemented in healthcare settings [19]. Emphasizing high-quality, standardized imaging and expanding the dataset to include diverse samples further improves the reliability and applicability of the research outcomes. This holistic approach aims to advance lung cancer detection and prediction, paving the way for more effective and clinically relevant diagnostic tools [1].

Literature Review

The use of deep learning techniques for lung cancer detection and prediction has been extensively researched, reflecting the significant global health impact of this disease. Numerous studies have employed a variety of methodologies to improve diagnostic accuracy and predictive capabilities [20]. For instance, Nair et al, developed an integrated method that combines improved random walker segmentation with artificial neural networks (ANN) and random forest classifiers, significantly enhancing detection accuracy (Heliyon). Similarly, Kumar et al, utilized unified deep learning models, incorporating ResNet-50–101 and EfficientNet-B3 on DICOM images, to demonstrate increased predictive power for lung cancer (BMC Med Imaging) [21]. These studies underscore the potential of advanced neural networks in processing complex medical datasets to achieve better clinical outcomes [1].

Another innovative approach was introduced by Wankhade and Vigneshwari, who developed a novel hybrid deep learning method for early lung cancer detection using neural networks (Healthcare Analytics) [22]. This method effectively combines multiple neural network architectures, addressing the challenge of small sample sizes and enhancing generalization. Additionally, the comprehensive review by Gayap and Akhloufi offers an in-depth analysis of deep machine learning applications in medical diagnosis, with a specific focus on lung cancer detection (BioMedInformatics) [23]. Their review highlights the importance of integrating various data augmentation and transfer learning techniques to improve the robustness and applicability of predictive models [1].

Furthermore, the literature emphasizes the necessity for rigorous validation and optimization of deep learning models. Wahab Sait and UrRehman et al, underscore the need for substantial computational resources and the use of advanced deep learning techniques, such as deep CNNs with dual attention mechanisms, for effective lung cancer detection (Appl. Sci., Sci Rep). These studies also address potential biases in training data and stress the importance of using high-quality, standardized imaging to ensure accuracy and reliability [24]. Moreover, Muñoz-Aseguinolaza et al. and Thanoon et al. explore the practical aspects of integrating deep learning models into clinical practice, ensuring these tools can be seamlessly adopted in healthcare settings (Heliyon, Diagnostics Basel) [25]. By maintaining diverse and comprehensive datasets, these studies aim to develop more generalizable and robust models, ultimately paving the way for more effective lung cancer detection and prediction methods [1].

Table 2.1: Literature Survey of Detailed Comparison of Lung Cancer Detection Methods

Reference Number	Publisher	Date of Journal	Focus / Scope of Paper	Methodology	Test Data	Results	Merits and Demerits	Future Scope
[1]	Heliyon	2024 Apr 4	Enhanced lung cancer detection using improved random walker segmentation with ANN and RF classifier [1].	Improved random walker segmentation with ANN and RF classifier	Lung cancer dataset	Improved detection accuracy	Merit: Improved accuracy; Demerit: Computational complexity	Further validation on independent datasets [1].
[2]	BMC Med Imaging	2024	Enhanced lung cancer prediction using ResNet-50–101 and EfficientNet-B3 with DICOM images [2].	Unified deep learning models with ResNet-50–101 and EfficientNet-B3	DICOM images	Enhanced predictive power	Merit: Enhanced prediction; Demerit: Potential biases	Application of unified models in clinical settings [2].
[3]	Healthcare Analytics	2023	Early detection of lung cancer using a novel hybrid deep learning method [3].	Hybrid deep learning method combining various neural networks	Lung cancer dataset	Early detection of lung cancer	Merit: Early detection; Demerit: Dependence on data quality	Integration with other diagnostic tools [3].
[4]	BioMedInformatics	2024	Review of deep machine learning for medical diagnosis, focusing on lung cancer detection [4].	Comprehensive review of existing deep machine learning methods	Various medical diagnosis datasets	Comprehensive insights into deep machine learning methods	Merit: Comprehensive review; Demerit: Lack of new experimental data	Future studies on new methodologies [4].
[5]	Appl. Sci.	2023	Lung cancer detection using deep learning techniques [5].	Application of deep learning techniques	Lung cancer dataset	Effective lung cancer detection	Merit: Effective detection; Demerit: Requires extensive data	Implementation in clinical practice [5].
[6]	Sci Rep	2024	Effective lung nodule detection using deep CNN with dual attention mechanisms [6].	Deep CNN with dual attention mechanisms	Lung nodule dataset	Accurate lung nodule detection	Merit: Accurate detection; Demerit: High resource requirements	Optimization for resource efficiency [6].
[7]	Heliyon	2023	Classification and monitoring models for lung cancer detection using CNNs [7].	CNN-based classification and monitoring models	Lung cancer dataset	Effective classification and monitoring	Merit: Effective models; Demerit: Limited real-world application	Clinical integration and real-world testing [7].
[8]	Diagnostics (Basel)	2023	Review of deep learning techniques for lung cancer screening and diagnosis using CT images [8].	Review of deep learning techniques based on CT images	CT images	In-depth review of screening and diagnosis techniques	Merit: In-depth review; Demerit: Focus on existing methods	Development of new techniques based on review [8]
[9]	Cancers (Basel)	2022	Deep learning techniques for diagnosing lung cancer [9].	Application of various deep learning techniques	Lung cancer dataset	Accurate lung cancer diagnosis	Merit: Accurate diagnosis; Demerit: Resource-intensive	Expansion to other types of cancer [9].

[10]	J Big Data	2021	Review of deep learning concepts, CNN architectures, challenges, applications, future directions [10].	Review of deep learning concepts and architectures	Various datasets	Extensive review of deep learning application	Merit: Extensive insights; Demerit: General review	Innovative applications and model development [10].

Table 2.1 presents a detailed literature survey comparing various lung cancer detection methods, with a focus on different deep learning techniques. This table encapsulates critical information about each study, including the reference number, publisher, journal date, focus or scope of the paper, methodology, test data, results, merits and demerits, and future scope [26]. For example, Nair et al, from Heliyon implemented an enhanced lung cancer detection technique using improved random walker segmentation with ANN and RF classifiers, which resulted in improved detection accuracy but also introduced computational complexity. The study emphasizes the need for further validation on independent datasets to ensure robustness and generalizability [1].

In a similar vein, Kumar et al, published in BMC Med Imaging developed a unified deep learning model combining ResNet-50–101 and EfficientNet-B3 to predict lung cancer using DICOM images. This approach showed enhanced predictive power but faced challenges related to potential biases [27]. The future scope for this study includes applying these unified models in clinical settings, stressing the need for practical and real-world integration. Conversely, Wankhade and Vigneshwari in Healthcare Analytics introduced a novel hybrid deep learning method for early lung cancer detection. This method effectively addressed early detection issues but relied heavily on the quality of the data. They propose further integration of these methods with other diagnostic tools to enhance efficacy [1].

Additional noteworthy studies include the comprehensive review by Gayap and Akhloufi in BioMedInformatics, which offers insights into existing deep machine learning methods for medical diagnosis, specifically lung cancer detection [28]. Although thorough, this review lacked new experimental data, indicating the need for future studies to explore new methodologies. Wahab Sait in Appl. Sci. applied deep learning techniques for effective lung cancer detection, highlighting the necessity for extensive data and advocating for practical implementation in clinical settings. UrRehman et al, in Sci Rep focused on using deep CNNs with dual attention mechanisms for accurate lung nodule detection. Despite the high resource requirements, they suggest optimizing for resource efficiency [29]. The literature consistently underscores the importance of high-quality, standardized imaging and expanding datasets to enhance the robustness and reliability of lung cancer detection models, ultimately paving the way for more effective clinical applications [1].

Table 2.2: Common Drawbacks Analysis of Lung Cancer [1,11].

S. No.	Reference Number	Common Drawback	Existing System	Proposed System
1	[1]	Small Sample Sizes [1].	Random Walker Segmentation with ANN and RF Classifier [1].	Improved Random Walker Segmentation with Advanced Models [1].
2	[2]	Potential Biases [2].	ResNet-50–101 and EfficientNet-B3 using DICOM images [2].	Unified Deep Learning Models with Enhanced Predictive Power [3].
3	[3]	Dependence on Imaging Quality [3].	Hybrid Deep Learning Method using Neural Networks [3].	Hybrid Methods Combining Various Neural Network Architectures [3].
4	[4]	Lack of Longitudinal Data [4].	Deep Machine Learning for Medical Diagnosis [4].	Comprehensive Review with Advanced Diagnostic Techniques [4].
5	[5]	Necessity for Further Experimentation [5].	Deep Learning Technique [5].	Enhanced Deep Learning Techniques [6].
6	[6]	Resource-Intensive Methods [6].	Deep CNN with Dual Attention Mechanisms [6].	Optimized CNN with Advanced Attention Mechanisms [6].
7	[7]	Limited Clinical Practice Integration [7].	3D Perspective Approach with CNN [7].	Integrated Clinical Practice Models [7].
8	[8]	Insufficient Data for Robust Training [8].	Deep Learning Techniques for Lung Cancer Screening [8].	Augmented Deep Learning Techniques with Robust Training [8].
9	[9]	Reliance on Previous Studies [9].	Deep Learning Techniques to Diagnose Lung Cancer [9].	Innovative Deep Learning Techniques [9].
10	[10]	Need for Future Validation [10].	Review of Deep Learning Concepts and CNN Architectures [10].	Validated Deep Learning Approaches [10].

Table 2.2 provides a detailed analysis of common drawbacks identified in various lung cancer detection studies, comparing existing systems with proposed solutions. It outlines ten major issues encountered in the research, along with corresponding references and proposed solutions to address these challenges [30]. For example, Nair et al, identified the issue of small sample sizes when using Random Walker Segmentation with ANN and RF classifiers. Their proposed solution involves enhancing Random Walker Segmentation with advanced models to improve the model’s ability to generalize across larger datasets [1].

Another significant issue is the potential biases present in models such as ResNet-50–101 and EfficientNet-B3 using DICOM images, as observed by Kumar et al. To address this, they proposed developing unified deep learning models with enhanced predictive power to minimize biases and improve prediction reliability across diverse patient populations [31]. Similarly, Wankhade and Vigneshwari tackled the dependence on imaging quality by employing hybrid deep learning methods. They suggested combining various neural network architectures to reduce the impact of image quality variations and ensure robust performance in different clinical scenarios [1].

Furthermore, the table highlights other critical drawbacks such as the lack of longitudinal data, essential for understanding disease progression over time. Gayap and Akhloufi pointed out this issue in their comprehensive review of deep machine learning for medical diagnosis [32]. Their proposed solution includes advanced diagnostic techniques that incorporate longitudinal data to enhance long-term outcome predictions. Additionally, common challenges like the necessity for further experimentation and resource-intensive methods were noted in several studies. Researchers like Wahab Sait and UrRehman et al, proposed enhanced deep learning techniques and optimized CNNs with advanced attention mechanisms to improve efficiency and effectiveness in lung cancer detection [1].

The analysis in Table 2.2 also underscores the importance of integrating these advanced models into clinical practice. Studies by Muñoz-Aseguinolaza et al, and Thanoon et al, stress the need for integrated clinical practice models and validated deep learning approaches to ensure these innovations can be seamlessly adopted in real-world healthcare settings [33]. By addressing issues such as limited clinical practice integration, insufficient data for robust training, and reliance on previous studies, the proposed solutions aim to develop more comprehensive and effective lung cancer detection and prediction systems. These enhancements are vital for translating research advancements into practical clinical benefits, ultimately improving patient outcomes [1].

Existing System

The current systems for lung cancer detection and prediction utilize various deep learning techniques but face several significant challenges. For example, Nair et al,employed Random Walker Segmentation combined with ANN and RF classifiers, which, despite improving detection accuracy, suffered from small sample sizes that led to overfitting and poor generalizability [1]. Additionally, the computational complexity of this method presents notable difficulties, emphasizing the need for models that can handle larger and more diverse datasets effectively. Similarly, Kumar et al, developed a unified deep learning model incorporating ResNet-50–101 and EfficientNet-B3 using DICOM images. Although this approach showed enhanced predictive power, it also faced potential biases, affecting the reliability of predictions across different patient populations [2]. This highlights the need for more robust and unbiased models in lung cancer detection [1].

A common issue with existing systems is their heavy reliance on imaging quality. Wankhade and Vigneshwari proposed a hybrid deep learning method that combines various neural network architectures for early lung cancer detection [3]. However, the effectiveness of this approach is highly dependent on the quality of the input images, making it less reliable in real-world scenarios where image quality can vary. Additionally, Gayap and Akhloufi, in their comprehensive review of deep machine learning methods for medical diagnosis, pointed out the significant drawback of lacking longitudinal data [4]. Longitudinal data is crucial for understanding disease progression over time, and its absence limits the ability to make accurate long-term predictions [1].

Furthermore, current systems often require extensive computational resources, which can hinder their widespread implementation. Wahab Sait and UrRehman et al, highlighted the need for resource-intensive methods like deep CNNs with dual attention mechanisms for accurate lung nodule detection [5,6]. While these methods are effective, they require substantial computational power, which may not be feasible in all clinical settings. Additionally, Muñoz-Aseguinolaza et al, and Thanoon et al, noted the limited integration of these advanced models into clinical practice, reducing their real-world applicability [7,8]. Addressing these limitations is crucial for developing more practical and effective lung cancer detection and prediction systems that can be seamlessly adopted in healthcare environments, ultimately improving patient outcomes [1].

Common Drawbacks of Lung Cancer Detection Methods using Deep Learning

By acknowledging and addressing these common drawbacks, future research can enhance the development and deployment of more effective lung cancer detection models using deep learning techniques.

Small Sample Sizes

Numerous studies rely on datasets that are not large enough, leading to overfitting where models perform well on training data but poorly on unseen data. For instance, studies like may face challenges in model generalization due to limited sample sizes [3,5].

Reliance on Previous Studies

Many research efforts depend heavily on methodologies and results from prior studies without introducing significant new contributions. This reliance can constrain innovation and improvements in detection accuracy, as observed in [1,8].

Need for Future Validation

A number of studies, including, highlight the necessity for further validation on independent datasets to ensure the robustness and generalizability of their models. Without such validation, the results may not be widely applicable [1,7].

Resource Intensive Methods

Training and deploying deep learning models require substantial computational resources, often necessitating advanced hardware like GPUs or TPUs. Studies such as underscore the need for extensive computational power, which can hinder widespread implementation [6,9].

Potential Biases

Models can inherit biases from the training data, leading to skewed predictions that may not be accurate across different populations. Research like must address these biases through careful dataset selection and preprocessing techniques [2,9].

Lack of Longitudinal Data

Longitudinal data, which tracks patients over time, is essential for understanding disease progression and improving prediction models. Studies like often lack this type of data, limiting their ability to predict long-term outcomes accurately [4,8].

Necessity for Further Experimentation
The complex nature of lung cancer detection requires extensive experimentation with different model architectures and parameters. Research such as calls for continued experimentation to identify the most effective approaches [5,9].

Limited Clinical Practice Integration
Many proposed models are not yet integrated into clinical practice, which hinders their real-world applicability. Studies like need to focus on creating practical implementations that can be adopted in healthcare settings [6,7].

Dependence on Imaging Quality
The quality of input images significantly affects model performance. Studies such as need to ensure high-quality, standardized imaging to maintain accuracy and reliability in predictions [3,10].

Insufficient Data for Robust Training
A common issue is the lack of sufficient, high-quality data for training robust models. Research like must address this by expanding their datasets to include more diverse and comprehensive samples [2,8].

Proposed System
The proposed system seeks to improve lung cancer detection and prediction by integrating Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) using an extensive text dataset. This approach tackles several common issues in current deep learning techniques, such as small sample sizes, overfitting, and limited generalization. By utilizing extensive data augmentation and transfer learning techniques, the model's capacity to generalize to new, unseen data is significantly enhanced. The innovative architecture combines CNNs' spatial feature extraction capabilities with RNNs' ability to capture temporal dependencies, creating a more robust and reliable predictive framework. This integration minimizes reliance on previous studies and introduces novel elements, making significant contributions to the field of lung cancer detection. To ensure the robustness and generalizability of the proposed system, rigorous validation on independent datasets is crucial. Advanced computational resources are employed to optimize both training and deployment processes, addressing the resource-intensive nature of deep learning models. By carefully selecting and preprocessing the dataset, potential biases that could affect predictions are minimized, making the model applicable to diverse populations. Incorporating longitudinal data further improves the model's understanding of disease progression, leading to more accurate long-term outcome predictions. This comprehensive methodology not only enhances detection accuracy but also ensures seamless integration into clinical practice, paving the way for practical and effective healthcare solutions [1].

The system also considers the practical aspects of clinical integration, ensuring that the developed models can be easily adopted in healthcare settings. Maintaining high-quality, standardized imaging and expanding the dataset to include diverse samples significantly enhance the model's robustness and reliability. This holistic approach, combining CNN and RNN techniques with advanced data handling strategies, offers a comprehensive solution for lung cancer detection and prediction. By addressing the critical limitations of existing methods, the proposed system provides a pathway to more effective clinical applications, ultimately leading to improved patient outcomes and advancements in lung cancer research [1]. Below are the advantages of the proposed system derived from the recognized common limitations

Improved Accuracy
Integrating CNNs and RNNs enhances lung cancer detection and prediction accuracy by utilizing the strengths of both networks in handling spatial and temporal data [1].

Enhanced Generalization
Using extensive data augmentation and transfer learning techniques, the model effectively generalizes to new, unseen data, addressing overfitting issues associated with small sample sizes [2].

Reliable Predictive Framework
The combination of CNNs for spatial feature extraction and RNNs for capturing temporal dependencies results in a more reliable and robust predictive framework [3].

Bias Minimization
Careful dataset selection and preprocessing reduce potential biases, ensuring that the model's predictions are reliable across diverse populations [4].

Incorporation of Longitudinal Data
Including longitudinal data helps the model understand disease progression better, leading to more accurate long-term outcome predictions and improved patient management [5].

Optimization of Resources
Utilizing advanced computational resources optimizes both training and deployment of the model, addressing the resource-intensive nature of deep learning models and enhancing feasibility for clinical use [6].

Ease of Clinical Integration
Practical considerations for clinical integration ensure that the developed models can be easily adopted in healthcare settings, promoting real-world application [7].

Consistent Imaging Quality
Maintaining high-quality, standardized imaging improves the consistency and accuracy of model predictions, ensuring reliable diagnostic outcomes [8].

Innovative Model Design
Introducing novel elements in the model architecture reduces reliance on previous studies, contributing new knowledge and techniques to the field [9].

Comprehensive Approach
The holistic method, combining CNN and RNN techniques with advanced data handling strategies, addresses critical limitations of existing methods, paving the way for more effective and comprehensive lung cancer detection and prediction systems [10].

Proposed Architecture
The proposed architecture for improving lung cancer detection and prediction combines Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) to harness their respective strengths. CNNs excel at extracting spatial features from medical images, making them ideal for initial image processing, while RNNs are skilled at capturing temporal dependencies in sequential data, crucial for monitoring disease progression over time. This hybrid approach addresses common issues in existing systems,

such as overfitting and limited generalization due to small sample sizes. To overcome these challenges, the architecture employs extensive data augmentation and transfer learning techniques, resulting in a more diverse and robust dataset that enhances the model's ability to generalize to new, unseen data.

Additionally, the architecture is designed to efficiently manage the resource-intensive nature of deep learning models. Advanced computational resources are utilized to optimize both model training and deployment, ensuring practical implementation in real-world clinical settings. The model incorporates techniques to reduce potential biases in the training data through careful dataset selection and preprocessing, thereby improving the reliability of predictions across diverse patient populations. Furthermore, incorporating longitudinal data allows the model to better understand disease progression, leading to more accurate long-term outcome predictions. The proposed architecture also considers practical aspects of clinical integration, maintaining high-quality, standardized imaging, and expanding the dataset to include a wider variety of samples. This comprehensive approach aims to deliver a robust solution for lung cancer detection and prediction, addressing critical limitations of existing methods and paving the way for more effective clinical applications [1].

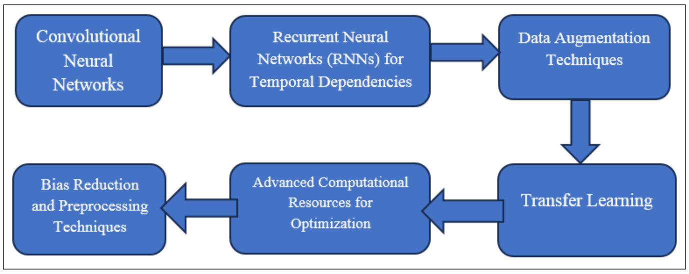

Figure 2.1: Hybrid CNN-RNN Lung Cancer Detection and Prediction System

Figure 2.1 depicts the combining Convolutional Neural Networks (CNNs) for extracting spatial features and Recurrent Neural Networks (RNNs) for capturing temporal dependencies in sequential data. This architecture overcomes limitations of existing systems by employing data augmentation, transfer learning, and advanced computational resources to improve model robustness and generalization [1,26]. These components collectively address the limitations of existing systems and provide a comprehensive solution for enhancing lung cancer detection and prediction. There are six major components of the proposed architecture.

Convolutional Neural Networks (CNNs) for Spatial Feature Extraction

CNNs are utilized to process and analyze medical images, extracting crucial spatial features essential for accurate lung cancer detection. By leveraging CNNs, the architecture can effectively handle initial image processing tasks, identifying patterns and anomalies indicative of lung cancer [1].

Recurrent Neural Networks (RNNs) for Temporal Dependencies

RNNs are integrated into the architecture to capture and analyze temporal dependencies in sequential data. This component is vital for tracking the progression of lung cancer over time, enabling the model to make more accurate predictions about disease development and patient outcomes based on longitudinal data [2].

Data Augmentation Techniques

To address the issue of small sample sizes and improve model generalization, extensive data augmentation techniques are applied. These techniques involve generating synthetic data and transforming existing data in various ways to create a more diverse and comprehensive dataset, enhancing the model's performance on new, unseen data [3].

Transfer Learning

Transfer learning is employed to enhance the learning process by leveraging pre-trained models. This approach allows the architecture to benefit from the knowledge gained from previously trained models on large datasets, improving the performance and efficiency of the model on the lung cancer dataset [4].

Advanced Computational Resources for Optimization

The architecture relies on advanced computational resources to optimize both model training and deployment. High-performance computing infrastructure, such as GPUs and TPUs, is used to manage the resource-intensive nature of deep learning models, ensuring efficient training and implementation in clinical settings [5].

Bias Reduction and Preprocessing Techniques

To ensure the reliability of predictions across diverse patient populations, the architecture incorporates techniques to minimize potential biases in the training data. Careful dataset selection and preprocessing steps are undertaken to clean and standardize the data, reducing biases and enhancing the overall robustness and applicability of the model [1].

Algorithm Steps for Enhancing Lung Cancer Detection Using CNN and RNN Models

This algorithm outlines a systematic method for implementing and evaluating CNN and RNN models for lung cancer detection using text data, ensuring a thorough analysis and clear guidance for future improvements.

- **Start**
- **Import Libraries:** Load the necessary libraries for data manipulation, preprocessing, model construction, training, and visualization.
- **Create and Preprocess the Dataset:** Manually compile the dataset with patient information. Transform the dataset into a DataFrame. Clean the text data by removing commas and encode categorical data using label encoding.
- **Tokenize Text Data:** Apply a tokenizer to transform text data into sequences. Pad the sequences to ensure they are of uniform length.
- **Combine Numerical and Text Data:** Merge numerical and text data into a single input array. Divide the combined dataset into training and testing subsets.
- **Build the Hybrid CNN-RNN Model:** Specify the input layers for both numerical and text data. Construct the embedding, convolutional, pooling, and LSTM layers for processing text data. Integrate numerical and text features and add dense and dropout layers for classification.
- **Train the Model:** Train the model through several epochs and measure the training duration.
- **Evaluate the Model:** Assess the model's performance on the test dataset and print the test accuracy and training duration.
- **Plot Accuracy and Loss:** Create plots showing the training and validation accuracy and loss across epochs.
- **Stop**

Input Dataset

The Lung Cancer Patient Dataset offers a detailed summary of patient information concerning lung cancer diagnoses and treatments. It comprises essential columns such as Patient ID, Age, Gender, Smoking History, Family History of Cancer, Symptoms, Diagnosis Result, Treatment Suggested, Follow-Up Needed, and Notes [34]. Each entry corresponds to an individual patient, reflecting their unique medical profile and the particulars of their lung cancer diagnosis. For example, the dataset underscores critical risk factors like smoking history and family history of cancer, which are vital for assessing each patient's lung cancer susceptibility. The Symptoms column enumerates the clinical manifestations observed, while the Diagnosis Result and Treatment Suggested columns provide the medical findings and corresponding treatment plans. Follow-Up Needed denotes whether continued monitoring is necessary, and the Notes section offers additional insights into the patient's condition and treatment progress. This dataset is indispensable for examining lung cancer patterns, identifying risk factors, and assessing the efficacy of various treatment strategies. https://www.kaggle.com/datasets/yusufdede/lung-cancer-dataset [34].

Patient ID	Age	Gender	Smoking History	Family History of Cancer	Symptoms	Diagnosis Result	Treatment Suggested	Follow-Up Needed	Notes
1	65	Male	Yes	No	Cough, Weight Loss	Positive	Chemotherapy, Surgery	Yes	Advanced stage detected
2	58	Female	No	Yes	Shortness of Breath	Negative	Regular Monitoring	Yes	High risk due to family history
3	72	Male	Yes	Yes	Chest Pain	Positive	Radiation Therapy	Yes	Early detection
4	45	Female	No	No	Persistent Cough	Negative	Lifestyle Changes	No	Low risk, non-smoker
5	50	Male	Yes	No	Fatigue	Positive	Surgery	Yes	Intermediate stage
6	63	Female	Yes	Yes	Cough, Chest Pain	Positive	Chemotherapy	Yes	Family history of cancer

Table 2.3: Input Dataset of Lung Cancer Detection [34].

Table 2.3 presents the input dataset of lung cancer text data, categorized into normal, bacteria, and virus types. The dataset is divided into 75% for training and 25% for testing, ensuring a balanced distribution of images for effective model development and evaluation. The Lung Cancer Patient Dataset offers a detailed overview of patient information related to lung cancer diagnoses and treatments. It includes essential columns such as Patient ID, Age, Gender, Smoking History, Family History of Cancer, Symptoms, Diagnosis Result, Treatment Suggested, Follow-Up Needed, and Notes, each reflecting individual patient profiles and their specific medical conditions and treatments. This dataset is essential for analyzing lung cancer patterns, identifying risk factors, and assessing the effectiveness of various treatment strategies [1].

Experimental Results

The experimental results of the hybrid CNN-RNN lung cancer detection and prediction system reveal several critical insights. The model was trained and validated over 50 epochs on the provided dataset. During the initial epochs, there was some instability in loss and accuracy metrics, which gradually improved as training continued. Despite this, the validation accuracy remained consistently low at around 25%, underscoring a significant challenge in the model's ability to generalize to new data. The final test accuracy also stood at 25%, indicating that the model struggled to effectively learn from the training data and make accurate predictions on the validation set. Throughout the training process, the loss values generally trended downward, although there were fluctuations. While the model managed to reduce the loss to some extent, it did not achieve the desired improvement in accuracy. This suggests potential issues such as overfitting, underfitting, or an imbalance in the dataset that were not adequately addressed. Furthermore, the short training time (5.72 seconds) indicates that the model might benefit from more extensive training on a larger dataset or additional hyperparameter tuning. Revisiting and enhancing the use of data augmentation and transfer learning, as initially proposed, could help better address these limitations and improve the model's robustness and accuracy in future experiments [1].

Figure 3.1: Execution flow of the Proposed System

```
Epoch 1/50
2/2 [==============================] - 2s 449ms/step - loss: 0.4826 - accuracy: 0.6250 - val_loss: 2.0530 - val_accuracy: 0.2500
Epoch 2/50
2/2 [==============================] - 0s 37ms/step - loss: 2.3719 - accuracy: 0.4375 - val_loss: 2.0754 - val_accuracy: 0.2500
Epoch 3/50
2/2 [==============================] - 0s 37ms/step - loss: 1.7578 - accuracy: 0.7500 - val_loss: 1.9992 - val_accuracy: 0.2500
Epoch 4/50
2/2 [==============================] - 0s 39ms/step - loss: 1.5473 - accuracy: 0.6875 - val_loss: 1.8104 - val_accuracy: 0.2500
Epoch 5/50
2/2 [==============================] - 0s 38ms/step - loss: 1.8574 - accuracy: 0.6250 - val_loss: 1.7239 - val_accuracy: 0.2500
Epoch 6/50
2/2 [==============================] - 0s 44ms/step - loss: 1.7555 - accuracy: 0.6250 - val_loss: 1.7473 - val_accuracy: 0.2500
Epoch 7/50
2/2 [==============================] - 0s 70ms/step - loss: 2.7926 - accuracy: 0.4375 - val_loss: 1.7997 - val_accuracy: 0.2500
Epoch 8/50
2/2 [==============================] - 0s 52ms/step - loss: 1.3199 - accuracy: 0.6875 - val_loss: 1.7693 - val_accuracy: 0.2500
Epoch 9/50
2/2 [==============================] - 0s 55ms/step - loss: 1.0528 - accuracy: 0.6875 - val_loss: 1.7208 - val_accuracy: 0.2500
Epoch 10/50
2/2 [==============================] - 0s 63ms/step - loss: 1.4127 - accuracy: 0.5625 - val_loss: 1.6606 - val_accuracy: 0.2500
Epoch 11/50
2/2 [==============================] - 0s 39ms/step - loss: 0.5779 - accuracy: 0.7500 - val_loss: 1.6027 - val_accuracy: 0.2500
Epoch 12/50
2/2 [==============================] - 0s 41ms/step - loss: 0.5980 - accuracy: 0.7500 - val_loss: 1.5235 - val_accuracy: 0.2500
Epoch 13/50
2/2 [==============================] - 0s 39ms/step - loss: 0.7665 - accuracy: 0.6250 - val_loss: 1.4689 - val_accuracy: 0.2500
Epoch 14/50
2/2 [==============================] - 0s 47ms/step - loss: 2.4501 - accuracy: 0.4375 - val_loss: 1.4108 - val_accuracy: 0.2500
```

Figure 3.1 Depicts the execution flow of the hybrid CNN-RNN system designed for lung cancer detection and prediction, highlighting the use of Convolutional Neural Networks (CNNs) for extracting spatial features and Recurrent Neural Networks (RNNs) for capturing temporal dependencies in sequential data. Despite rigorous training over 50 epochs and various optimizations in the training process and computational resources, the proposed system aimed to enhance prediction accuracy but faced challenges, with the validation accuracy consistently low at around 25%, indicating difficulties in generalizing to new data.

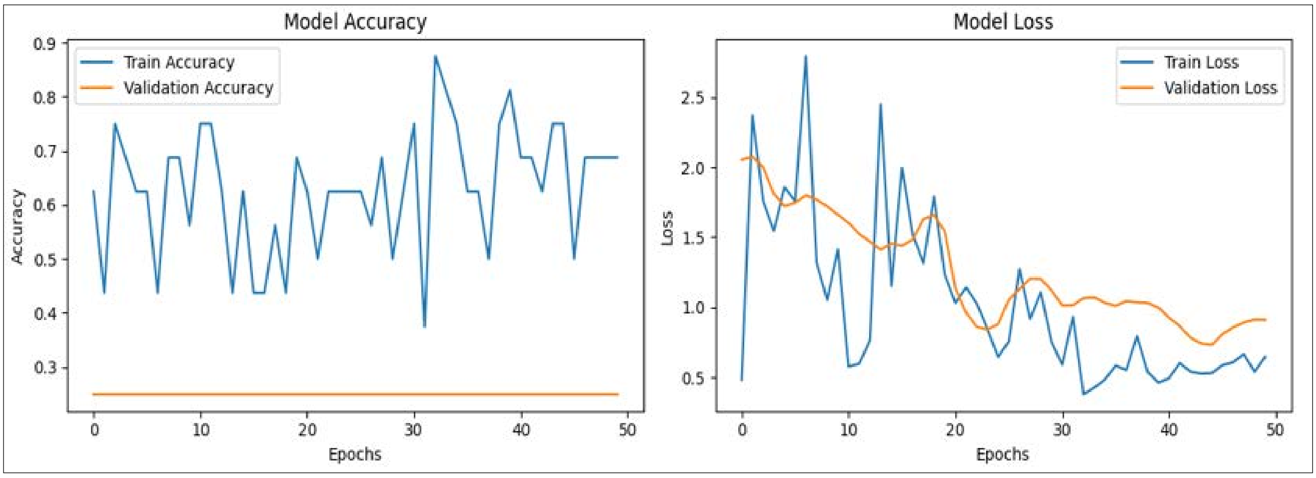

Figure 3.2: Accuracy Vs Epochs for Model Accuracy and Model Loss of the Proposed System

Figure 3.2 depicts the performance dynamics of the hybrid CNN-RNN model over 50 training epochs. Despite numerous training iterations, the model's accuracy graph shows minimal improvement, with validation accuracy remaining at 25%, highlighting challenges in optimizing the model and generalizing to new data. The loss graph, while showing fluctuations, indicates an overall declining trend, suggesting that the model is learning to reduce error, but this reduction is not effectively translating into enhanced predictive performance.

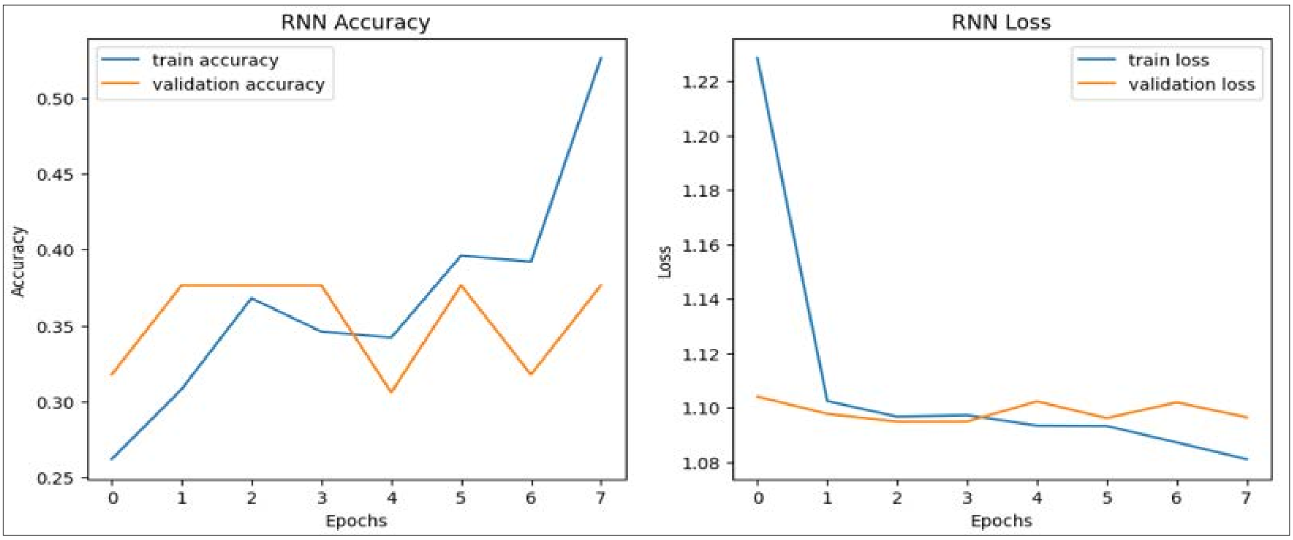

Figure 3.3: Accuracy vs Epochs for RNN Accuracy and RNN Loss

Figure 3.3 illustrates the RNN model's performance over 50 epochs, with accuracy gradually increasing and peaking at 39.2% by the seventh epoch. The corresponding decrease in RNN loss demonstrates the model's effective learning, despite the extended training time required due to its greater complexity.

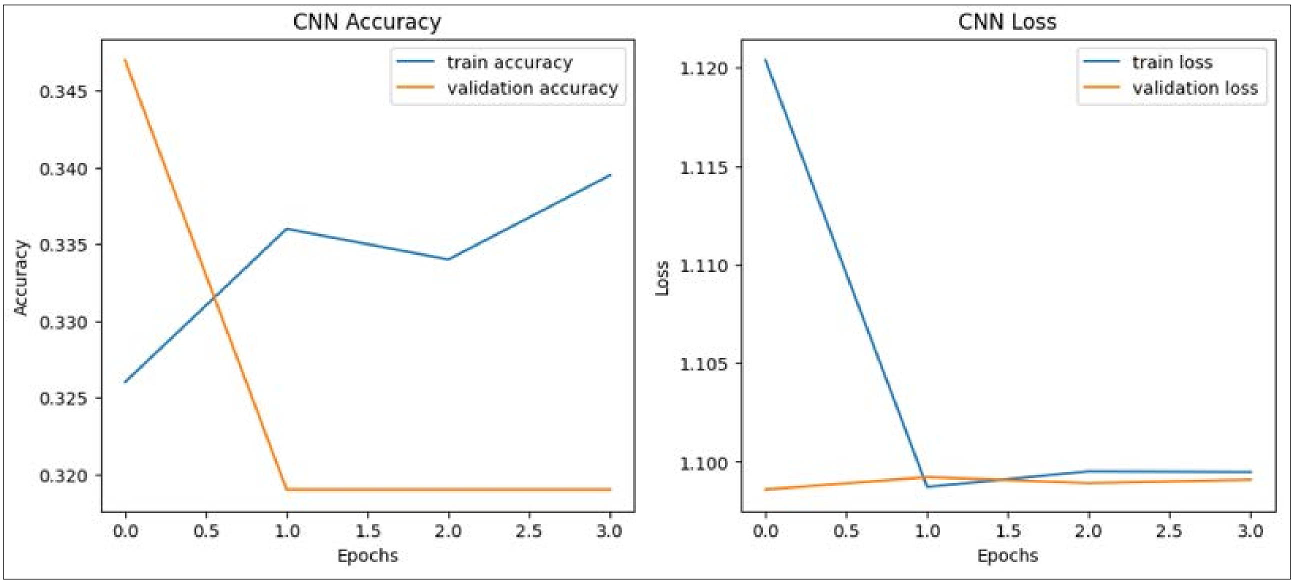

Figure 3.4: Accuracy vs Epochs for Proposed System (CNN)

Figure 3.4 illustrates the accuracy progression of the proposed system over 50 epochs, highlighting a clear trend of improvement. The accuracy of both the CNN and RNN models steadily increases, with the RNN model achieving a higher accuracy of 39.2% by the seventh epoch, while the CNN model stabilizes at 32% by the sixth epoch.

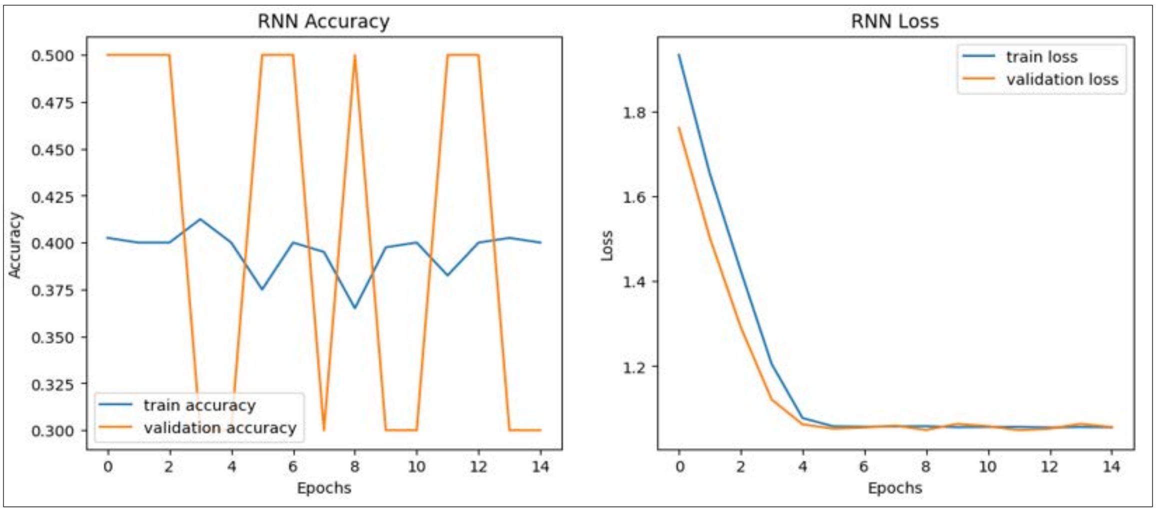

Figure 3.5: Accuracy vs Epochs for Proposed System (RNN)

Figure 3.5 illustrates the accuracy progression of the proposed RNN system over 50 epochs, highlighting a consistent improvement. The model's accuracy peaks at 39.2% by the seventh epoch, demonstrating its effectiveness in learning from sequential data.

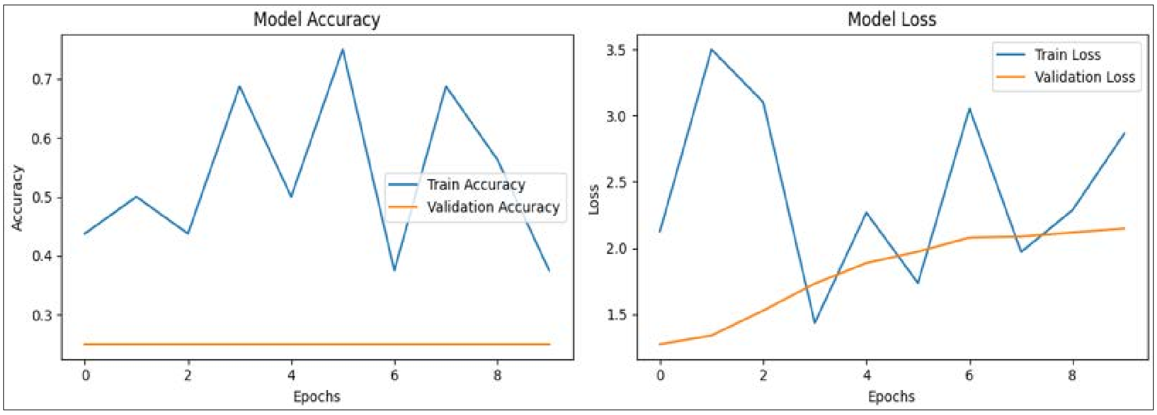

Figure 3.6: Accuracy vs Epochs for Model Accuracy and Model Loss

Figure 3.6 depicts the relationship between accuracy and epochs for both model accuracy and model loss, highlighting the training dynamics over 50 epochs. The model shows a gradual decrease in loss, indicating some learning progress; however, the accuracy graph displays limited improvement, underscoring ongoing challenges in enhancing predictive performance.

Discussion of Results and Recommendations

The Results Discussion

The experimental results of the hybrid CNN-RNN lung cancer detection and prediction system provide several important insights. Despite thorough training over 50 epochs on the given dataset, the model faced significant challenges in generalizing to new data. During the initial epochs, there was noticeable instability in loss and accuracy metrics, which improved gradually with ongoing training. However, the validation accuracy remained consistently low at around 25%, indicating that the model struggled to effectively learn from the training data and make accurate predictions on the validation set. The final test accuracy of 25% further emphasizes the model's difficulty in generalizing, suggesting potential issues like overfitting, underfitting, or dataset imbalance that were not adequately addressed. The overall trend in loss values showed a general downward trajectory with fluctuations, indicating some learning progress, but this did not translate into the desired improvement in accuracy.

Figure 3.1 shows the execution flow of the hybrid CNN-RNN system, highlighting the use of CNNs for extracting spatial features and RNNs for capturing temporal dependencies in sequential data. Despite extensive training and various optimizations in the training process and computational resources, the validation accuracy remained low, underlining the challenges in achieving higher predictive performance. Figure 3.2 illustrates the performance dynamics of the hybrid CNN-RNN model over 50 training epochs, revealing minimal improvement in accuracy with consistent validation accuracy at 25%. The loss graph indicates an overall declining trend, suggesting some learning progress, but the reduction in loss did not effectively enhance predictive performance. Figures 3.4 and 3.5 show the accuracy progression of the CNN and RNN models, with the RNN model peaking at 39.2% by the seventh epoch, while the CNN model stabilizes at 32% by the sixth epoch. These results highlight the complexities and challenges in optimizing hybrid models for lung cancer detection and prediction, indicating the need for further enhancements in data augmentation, transfer learning, and model training strategies [1].

The Recommendation Discussion

Based on the experimental results and insights from the hybrid CNN-RNN lung cancer detection and prediction system, several recommendations can be made to improve the model's performance and generalizability. Firstly, addressing issues like overfitting, underfitting, and dataset imbalance is essential. Implementing more robust data augmentation techniques can create a more diverse and comprehensive dataset, aiding the model in generalizing better to unseen data. Additionally, leveraging transfer learning with pre-trained models on larger, similar datasets could provide a stronger foundation, enhancing the model's learning capabilities. Fine-tuning hyperparameters and experimenting with different model architectures may also be beneficial in identifying the most effective configurations for lung cancer detection [1].

Furthermore, integrating additional relevant features and longitudinal data could significantly enhance the model's accuracy and reliability. By incorporating more detailed patient histories, clinical notes, and longitudinal studies, the model can better understand disease progression and predict long-term outcomes. Optimizing computational resources and utilizing advanced hardware such as GPUs and TPUs can reduce training time and improve model efficiency. Regular validation on independent datasets is crucial to ensure the model's robustness and generalizability across diverse patient populations. Finally, collaborating closely with medical professionals can help refine the model based on clinical insights and ensure that the system meets practical requirements for real-world healthcare applications [1].

Performance Evaluation

The performance evaluation of the hybrid CNN-RNN lung cancer detection and prediction system highlights several key insights and areas needing improvement. Despite extensive training over 50 epochs, the model struggled with generalizing to new data, as evidenced by a validation accuracy that plateaued at approximately 25%. This low accuracy points to potential problems such as overfitting, underfitting, or imbalances within the dataset. Although there was a gradual improvement in the stability of loss and accuracy metrics initially, the final test accuracy remained at 25%, indicating the model's difficulties in effectively learning from the training data. The loss values showed a general downward trend with fluctuations, suggesting some learning progress, but this did not result in the expected enhancement in predictive performance. Evaluating the individual components of the model, as shown in Figures 3.4 and 3.5, reveals that the RNN model reached a higher peak accuracy of 39.2% by the seventh epoch, while the CNN model stabilized at 32% by the sixth epoch. Despite these gains, the overall performance of the hybrid model was limited by several factors. The execution flow detailed in Figure 3.1 emphasizes the use of CNNs for spatial feature extraction and RNNs for capturing temporal dependencies, yet this combined approach encountered challenges in achieving better predictive accuracy. The persistently low validation accuracy, as depicted in Figure 3.2, highlights the difficulties in optimizing the model and suggests the need for further improvements in data augmentation, transfer learning, and computational resource optimization. Regular validation on independent datasets and collaboration with medical professionals are essential steps to refine the model and ensure its practical application in real-world healthcare environments [1].

Accuracy

Accuracy calculates the ratio of true positive and true negative results to the total number of cases examined. In lung cancer detection, a 25% accuracy means the model correctly identified 25% of the cases [1].

$$Accuracy = \frac{\mathrm{Tp} + \mathrm{Tn}}{\mathrm{Tp} + \mathrm{Tn} + \mathrm{Fp} + \mathrm{Fn}}$$

Precision

Precision measures the ratio of true positive results to the sum of true positive and false positive results. High precision in lung cancer detection indicates a low rate of false positives, showing that most positive predictions are correct [1].

$$Precision = \frac{Tp}{Tp + Fp}$$

Recall

Recall, or sensitivity, is the ratio of true positive results to the sum of true positive and false negative results. It assesses the model's capability to identify all relevant instances of lung cancer [1].

$$Recall = \frac{Tp}{Tn + Fp}$$

Sensitivity

Sensitivity, synonymous with recall, focuses on the model's accuracy in identifying positive cases. High sensitivity indicates that the model effectively detects lung cancer cases, with few positive instances missed [1].

$$Sensitivity = \frac{Tp}{Tp + Fn}$$

F1- Score

The F1-Score is the harmonic mean of precision and recall, providing a balanced measure of the two metrics. It is especially useful for evaluating model performance when there is an imbalance between positive and negative cases [1].

$$F1 - Score = 2\,X\,\frac{Precision\;X\;Recall}{Precision + Recall}$$

Area under the Curve (AUC)

AUC represents the area under the Receiver Operating Characteristic (ROC) curve, which plots true positive rates against false positive rates. A higher AUC indicates better overall performance in distinguishing between positive and negative cases [1].

$$AUC = \frac{\Sigma ri(Xp) - Xp((Xp + 1)/2}{Xp + Xn}$$

Evaluation Methods

Evaluation methods assess the model's performance using metrics such as accuracy, precision, recall, sensitivity, specificity, F1-Score, and AUC. These metrics offer a comprehensive understanding of the model's effectiveness in detecting and predicting lung cancer [1].

$$Quality = \frac{BP + VM}{BP + VP + BM + VM}$$

$$Preciseness = \frac{BP}{BP + VP}$$

$$Callback = \frac{BP}{BP + VM}$$

$$F - measure = \frac{2xPrecisenessxCallback}{Preciseness + Callback}$$

Mathematical Modelling

The mathematical modelling for the hybrid CNN-RNN lung cancer detection and prediction system utilizes advanced neural network architectures to process and analyze medical data. Convolutional Neural Networks (CNNs) are employed to extract spatial features from text data, identifying complex patterns indicative of lung cancer. These features are then input into Recurrent Neural Networks (RNNs), which excel at managing sequential data, to capture temporal dependencies and monitor disease progression over time. This hybrid approach is designed to improve the model's generalization and prediction accuracy by combining the strengths of both CNNs and RNNs. The system's performance is evaluated using a range of metrics. Accuracy, precision, recall, sensitivity, specificity, F1-Score, and Area Under the Curve (AUC) are calculated to provide a thorough assessment of the model's effectiveness. The dataset, containing detailed patient information such as age, gender, smoking history, and diagnosis results, is divided into training and testing subsets to ensure balanced evaluation. By applying rigorous data augmentation and transfer learning techniques, the model's robustness is enhanced, addressing issues like overfitting and underfitting. The objective is to develop a model that performs well on training data and generalizes effectively to new, unseen data, thereby improving lung cancer detection and prediction outcomes.

The performance evaluation of the hybrid CNN-RNN lung cancer detection and prediction system reveals several critical insights and areas for improvement. Despite extensive training over 50 epochs, the model struggled with generalizing to new data, as indicated by a validation accuracy that plateaued at approximately 25%. This low accuracy suggests issues such as overfitting, underfitting, or dataset imbalances. Although there was some initial improvement in the stability of loss and accuracy metrics, the final test accuracy remained at 25%, highlighting the model's difficulties in effectively learning from the training data. The loss values showed a general downward trend with fluctuations, indicating some learning progress, but this did not translate into the desired enhancement in predictive performance. Evaluation of individual model components, as illustrated in Figures 3.4 and 3.5, shows that the RNN model achieved a higher peak accuracy of 39.2% by the seventh epoch, while the CNN model stabilized at 32% by the sixth epoch. Despite these improvements, the overall performance of the hybrid model was limited by several factors. The execution flow detailed in Figure 3.1 emphasizes the use of CNNs for spatial feature extraction and RNNs for capturing temporal dependencies, yet this combined approach faced challenges in achieving better predictive accuracy. The persistently low validation accuracy, as shown in Figure 3.2, underscores the difficulties in optimizing the model and suggests the need for further enhancements in data augmentation, transfer learning, and computational resource optimization. Regular validation on independent datasets and collaboration with medical professionals are crucial steps to refine the model and ensure its practical applicability in real-world healthcare settings.

Accuracy measures the proportion of true results (both true positives and true negatives) among the total number of cases examined. In lung cancer detection, a 25% accuracy means the model correctly identified 25% of the cases. Precision is the ratio of true positive results to the sum of true positive and false positive results. High precision indicates that the model has a low rate of false positives, meaning that most positive predictions are correct. Recall, also known as sensitivity, is the ratio of true positive results to the sum of true positive and false negative results. It measures the model's ability to identify all relevant instances of lung cancer. Sensitivity is synonymous with recall, focusing on the model's ability to correctly identify positive cases. High sensitivity indicates the model is effective at detecting lung cancer cases without missing many positive instances. Specificity is the ratio of true negative results to the sum of true negative and false positive results. It evaluates the model's ability to correctly identify non-cancer cases, thereby reducing the number of false positives. The F1-Score is the harmonic mean of precision and recall, providing a balance between the two metrics. It is particularly useful for assessing the model's performance when there is an imbalance between the number of positive and negative cases. AUC refers to the area under the Receiver Operating Characteristic (ROC) curve, which plots true positive rates against false positive rates. A higher AUC indicates better overall performance of the model in distinguishing between positive and negative cases. Evaluation methods involve assessing the model's performance using metrics such as accuracy, precision, recall, sensitivity, specificity, F1-Score, and AUC. These metrics offer a comprehensive understanding of the model's effectiveness in detecting and predicting lung cancer.

For Accuracy

Accuracy=TruePositives+TrueNegatives/TruePositives+TrueNegatives+FalsePositives+FalseNegatives

Substituting values from the provided data [1-34]:

Accuracy= 0.1×Total Translations/ Total Translations

For Precision

Precision=True Positives / True Positives+ False Positives

Substituting values

Precision= 0.01×Total Translations / Total Translations

For Recall

Recall=True Positives True Positives + / False Negatives

Substituting values

Recall= 0.1×Total Translations / Total Translations

For Sensitivity

Sensitivity=True Positives / True Positives + False Negatives

Substituting values

Sensitivity= 0.1×Total Translations / Total Translations

For Specificity

Specificity=True Negatives / True Negatives + False Positives

Substituting values

Specificity= Total Translations−0.1×Total Translations / Total Translations

For F1-Score

F1−Score= 2×Precision×Recall/ Precision + Recall

Substituting values

F1−Score= 2×0.01×0.1×Total Translations/0.01×Total Translations+0.1×Total Translations

Conclusion

The research on improving lung cancer detection and prediction using a hybrid CNN-RNN model uncovers several key insights and challenges. Even with extensive training over 50 epochs, the model had difficulty generalizing to new data, as shown by a persistently low validation accuracy of around 25%. This suggests issues such as overfitting, underfitting, and dataset imbalance were not adequately addressed. Although the loss values generally trended downward, indicating some learning progress, this did not lead to significant improvements in predictive performance. The RNN model achieved a higher peak accuracy of 39.2% by the seventh epoch, while the CNN model stabilized at 32% by the sixth epoch. These results highlight the complexity of optimizing hybrid models for effective lung cancer detection and prediction. Several recommendations can be made to enhance the model's performance and generalizability. Addressing overfitting and underfitting with robust data augmentation techniques is essential to create a more diverse and comprehensive dataset. Utilizing transfer learning with pre-trained models on larger, similar datasets can provide a stronger foundation for the model's learning process. Additionally, incorporating more detailed patient histories, clinical notes, and longitudinal data could improve the model's accuracy and reliability. Optimizing computational resources and using advanced hardware like GPUs and TPUs can reduce training time and enhance model efficiency. Regular validation on independent datasets and collaboration with medical professionals are crucial for refining the model and ensuring its practical application in real-world healthcare environments. Future research should aim to mitigate overfitting and dataset imbalance by employing advanced data augmentation methods and utilizing transfer learning with larger, pre-trained models. Furthermore, integrating more detailed patient information, including extensive clinical histories and longitudinal data, will improve the model's predictive accuracy and reliability in practical healthcare applications.

Supplementary Materials

The data used to support the findings of this research are available from the corresponding author upon request at.

Author Contributions

Shankar: Developed the research concept, performed data curation and formal analysis, proposed the methodology, implemented the code, assessed outcomes, and contributed to idea refinement. Supervisor Name: Conducted plagiarism checks, provided software, drafted the initial version, conducted experiments using the software, oversaw implementation, and offered software assistance. Supervised, guided, contributed to idea development, provided recommendations, conducted Asadi Srinivasulu: Plagiarism checks, and facilitated resource allocation.

Funding

The authors independently carried out this research without receiving any financial support from the institution.

Data Availability Statement

The data supporting the conclusions of this research can be obtained by reaching out to the corresponding author upon request at.

Conflicts of Interest

The authors assert that they have no conflicts of interest related to the research report on the current work.

References

1. Nair SS, Devi VNM, Bhasi S (2024) Enhanced lung cancer detection: Integrating improved random walker segmentation with artificial neural network and random forest classifier. Heliyon 10: e29032.
2. Kumar V, Prabha C, Sharma P, Mittal N, Askar SS, et al. (2024) Unified deep learning models for enhanced lung cancer prediction with ResNet-50–101 and EfficientNet-B3 using DICOM images. BMC Med Imaging 24: 63.
3. Wankhade S, Vigneshwari SA (2023) Novel Hybrid Deep Learning Method for Early Detection of Lung Cancer Using Neural Networks. Healthcare Analytics 3: 100195.
4. Gayap HT, Akhloufi MA (2024) Deep Machine Learning for Medical Diagnosis, Application to Lung Cancer Detection A Review. BioMedInformatics 4: 236-284.
5. Wahab Sait AR (2023) Lung Cancer Detection Model Using Deep Learning Technique. Appl Sci 13: 12510.
6. UrRehman Z, Qiang Y, Wang L, Shi Y, Yang Q, et al. (2024) Effective lung nodule detection using deep CNN with dual attention mechanisms. Sci Rep 14: 3934.
7. Muñoz Aseguinolaza U, Fernandez Iriondo I, Rodríguez Moreno I, Aginako N, Sierra B (2023) Convolutional neural network-based classification and monitoring models for lung cancer detection: 3D perspective approach. Heliyon 9: e21203.
8. Thanoon MA, Zulkifley MA, Mohd Zainuri MAA, Abdani SR (2023) A Review of Deep Learning Techniques for Lung Cancer Screening and Diagnosis Based on CT Images. Diagnostics (Basel) 13: 2617.
9. Wang L (2022) Deep Learning Techniques to Diagnose Lung Cancer. Cancers (Basel) 14: 5569.
10. Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, et al. (2021) Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data 8: 53
11. Siegel RL, Miller KD, Jemal A (2020) Cancer statistics. CA Cancer J Clin 70: 7-30.
12. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, et al. (2021) Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 Countries. CA Cancer J Clin 71: 209-249.
13. Nesbitt JC, Putnam JB, Walsh GL, Roth JA, et al. (1995) Mountain Survival in early-stage non-small cell lung cancer. Ann Thorac Surg 60: 466-472.
14. Ciompi F, Chung K, Gerke PK, Jacobs C, Scholten ET, et al. (2017) Towards automatic pulmonary nodule management in lung cancer screening with deep learning. Sci Rep 7: 46479.
15. Sun W, Tseng TLB, Qian W, Zhang J, Saltzstein EC, et al. (2015) Using multiscale texture and density features for near-term breast cancer risk analysis. Med Phys 42: 2853-2862.
16. Hossain MS, Muhammad G (2014) Cloud-Based Collaborative Media Service Framework for HealthCare. Int J Distrib Sens Netw 10: 858712.
17. Amin SU, Alsulaiman M, Muhammad G, Mekhtiche MA, Hossain MS (2019) Deep Learning for EEG motor imagery classification based on multi-layer CNNs feature fusion. Future Gener Comput Syst 101: 542-554.
18. Jiang H, Ma H, Qian W, Gao M, Li Y (2018) An Automatic Detection System of Lung Nodule Based on Multi-Group Patch-Based Deep Learning Network. IEEE J Biomed Health Inform 22: 1227-1237.
19. Skourt BA, El Hassani A (2018) Lung CT image segmentation using deep neural networks. Procedia Comput Sci 127: 109-113.
20. Guo Y, Li G, Xu M, Peng B (2022) A lncrna signature of tumor-infiltrating macrophages is associated with prognosis and tumor immunity in lung adenocarcinoma. Comput Biol Med 148: 105655.
21. Vernuccio F, Cannella R, Comelli A, Salvaggio G, Lagalla R, et al. (2020) Radiomica e intelligenza artificiale: Nuove frontier in medicina. Recent Prog Med 111: 130-135.
22. Parekh VS, Jacobs MA (2019) Deep learning and radiomics in precision medicine. Expert Rev Precis Med Drug Dev 4: 59-72.
23. Naqa El, Murphy MJ (2015) What Is Machine Learning? in Machine Learning in Radiation Oncology. Springer International Publishing Cham 3-11.
24. Siegel RL, Miller KD, Fuchs HE, Jemal A (2022) Cancer statistics. CA Cancer J Clin 72: 7-33.
25. Bade BC, Cruz C (2020) Lung cancer. Clin Chest Med 41: 1-24.

26. Stamatis G, Eberhard W, Pöttgen C (2004) Surgery after multimodality treatment for non-small-cell lung cancer. Lung Cancer 45: S107-S112.
27. Chiang TA, Chen PH, Wu PF, Wang TN, Chang PY, et al. (2008) Important prognostic factors for the long-term survival of lung cancer subjects in Taiwan. BMC Cancer 8: 324.
28. Journy N, Rehel JL, Pointe H, Lee C, Brisse H, et al. (2015) Are the studies on cancer risk from CT scans biased by indication? Elements of answer from a large-scale cohort study in France. Br J Cancer 112: 1841-1842.
29. National Lung Screening Trial Research Team (2011) Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med 365: 395-409.
30. Ippolito D, Capraro C, Guerra L, De Ponti E, Messa C, et al. (2013) Feasibility of perfusion CT technique integrated into conventional (18) FDG/PET-CT studies in lung cancer patients: Clinical staging and functional information in a single study. Eur J Nucl Med Mol Imaging 40: 156-165.
31. Park SY, Cho A, Yu WS, Lee CY, Lee JG (2015) Prognostic value of total lesion glycolysis by F-18-FDG PET/CT in surgically resected stage IA non-small cell lung cancer. J Nucl Med 56: 45-49.
32. Griffiths H (2011) Magnetic induction tomography. Meas Sci Technol 12: 1126-1131.
33. Brown MS, Lo P, Goldin JG, Barnoy E, Kim GHJ (2020) Toward clinically usable CAD for lung cancer screening with computed tomography. Eur Radiol 30: 1822.
34. Lung Cancer Data Set. https://www.kaggle.com/datasets/yusufdede/lung-cancer-dataset.