IJRAR.ORG E-ISSN: 2348-1269, P-ISSN: 2349-5138

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYTICAL REVIEWS (IJRAR) | IJRAR.ORG

An International Open Access, Peer-reviewed, Refereed Journal

Effective Framework for Early Lung Disease Prediction Using Deep Learning Approach

[1]Chaitra M, [2]Sadiyah Sarwat Iqbal, [3]Sushmitha Shree S, [4]Suhas T S

[1]Associate Professor, [2,3,4]Students

[1,2,3,4]Dept. of Computer Science and Engineering,

[1,2,3,4]BNM Institute of Technology, Bangalore, India

Abstract : This work presents an innovative framework for the early prediction of lung diseases employing a deep learning approach. Lung diseases such as lung cancer, pneumonia, tuberculosis, all these conditions impact the normal functioning of lungs leading to symptoms such as shortness of breath, cough and chest pain. Lung diseases poses significant global health challenges, necessitating timely and accurate diagnostic solutions. The proposed framework leverages advanced deep learning techniques, specifically convolutional neural networks (CNNs), to analyze medical images such as chest X-rays and CT scans. By harnessing the power of deep learning, the framework aims to enhance diagnostic accuracy and enable proactive intervention, ultimately improving patient outcomes. The study acknowledges the current limitations in lung disease diagnostics which mainly focuses on the need for accurate precision and early detection of lung diseases. Through comprehensive experimentation and validation, this research aims to contribute a robust and effective tool to the medical community, fostering advancements in the timely identification and management of diverse lung conditions. The anticipated impact of this framework extends to its potential integration into healthcare systems, supporting clinicians in making informed decisions and ultimately benefiting individuals susceptible to respiratory ailments.

***IndexTerms* - Deep Learning, CNN, Lung disease detection, X-rays**

I. INTRODUCTION

Global health is faced with major problems due to lung diseases, which impact millions of people globally and cause conditions like lung cancer, pneumonia, and tuberculosis. Breathless-ness, coughing, and chest pain are just a few of the symptoms that indicate that serious long-term conditions affect respiratory function. A deep learning technique is used in this research to propose a unique framework for the early prediction of lung diseases, recognizing the crucial need for timely and accurate diagnostic solutions. The usual method for diagnosis of lung diseases often faces limitations, particularly in terms of precision and early detection. To overcome these challenges, our proposed framework uses deep learning techniques, particularly focusing on convolutional neural networks (CNNs), to analyze medical images such as chest X-rays and CT scans. This study recognizes the critical need to create instruments that address existing diagnostic gaps and aid in the early diagnosis and treatment of a wide variety of lung diseases. By exploring of deep learning, we target to significantly enhance diagnostic accuracy, enabling prepared intervention that can finally lead to improved patient outcomes. Our study is to provide the medical community with a reliable and efficient solution through thorough testing and validation. This framework is expected to have an influence that goes beyond research, with the goal of being easily integrated into healthcare systems. This integration will ultimately help those who are likely to get infected respiratory conditions as it gives the way for future developments in case of lung health by assisting doctors in making well-informed conclusions. Patient outcomes for lung cancer can be greatly enhanced by early identification and precise diagnosis. After diagnosis, lung cancer patients have a 10 to 20 percent chance of surviving for five years. Magnetic resonance imaging (MRI) and computed tomography (CT) are frequent medical treatments for early detection that increase patient survival. Many medical image-processing approaches have been effectively implemented as a result of recent advancements in deep learning technology, which have allowed CAD systems to independently identify visual elements.

II. RELATED WORK

The field of early lung disease prediction has made substantial progress by incorporating deep learning methods, providing a more precise and advanced approach to detect potential respiratory abnormalities. A key focus has been on leveraging deep learning models for image-based detection, especially in analyzing medical images like chest X-rays and CT scans. Convolutional Neural Networks (CNNs), frequently enhanced with transfer learning methods, have shown re- markable effectiveness in classifying images and detecting subtle abnormalities associated with various lung diseases, such as pneumonia, tuberculosis, and lung cancer.

In [1], the paper discusses detecting and categorizing lung ailments, including pneumonia, tuberculosis (TB), and COVID-19, using deep learning-based pretrained models and ensembling techniques. It highlights the global significance of lung diseases on public health and emphasizes the need to address environmental factors, particularly air pollution and climate change, to mitigate their impact. The paper underscores the worldwide importance of lung diseases on public health and stresses the urgency of tackling environmental factors, specifically air pollution and climate change, to reduce their impact. Additionally, it introduces a method for identifying lung diseases that utilizes multichannel EfficientNet ImageNet based pretrained models and a stacking ensembling strategy, demonstrating superior performance compared to current methods. It offers in- sights into ongoing research and advancements in lung disease detection, demonstrating the potential of advanced computational techniques in addressing critical healthcare challenges.

In [2], the paper examines lung disease classification, particularly COVID-19, through chest X-ray images, employing deep learning and machine learning techniques. It reviews recent advancements, identifies research gaps, and proposes contributions, emphasizing the fusion of deep learning, machine learning, and computer vision for classification. Limitations include a lack of disease severity analysis, difficulty in estimating the Region of Interest (ROI), and absence of real-time X-ray analysis, indicating a need for refinement. In conclusion, the authors propose HDLA, a framework integrating preprocessing and classification steps to detect lung diseases using deep learning and machine learning techniques applied to chest X-ray images.

In [3], the paper explores the utilization of artificial intelligence (AI) and deep learning (DL) methods in detecting Covid-19 from X-ray images. The Covid-19 pandemic, caused by the SARS-CoV-2 virus, has prompted extensive research because to its global impact. Traditional testing methods, like reverse transcription-polymerase chain reaction (RT-PCR), have limitations, such as time delays and the requirement of specialized equipment. CT and X-ray imaging, target on lung infections, offer alternative diagnostic approaches. The rise of AI and DL in this context starts from their ability to assist in early and accurate disease detection, overcoming challenges associated with RT-PCR. Researchers have established various deep learning models, such as CovNNet, ResNet50V2, and different hybrid models, to analyze radiological images. Some studies criticize the sensitivity of RT-PCR, prompting the exploration of AI-based alternatives. Notable models like CoroNet, CovidXrayNet, and CovXNet showcase the diversity of approaches. Challenges include model complexity and the need for extensive datasets. The literature survey emphasizes the growing significance of AI and DL in enhancing the efficiency and accuracy of COVID-19 detection from medical images. These advancements demonstrate a shift towards leveraging AI and DL not only for improved accuracy but also for overcoming practical challenges associated with traditional testing methods.

In [4], the paper delves into the implementation of the VDSNet algorithm, a deep learning approach that combines CNN, VGG, data augmentation, and spatial transformer network (STN). The study applies this method to a X-ray dataset from Kaggle, which includes attributes like age, X-ray images, gender, and view position. The dataset's complexity, substantial size (considered big data), and noise pose significant data processing challenges. The primary objective is binary classification to predict lung diseases, labeled as ”Yes” or ”No.” The focus is on patient classifying using CNN deep learning methods for X-ray images. The survey introduces Capsule Network (CapsNet) as an effective algorithm with generative and deterministic capabilities, though noted for its image sensitivity compared to simpler CNN structures. Section 7 of the paper conducts a thorough comparative analysis, evaluating the performance of VDSNet against CapsNet, modified CapsNet, and other established deep learning techniques. The results highlight the superior predictive accuracy of VDSNet in diagnosing lung diseases from X-ray images, contributing significantly to the field's advancement. The paper's significance lies in the development of VDSNet, providing an effective approach surpassing existing methods in accuracy for lung disease prediction.

In [5], this paper explores the landscape of lung diseases, considering the rise in cases globally and the challenges faced by developing countries, like India, in providing adequate medical support. It emphasizes the diversity of lung diseases, including an upsurge in lung cancer cases, Chronic Obstructive Pulmonary Disease (COPD), and lower respiratory diseases. Factors such as exposure to toxic elements from smoking contribute to the destruction of airways and the developing of COPD. The text underscores the effect of travel on the spread of diseases like Middle East Respiratory Syndrome (MERS) and the consideration of viruses associated with travel destinations. Pneumonia, caused by infections in the lower respiratory tract, is prevalent, and exposure to smoking toxins contributes to its development. The paper shifts focus to machine learning (ML) techniques, particularly deep learning, in addressing lung disease detection. It acknowledges the challenges in data collection for training deep learning models, especially in medical diagnoses where data confidentiality is a concern. The importance of choosing a suitable model architecture and hyperparameters is highlighted, with a recognition of the need for high-quality training data. The primary goal of the paper is to analyze various deep learning models for lung disease detection, examining related works, discussing different model insights, and ultimately proposing an optimal model based on accuracy and performance.

In [6], the paper suggests a novel framework for predicting lung conditions like pneumonia and COVID-19 from chest X-ray images of patients. With the emergence of COVID-19, precise prognosis of illness has become paramount, especially given its similarities with pneumonia. The framework comprises several key steps: dataset acquisition utilizing publicly available chest X-ray datasets, image quality enhancement to address degradation in image quality, adaptive and accurate region of interest (ROI) estimation, feature extraction including visual, shape, texture, and intensity features, and disease anticipation. In order to improve the accuracy of disease detection, for precise ROI extraction, a modified region-growing approach is introduced, along with normalization techniques to improve feature identification and categorization. The paper employs various techniques for soft computing like artificial neural networks (ANN), support vector machines (SVM), K-nearest neighbor (KNN), ensemble classifiers, and deep learning classifiers, recurrent neural networks (RNN) among them with long short-term memory (LSTM), for disease classification. In [7], the main topic of this paper is the detection of lung conditions with the use of various deep learning algorithms. The authors likely explore the application of varying deep learning techniques to identify and classify lung diseases, such as pneumonia or COVID-19, according to medical imaging data like chest X-rays or CT scans. These algorithms may include convolutional neural networks (CNNs), recurrent neural networks (RNNs), or other architectures specifically designed for medical image analysis. The work likely discusses the methodology involved in training these algorithms on datasets containing labeled examples of lung images with various disease conditions. The main aim is to develop accurate and efficient models that can help healthcare professionals in diagnosing lung diseases more effectively. By leveraging deep learning techniques, the authors aim to contribute to advancements in medical imaging and improve patient care in the context of respiratory health.

III. PROPOSED METHOD

This methodology integrates database image selection, image collection, image processing, image segmentation, feature extraction, and classification to develop a CNN-based framework for early lung disease prediction from chest X-ray images. The approach emphasizes robustness, interpret-ability and validation on diverse datasets to enhance the effectiveness of the prediction model. This methodology establishes a holistic approach, seamlessly integrating database image selection, advanced image processing, and deep learning techniques for lung disease prediction. The incorporation of image segmentation and feature extraction enriches the model's ability to discern intricate patterns associated with early-stage lung diseases. The model's interpretability is heightened through visualization, ensuring transparency in decision-making and its robustness is validated across diverse datasets, affirming its potential as an effective framework for early lung disease prediction using a CNN-based approach. Furthermore, the systematic integration of image collection and classification techniques enhances the model's scalability and adaptability, fostering its utility in diverse clinical scenarios for early detection of lung diseases. Figure 1 shows the system architecture of the proposed method.

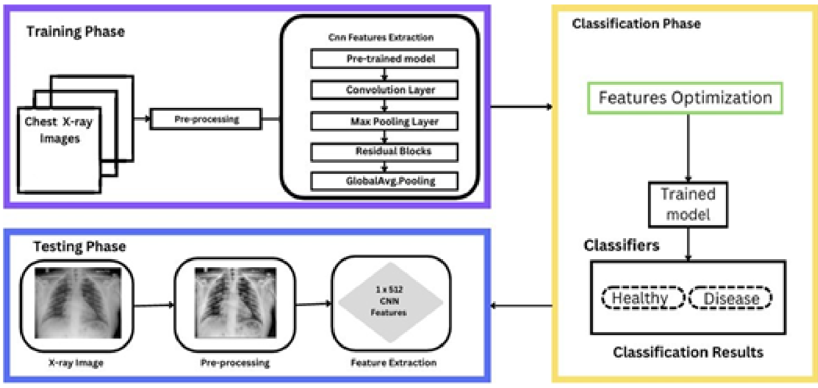

Fig. 1. System Architecture for Lung Disease Prediction

The workflow of the proposed method is as follows:

A. Database Image Selection

Curate a comprehensive database of chest X-ray images containing diverse cases, emphasizing early-stage lung diseases. Consider datasets such as Kaggle.

B. Image Collection and Annotation

Collect a large set of chest X-ray images from diverse sources, ensuring the inclusion of both normal and diseased cases. Annotate the images with ground truth labels, specifying the presence and type of lung disease, to facilitate supervised learning.

C. Image Preprocessing

Apply image preprocessing techniques, including resizing, normalization, and histogram equalization, to enhance the quality and consistency of the dataset. Explore data augmentation methods to artificially increase the dataset size and improve model generalization.

D. Image Segmentation

Implement image segmentation techniques to focus on the region of interest, i.e., the lung area, within the chest X-ray images. Utilize segmentation algorithms, possibly based on thresholding or deep learning approaches, to isolate the lungs from the background.

E. Feature Extraction from Segmented Images:

Extract relevant features from the segmented lung regions to capture disease-specific patterns. Leverage techniques such as texture analysis, edge detection, or deep feature extraction through pre-trained CNN layers for informative feature representation.

F. Classification Model Design:

Select a suitable CNN architecture, such as a modified VGG or ResNet, for image classification based on the extracted features Design the classification model with input layers accommodating the extracted features and output layers corresponding to different lung disease classes.

G. Training the Classification Model:

Split the dataset into training, validation, and testing subsets for training and evaluating the classification model. Train the CNN-based classifier using the training set, optimizing for disease classification based on the extracted features.

H. Performance Evaluation Metrics:

Evaluate the classification model's performance using metrics such as accuracy, precision, recall, F1-score, and AUC- ROC, specifically considering early-stage disease prediction metrics.

I. Interpretability and Visualization:

Implement visualization techniques, like Grad-CAM, to interpret the model's decision-making process and understand the regions crucial for classification. Provide visual explanations to enhance the interpretability of the deep learning model.

J. Validation on External Databases:

Validate the trained model on external chest X-ray databases to ensure its robust- ness and generalizability across different datasets.

IV. IMPLEMENTATION

The implementation process for the Front-End Developmentusing Python Tkinter involves leveraging Tkinter, the standard GUI library for Python. Tkinter simplifies the creationof GUI ap- plications by offering an object-oriented interface to the Tk GUI toolkit. Its cross-platform compatibility ensures that applications developed using Tkinter seamlessly run on Windows, macOS, and Linux. Additionally, Tkinter utilizes native operating system elements for rendering visual elements, ensuring that applications blend seamlessly with the platform's aesthetics. This approach not only streamlines the development process but also enhances the user experienceby providing a familiar interface regardless of the underlying operating system.

During the implementation phase, attention is directed to- wards translating design specifications into source code effectively. This entails writing clear and straightforward source code accompanied by internal documentation to facilitate easy verification of code conformance to specifications. By prioritizing simplicity, clarity, and elegance in each program module, the implementation aims to minimize complexities and streamline debugging, testing, and modification processes. Furthermore, adhering to these principles enhances the maintainability and scalability of the codebase, ensuring that future development efforts are facilitated rather than impeded.

The implementation of the algorithm involves several key steps, including data processing, model creation, and disease stage analysis. Data processing involves preparing the verification data by resizing images and organizing them for analysis. Subsequently, a convolutional neural network (CNN) model is constructed using TensorFlow and TFLearn, incorporating layers for image convolution, pooling, and fully connected layers. Once the model is trained, it is utilized for disease stage analysis, where images are processed to identify specific features indicative of disease progression. By leveraging advanced image processing techniques and machine learning algorithms, the implementation facilitates automated disease diagnosis, contributing to the advancement of medical tech- nology and healthcare delivery. Figure 2 shows the Register Module.

Fig. 2. Register Module

This module is responsible for users to register and login using their desired credentials which includes User Name, First Name, Last Name and password. After the verification of the login credential, and if it is found correct, the user will be taken to the home page or Dashboard and the session is started. Figure 3 shows the representation of Detection Module.

Fig. 3. Detection module

This module is responsible for taking X-ray image and use that image to analyze whether covid is present or not. It takes X-ray images as a input and it gives the prediction based on the input. It imports essential libraries for image processing, data handling, and deep learning tasks. It sets up constants such as directories for training and testing data, image size, learning rate, and model name. Data processing is then performed, where images from the training and testing directories are read, resized to a specified dimension, and labeled based on their filenames. These processed images are stored as NumPy arrays, ready for training. The first step in pre-processing is converting the image from RGB to Grayscale. It can be obtained by applying the below formula to the RGB image. Figure 4 depicts the Conversion from RGB to grayscale.

Noise removal algorithm is the process of removing or reducing the noise from the image. The noise removal algorithms reduce or remove the visibility of noise by smoothing the entire image leaving areas near contrast boundaries. Noise removal is the second step in image pre-processing. Here the grayscale image which was obtained in the previous step is given as input. Here we are making use of Median Filter which is a Noise Removal Technique.

The median filter is a non-linear digital filtering technique, often used to remove noise from an image or signal. Here 0's are appended at the edges and corners to the matrix which is the representation of the grey-scale image. Then for every 3*3 matrix, arrange elements in ascending order, then find the median/middle element of those 9 elements, and write that median value to that particular pixel position. Thresholding is a type of image segmentation, where we change the pixels of an image to make the image easier to analyze. A(i,j) is greater than or equal to the threshold T, retain it. Else, replace the value by 0. Here, the value of T can be manipulated in the frontend, to suit the varying needs of different images. We use trial and error method here to obtain threshold value which may be best suited for us. Thresholding using basic global thresholding. Image sharpening refers to any enhancement technique that highlights edges and fine details in an image, Increasing yields a more sharpened image. A high-pass filter can be used to make an image appear sharper. These filters emphasize fine details in the image. Here the output from the thresholding is given as input. Here, we are making use of a filter, first we append the nearest values to pixels at the boundary pixels. The convolutional neural network (CNN) model architecture is then defined using tflearn. It consists of several convolutional layers for feature extraction, followed by max-pooling layers to reduce spatial dimensions. Fully connected layers are added for classification and a softmax output layer is used for multi-class prediction.

The model training phase begins by checking if a pre-trained model exists. If found, it loads the pre-trained weights; otherwise, it initializes a new model. The training data is split into training and validation sets. The model is trained for 100 epochs using the Adam optimizer, with categorical cross-entropy as the loss function. Training progress is monitored, and validation is performed on a separate dataset. Once training completes, the trained model is saved for future use, enabling easy deployment and inference on new data. Figure 4-8 represents a result of different lungs affected images.

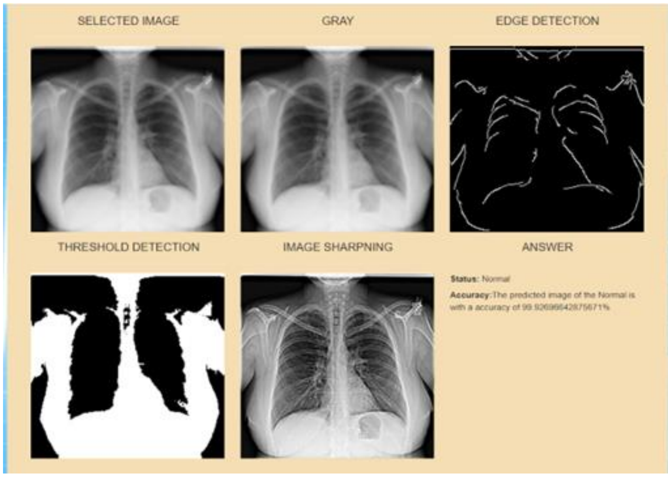

Fig. 4. Results of a normal lung

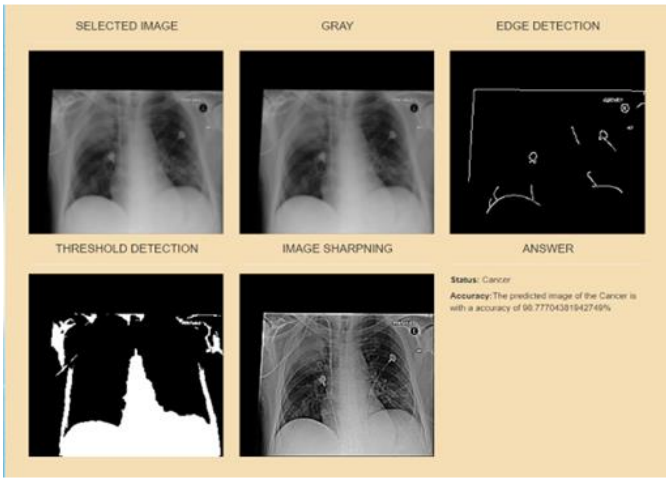

Fig. 5. Results of a cancer-affected lung

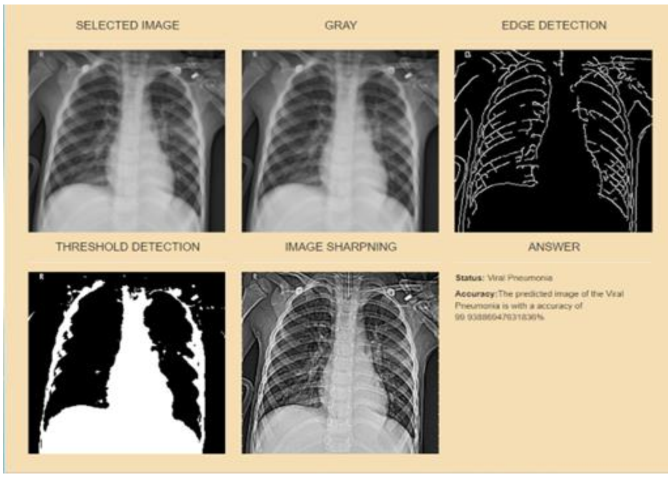

Fig. 6. Results of a pneumonia affected lung

Fig. 7. Results of a tuberculosis-affected lung

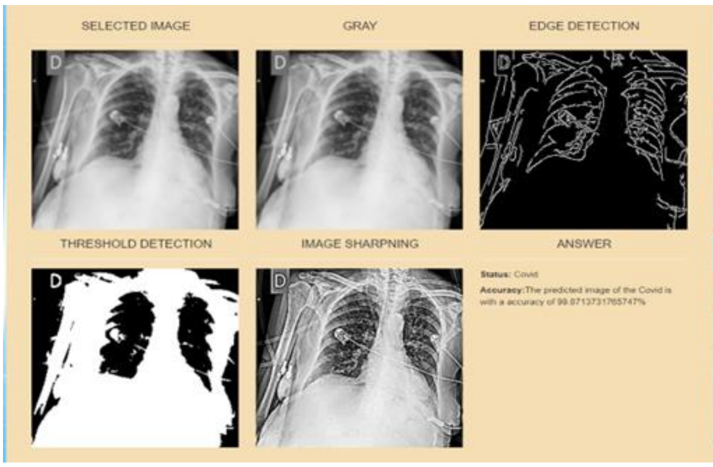

Fig. 8. Results of a covid affected lung

V. CONCLUSION AND FUTURE SCOPE

In essence, the proposed methodology uses advanced techniques in deep learning and image processing to create an integrated approach to early lung disease prediction. A robust and relevant dataset is the result of the comprehensive selection and gathering of a different chest X-ray database that includes cases of early-stage lung diseases. Accurate model training is made possible by the subsequent steps of image annotation and collecting, which are followed by thorough image preparation. These procedures assure the consistency and quality of the dataset. In conclusion, flexible and comprehensible model is produced by the complete integration of image processing, segmentation, feature extraction, and classification inside a CNN-based framework. This methodology adds to the progress of diagnostic functions by dealing with the challenges of early-stage lung disease prediction. It provides an appealing path for improving the outcomes of patients through accurate and timely procedures.

In the future, the framework might explore an assortment of patient-focused data, including genetic profiles, physician notes, and electronic medical records, in along with chest X-ray pictures. This expansion could improve diagnostic precision by presenting a more complex image of particular patients. Real-time insights could be gained by analysing data inputs from wearable devices or continuous monitoring systems, which allows for a more flexible and adaptable approach.

One of our top priorities for further improvements is to make sure that our deep learning framework is reliable and easy to understand. Transparent AI systems are necessary given the changing healthcare environment. Thus, future iterations might focus on designing deep learning frameworks that are naturally accessible. Methods like layer-wise relevance propagation or attention processes could act as arrows, revealing perspective on the model's decision-making process. This kind of strategy builds end-user trust while also helping healthcare professionals understand predictions. Deep learning approach to early lung disease prediction will ultimately depend on how strategically we approach transparency and adaptability going ahead. Through expanded data integration, improved interpretability, and focusing on ongoing learning, our framework aims to transform into a reliable partner that guides diagnosis precision and improves patient outcomes in the complex field of lung health.

REFERENCES

[1] Ravi, V., Acharya, V. & Alazab, M. A multichannel EfficientNet deep learning-based stacking ensemble approach for lung disease detection using chest X-ray images. Cluster Comput 26, 2023. https://doi.org/10.1007/s10586-022-03664-6

[2] Farhan, A.M.Q., Yang, S. Automatic lung disease classification from the chest X-ray images using hybrid deep learning algorithm. Multimed Tools Appl 82, 38561–38587, 2023. https://doi.org/10.1007/s11042-023-15047-z

[3] Gaffari Celik, Detection of Covid-19 and other pneumonia cases from CT and X-ray chest images using deep learning based on feature reuse residual block and depthwise dilated convolutions neural network, Applied Soft Computing, Volume 133, 2023, https://doi.org/10.1016/j.asoc.2022.109906

[4] Subrato Bharati, Prajoy Podder, M. Rubaiyat Hossain Mondal, Hybrid deep learning for detecting lung diseases from X-ray images,Informatics in Medicine Unlocked, Volume 20, 2020, https://doi.org/10.1016/j.imu.2020.100391

[5] Deepapriya, B.S., Kumar, P., Nandakumar, G. et al. Performance evaluation of deep learning techniques for lung cancer prediction. Soft Comput 27, 9191–9198, 2023. https://doi.org/10.1007/s00500-023-08313-7

[6] Goyal, S., Singh, R. Detection and classification of lung diseases for pneumonia and Covid-19 using machine and deep learning techniques. J Ambient Intell Human Comput 14, 3239–3259, 2023. https://doi.org/10.1007/s12652-021-03464-7

[7] M. Jasmine Pemeena Priyadarsini, Ketan kotecha, G. K. Rajini, K. Hariharan, K. Utkarsh Raj, K. Bhargav Ram, V. Indragandhi, V. Subramaniyaswamy, Sharnil Pandya, "Lung Diseases Detection Using Various Deep Learning Algorithms", Journal of Healthcare Engineering, vol. 2023, Article ID 3563696, 13 pages, 2023. https://doi.org/10.1155/2023/3563696

[8] Chaitra, Y.L., Dinesh, R., Gopalakrishna, M.T. et al. Deep-CNNTL: Text Localization from Natural Scene Images Using Deep Convolution Neural Network with Transfer Learning. Arab J Sci Eng 47, 9629–9640, 2022. https://doi.org/10.1007/s13369-021-06309-9

[9] Chaitra Y.L., Dinesh R., An Impact of Radon Transforms and Filtering Techniques for Text Localization in Natural Scene Text Images. ICT with Intelligent Applications. Smart Innovation, Systems and Technologies, vol 248. Springer, Singapore, pp. 563-573, 2022. https://doi.org/10.1007/978-981-16-4177-0_55