CNN-based Diagnosis from Medical Imaging: Leveraging Transfer Learning for Enhanced Accuracy

Nathan Wan
University of Western Ontario
nwan23@uwo.ca

Kevin Du
University of Western Ontario
kdu22@uwo.ca

Matthew Louis Li
University of Western Ontario
mli2646@uwo.ca

Juna Kim
University of Western Ontario
jkim2983@uwo.ca

Millicent Song
University of Western Ontario
msong257@uwo.ca

Besma Serrai
University of Western Ontario
bserrai@uwo.ca

Liam McQuay
University of Western Ontario
lmcquay2@uwo.ca

Artemiy Vishnyakov
University of Western Ontario
avishny@uwo.ca

Sabrina Lee
University of Western Ontario
slee2893@uwo.ca

***Abstract*—Medical imaging plays a crucial role in disease diagnosis and treatment planning, yet the increasing volume of imaging data poses challenges for radiologists and healthcare systems. This study investigates the application of Convolutional Neural Networks (CNNs) for automated medical image classification, leveraging transfer learning to enhance diagnostic accuracy. Using a pre-trained DenseNet-121 model, we developed and evaluated CNN-based classifiers for lung cancer, pneumonia, and tuberculosis detection.**

Our models achieved high classification accuracy, with 90.48% for lung cancer, 91.83% for pneumonia, and 99.84% for tuberculosis, demonstrating their effectiveness in distinguishing between normal and pathological cases. The results highlight the potential of AI-driven diagnostics to assist medical professionals by reducing workload, improving diagnostic speed and accuracy, and addressing the shortage of radiologists.

Despite promising performance, challenges such as dataset variability, potential biases, and real-world deployment remain. Future work will focus on expanding datasets, improving model interpretability, and integrating AI-assisted tools into clinical workflows to enhance reliability and accessibility in medical imaging.

I. Introduction

Medical imaging is a cornerstone of modern healthcare, enabling the early detection and diagnosis of a wide range of diseases. However, the increasing volume of medical imaging data presents challenges for radiologists and healthcare providers, who must analyze complex scans under time constraints. Artificial intelligence (AI), particularly deep learning, with its advantage of end-to-end processing, has emerged on a large scale in recent medical diagnosis studies [1], offering the potential for faster, more accurate, and more efficient analysis. This research explores the development of an AI-driven medical imaging diagnostic system using convolutional neural networks (CNNs) to detect anomalies in X-rays, magnetic resonance imaging (MRI), and computed tomography (CT) scans.

Deep learning, a subset of machine learning, has demonstrated remarkable capabilities in image recognition and classification tasks, making it an ideal tool for medical imaging diagnostics. Convolutional neural networks (CNNs) are particularly well-suited for this purpose because they automatically extract hierarchical features from medical images, distinguishing between normal and pathological conditions with high accuracy. One such model in skin cancer achieved levels of competence comparable to dermatologists [2]. Unlike traditional image processing techniques that rely on handcrafted feature extraction, CNNs learn spatial hierarchies of features directly from raw imaging data [3], leading to superior performance concerning old approaches [4].

Convolutional neural networks (CNNs) have empowered researchers to develop advanced models that accurately differentiate between healthy patients and those showing signs of cancer [5]. These sophisticated algorithms have improved speed and accuracy, marking a substantial technological breakthrough in medical diagnostics. This advancement has the potential to significantly transform how cancer is diagnosed and can greatly improve patient outcomes by facilitating earlier and more accurate detection, thereby enabling prompt and effective treatment. Incorporating CNNs into cancer diagnostics signifies a crucial milestone in integrating artificial intelligence within healthcare, promising to boost diagnostic accuracy and streamline medical interventions. Consequently, this technology could substantially reduce the mortality rates and overall global burden associated with cancer.

A. Motivation

The growing demand for medical imaging diagnostics has placed significant pressure on healthcare systems, leading to increased workloads for radiologists and the risk of diagnostic errors. AI, particularly deep learning models like CNNs, has demonstrated immense potential in enhancing diagnostic accuracy and efficiency [2]. AI can rapidly analyze complex imaging data, detect abnormalities with high precision, and support medical professionals in making timely and informed decisions [6]. This project is motivated by the need to integrate AI into medical imaging workflows to improve patient outcomes. Studies have shown that AI-assisted diagnostics can significantly reduce interpretation time while maintaining or exceeding human-level accuracy [7]. Moreover, AI can address the shortage of radiologists in many regions, ensuring broader access to high-quality healthcare services.

By developing a CNN-based diagnostic system for X-rays, MRIs, and CT scans, this study aims to contribute to the ongoing transformation of medical imaging. The project focuses on improving the interpretability and reliability of AI models, ensuring they can be effectively integrated into clinical practice [8]. Ultimately, this work seeks to demonstrate that AI-powered diagnostics can enhance healthcare efficiency, reduce diagnostic errors, and support medical professionals in providing better patient care.

B. Related Works

Our research focuses on adapting a pre-trained convolutional neural network (CNN) model for general-use medical diagnostics, specifically for detecting anomalies in X-rays, MRIs, and CT scans using PyTorch. The growing body of research on CNN applications in medical imaging underscores the potential of deep learning models to enhance diagnostic accuracy, streamline workflows, and assist radiologists in detecting diseases more efficiently. A key component of our approach is **transfer learning**, which allows us to fine-tune a pre-trained CNN for medical imaging tasks, reducing the need for extensive labeled datasets while improving model performance. This method aligns with previous work, such as that by Varma et al., who demonstrated that transfer learning significantly enhances CNN-based image classification [8].

Beyond transfer learning, CNNs have shown expert-level performance across multiple medical imaging domains. In dermatology, CNN-based models have matched or surpassed human specialists in classifying skin conditions, proving that deep learning can be adapted beyond traditional radiology applications [2]. Similarly, in oncology, AI models such as the **Chief AI system** developed at Harvard achieved a **94 % accuracy rate** in cancer detection, demonstrating the capability of CNNs to assist in early cancer diagnosis with high precision [9]. CNNs have also been widely explored in neurology, where they have been successfully applied to detect **Alzheimer's-related tauopathy**, showcasing their ability to interpret intricate pathological patterns in complex neuroimaging data [10].

Additionally, CNNs have contributed to advancements in MRI imaging, particularly in **motion correction and artifact reduction**, which improve the quality of MRI scans and minimize the need for repeat imaging, as shown in recent studies [12]. These collective findings reinforce the **broad applicability of CNNs in medical imaging diagnostics** and validate our approach in leveraging deep learning for automated anomaly detection. By incorporating these methodologies, our research aims to develop an AI-driven diagnostic tool that enhances medical imaging analysis, improves interpretability, and supports clinical decision-making.

C. Problem Definition

Our research explores the deployment of an image recognition algorithm, specifically a convolutional neural network (CNN), within the healthcare industry to enhance medical imaging diagnostics. The distinctiveness of this project lies in its approach: leveraging transfer learning by adapting pre-trained CNN models to create a general-use diagnostic tool

To facilitate model training, we utilize a medical imaging dataset, ensuring access to high-quality labeled data. The images undergo preprocessing to enhance compatibility with our CNN, which is implemented using PyTorch. We seek to maximize model accuracy and generalizability across different imaging modalities through careful fine-tuning, including hyperparameter optimization and augmentation techniques.

II. Methodology

A. Dataset

This study utilizes three distinct medical imaging datasets for disease classification:

- **Lung Cancer Image Dataset: A Comprehensive Collection 2024** [11] – Consists of **high-resolution CT scan images** for lung cancer classification. The dataset is divided into four distinct classes:
  - **Adenocarcinoma** – A common form of lung cancer originating in mucus-producing cells.
  - **Large Cell Carcinoma** – A rapidly growing lung cancer type appearing in any lung region.
  - **Normal** – Healthy lung images serving as control samples.
  - **Squamous Cell Carcinoma** – A cancer developing in the flat cells lining the airways.
- **Pneumonia Chest X-ray Dataset** – Contains 5,863 X-ray images (JPEG) classified into two categories: Pneumonia and Normal. The dataset is organized into three subsets (train, test, validation). The images were selected from retrospective cohorts of pediatric patients aged one to five years old from Guangzhou Women and Children's Medical Center. [12]
- **Tuberculosis Chest X-ray Dataset** – Contains 7,000 X-ray images (3,500 TB-positive and 3,500 normal). The dataset was compiled by researchers from Qatar University, the University of Dhaka, and their collaborators. The TB images were sourced from publicly accessible datasets and the NIAID TB portal program. [13]

Each dataset was split into training, validation, and testing sets to facilitate model evaluation.

B. Data Preprocessing

To optimize the datasets for **CNN-based image recognition**, a preprocessing pipeline was implemented using **PyTorch's torchvision.transforms**. The key preprocessing steps include:

- **Resizing**: Images were resized to ensure uniform input dimensions.
- **Grayscale Conversion**: Images were converted to a **single-channel grayscale format** to reduce computational complexity while preserving diagnostic features.
- **Normalization**: Pixel values were **scaled to the [0,1] range** by dividing by 255 to aid neural network convergence.
- **Data Augmentation**: To enhance model generalization, **random horizontal flipping** and **random rotation** were applied.
- **Mean-Std Normalization**: The dataset **mean and standard deviation** were computed and applied for stabilization during training.

These preprocessing steps ensure that the datasets are **well-structured, optimized for deep learning, and suitable for medical image classification**.

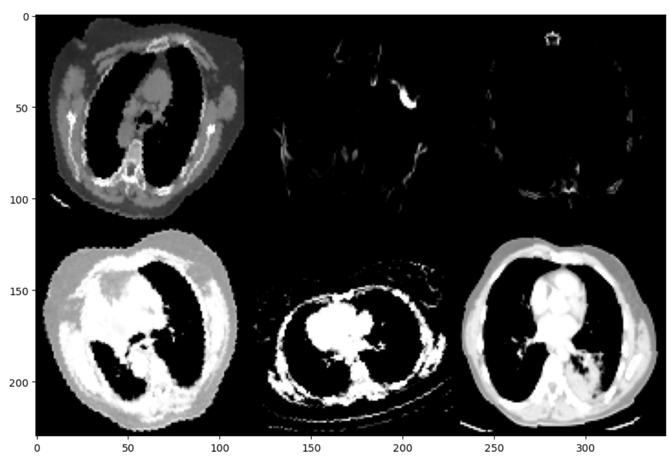

Fig. 1. Preprocessed lung CT scan images used for CNN-based image recognition. The preprocessing pipeline, implemented using PyTorch's torchvision.transforms, includes resizing for uniform input dimensions, grayscale conversion for reduced computational complexity, and pixel normalization to the [0,1] range for improved neural network convergence. Additionally, data augmentation techniques such as random horizontal flipping and rotation were applied to enhance model generalization. Mean-standard deviation normalization (mean = 0.3230, std = 0.2176) was performed to stabilize training. These steps optimize the dataset for deep learning-based lung cancer classification.

C. Model Architecture and Training

A **Convolutional Neural Network (CNN)-based approach** was employed using **transfer learning** with a **pre-trained DenseNet-121 model**. Three separate models were trained, each specialized for one of the three datasets (lung cancer, pneumonia, tuberculosis). The modifications and training approach are detailed below:

1) Model Architecture:

- The **pre-trained DenseNet-121** model was adapted.
- The **fully connected classifier layer** was replaced with a **linear classifier** for multi-class output (four classes for lung cancer, two classes each for pneumonia and tuberculosis).
- A **cross-entropy loss function** was utilized for classification.
- The **SGD optimizer with momentum (0.9) and learning rate decay** was applied for stable training.

2) Training Process:

- Each model was trained separately on its respective dataset using a **batch size of 32**.
- Training was conducted on a **GPU** (if available) for faster convergence.
- A **learning rate scheduler** adjusted the learning rate every 7 epochs to mitigate overfitting.
- The validation sets were used for hyperparameter tuning.

D. Evaluation Metrics

To rigorously assess each model's ability to differentiate between true-positive and true-negative cases, this study employs a comprehensive set of evaluation metrics. Specifically, we focus on **precision, recall (sensitivity), F1-score**, and insights derived from the **confusion matrix**.

Precision and recall are fundamental in evaluating the model's ability to correctly identify true-positive cases. Their respective formulas are:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{1}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{2}$$

The **F1-score**, which is the harmonic mean of precision and recall, quantifies the trade-off between false positives and false negatives and is defined as:

$$\text{F1-score} = 2 \cdot \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3}$$

Additionally, the **confusion matrix** provides a detailed breakdown of the model's performance, offering valuable insights into its strengths and weaknesses. This tool is especially crucial in classification tasks where the cost of misclassification varies in severity, such as medical diagnosis.

III. Results

A. Overall Model Performance

Three separate CNN-based models were trained for lung cancer, pneumonia, and tuberculosis classification. The overall test accuracies achieved for each model are:

- **Lung Cancer Model**: 90.48%
- **Pneumonia Model**: 91.83%
- **Tuberculosis Model**: 99.84%

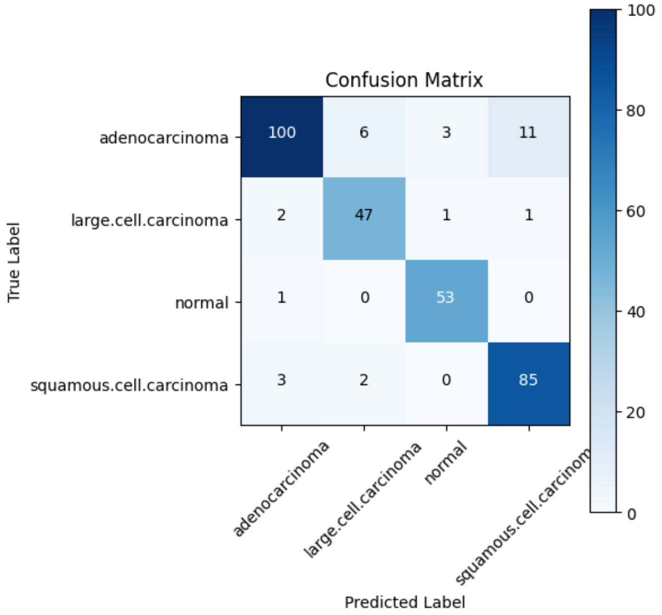

Fig. 2. Confusion matrix illustrating the performance of the CNN-based lung cancer classification model. The model demonstrates high classification accuracy for normal cases (98%) and squamous cell carcinoma (94%), but adenocarcinoma exhibits a slightly lower recall (83%).

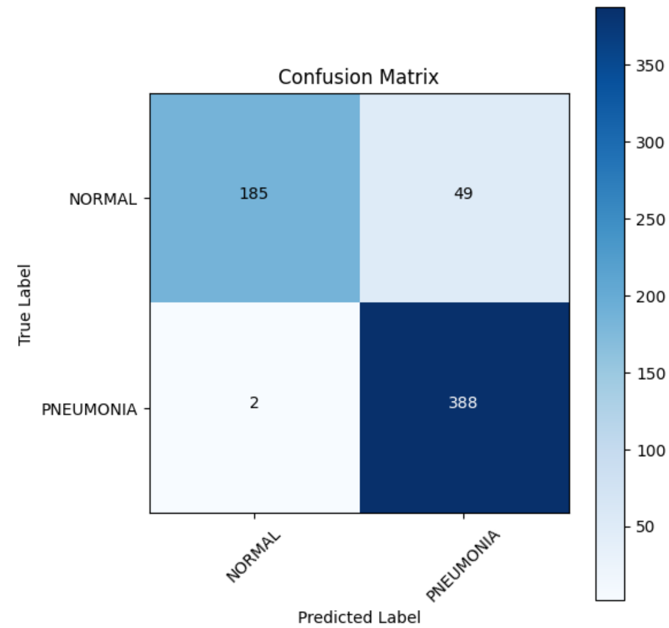

Fig. 3. Confusion matrix illustrating the performance of the CNN-based pneumonia classification model. The model achieved a recall of 99% for pneumonia cases, ensuring a minimal number of false negatives. However, normal cases exhibited a lower recall (79%), with some misclassification occurring.

These results demonstrate the effectiveness of the models in distinguishing between various medical conditions based on imaging data.

B. Lung Cancer Classification Results

The performance of our CNN-based lung cancer classification model was evaluated using a test dataset of 315 images, achieving an overall test accuracy of extbf90.48%. This demonstrates the model's effectiveness in distinguishing between adenocarcinoma, large cell carcinoma, squamous cell carcinoma, and normal lung conditions using CT scans.

1) Confusion Matrix Analysis: The confusion matrix provides insights into the model's classification performance:

- **Adenocarcinoma**: Correctly classified: 100, Misclassified as large cell carcinoma: 6, normal: 3, and squamous cell carcinoma: 11.
- **Large Cell Carcinoma**: Correctly classified: 47, Misclassified as adenocarcinoma: 2, normal: 1, squamous cell carcinoma: 1.
- **Normal Cases**: Correctly classified: 53, Misclassified as adenocarcinoma: 1.
- **Squamous Cell Carcinoma**: Correctly classified: 85, Misclassified as adenocarcinoma: 3, large cell carcinoma: 2, normal: 0.

C. Pneumonia Classification Results

The pneumonia classification model was evaluated using a test dataset of 624 images, achieving an overall test accuracy of **91.83%**.

- **Normal**: Correctly classified: 185, Misclassified as pneumonia: 49.
- **Pneumonia**: Correctly classified: 388, Misclassified as normal: 2.

D. Tuberculosis Classification Results

The tuberculosis classification model was evaluated using a test dataset of 630 images, achieving an outstanding test accuracy of **99.84%**.

- **Normal**: Correctly classified: 525, Misclassified as tuberculosis: 0.
- **Tuberculosis**: Correctly classified: 104, Misclassified as normal: 1.

E. Summary of Results

The tuberculosis classification model achieved near-perfect accuracy, indicating excellent performance in distinguishing between normal and tuberculosis cases. The pneumonia model exhibited strong performance, though with a slightly lower recall for normal cases. The lung cancer classification model performed well overall, though it showed some misclassification for adenocarcinoma cases. These results highlight the strengths and potential areas for improvement in medical image-based classification models.

IV. CONCLUSION

This study explored the application of **CNN-based deep learning models** for medical image classification, specifi-

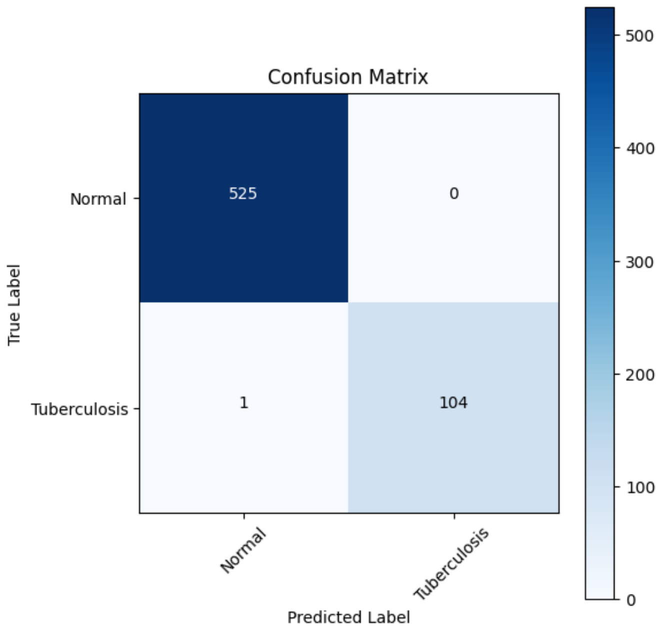

Fig. 4. Confusion matrix illustrating the performance of the CNN-based tuberculosis classification model. The model achieved near-perfect classification accuracy, with only one misclassification observed in the tuberculosis class.

cally targeting **lung cancer, pneumonia, and tuberculosis detection**. By leveraging transfer learning with a pre-trained **DenseNet-121** model, the study demonstrated that deep learning can achieve **high accuracy** in diagnosing medical conditions from X-ray and CT scan images.

The **tuberculosis classification model** achieved an outstanding **99.84%** accuracy, highlighting the potential of CNNs for real-world medical applications. The pneumonia model performed well, though it exhibited **slightly lower recall** for normal cases, indicating potential room for improvement. The lung cancer model successfully classified different cancer types but faced **challenges in distinguishing adenocarcinoma from other forms**. These findings emphasize the effectiveness of AI-powered diagnostics in assisting radiologists and healthcare professionals by reducing interpretation time and improving diagnostic accuracy.

Despite these promising results, several challenges remain. The **variability in real-world medical imaging datasets**, potential model biases, and the need for external validation suggest that further refinement is required before deployment in clinical settings. However, this research reinforces **the transformative potential of AI** in healthcare and sets the stage for **future advancements in automated medical diagnostics.**

V. Future Work

Future research should focus on improving the generalizability of the models by incorporating larger and more diverse datasets. Enhancing the model's robustness through additional data augmentation techniques and domain adaptation methods could reduce misclassification rates. Future studies could also explore the potential of multi-modal learning, where CT scans, X-ray images, and clinical data are combined to enhance diagnostic accuracy. Lastly, deploying these models in real-world clinical settings and evaluating their impact on healthcare workflows would be a crucial step toward practical implementation.

VI. Limitations

Despite the promising results, the models presented in this study have several limitations. First, the datasets used may not fully represent the variability found in real-world clinical settings, leading to potential biases in model predictions. Additionally, while deep learning models excel at pattern recognition, they still struggle with rare cases and subtle abnormalities that require human expertise. The lack of external validation on independent datasets limits the model's generalizability. Computational resource constraints may also hinder real-time deployment, particularly in low-resource healthcare settings.

VII. Ethical Considerations

AI-based diagnostic models must be deployed responsibly to ensure fairness and transparency. Bias in training data can lead to disparities in diagnostic accuracy across different patient groups. Ensuring patient privacy and compliance with regulations is crucial when handling medical data. AI should assist, not replace, human expertise to prevent over-reliance and diagnostic errors. Clear guidelines for AI use and continuous monitoring in clinical settings are necessary to maintain ethical and responsible deployment.

VIII. Acknowledgments

The model was curated by the team consisting of Nathan Wan, Kevin Du, Millicent Song, Juna Kim, Besma Serrai, Liam McQuay, Matthew Louis Li, Artemiy Vishnyakov and Sabrina Lee who helped this project come to fruition.

References

[1] T. Pang, P. Li, and L. Zhao, "A survey on automatic generation of medical imaging reports based on deep learning," *Biomedical engineering online*, vol. 22, no. 1, pp. 48–16, 2023.

[2] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature (London)*, vol. 542, no. 7639, pp. 115–118, 2017.

[3] K. A. P. Razzaghi and P. Bayat, "Learning spatial hierarchies of high-level features in deep neural network," *Journal of visual communication and image representation*, vol. 70, pp. 102 817–, 2020.

[4] J. K. Winkler *et al.*, "Association between surgical skin markings in dermoscopic images and diagnostic performance of a deep learning convolutional neural network for melanoma recognition," *JAMA dermatology (Chicago, Ill.)*, vol. 155, no. 10, pp. 1135–1141, 2019.

[5] F. W. Narongrit and J. V. Rispoli, "Editorial for 'preoperative prediction of axillary lymph node metastasis in breast cancer using cnn based on multiparametric mri'," *Journal of magnetic resonance imaging*, vol. 56, no. 3, pp. 710–711, 2022.

[6] X. Liu *et al.*, "A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis," *The Lancet. Digital health*, vol. 1, no. 6, pp. e271–e297, 2019.

[7] K. Khalili *et al.*, "Convolutional neural networks versus radiologists in characterization of small hypoattenuating hepatic nodules on ct: a critical diagnostic challenge in staging of colorectal carcinoma," *Scientific reports*, vol. 10, no. 1, pp. 15 248–15 248, 2020.

[8] S. M. B. S. R. P. M. K. P. B. S. Varma, S. Paturu and N. V. Krishna, "Sldcnet: Skin lesion detection and classification using full resolution convolutional network-based deep learning cnn with transfer learning," *Expert systems*, vol. 39, no. 9, 2022.

[9] M. B. Nierengarten, "New ai model shows promise for cancer diagnosis," *Cancer*, vol. 131, no. 3, pp. e35 715–n/a, 2025.

[10] R. Divya and R. S. S. Kumari, "Detection of alzheimer's disease from temporal lobe grey matter slices using 3d cnn," *The imaging science journal*, vol. 70, no. 8, pp. 578–587, 2022.

[11] Kabil, "Lungcancer4types-imagedataset multiclass lung cancer image dataset for research and analysis," *Kaggle [online]*, 2024.

[12] D. S. K. et al., "Identifying medical diagnoses and treatable diseases by image-based deep learning," cell," *Cell*, vol. 172, pp. 1122–1131, 2018.

[13] M. A. K. K. R. I. K. F. I. Z. B. M. M. A. A. M. E. H. C. T. Rahman, A. Khandakar, ""reliable tuberculosis detection using chest x-ray with deep learning, segmentation and visualization"." *IEEE Access 2020*, vol. 8, pp. 191 586 – 191 601, 2020.