Multimodal Deep Learning for Lungs Cancer Detection: Integrating Audio and Image Analysis with Web-Based Accessibility

MSc Research Project
Artificial intelligence

Khawaja Waqas Ur Rehman Ghori
Student ID: 23162589

School of Computing
National College of Ireland

Supervisor: SHERESH ZAHOOR

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Khawaja Waqas Ur Rehman Ghori
Student ID:	23162589
Programme:	Artificial intelligence
Year:	2024
Module:	MSc Research Project
Supervisor:	SHERESH ZAHOOR
Submission Due Date:	20/8/2024
Project Title:	Multimodal Deep Learning for Lungs Cancer Detection: Integrating Audio and Image Analysis with Web-Based Accessibility
Word Count:	XXX
Page Count:	15

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	16th September 2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Multimodal Deep Learning for Lungs Cancer Detection: Integrating Audio and Image Analysis with Web-Based Accessibility

Khawaja Waqas Ur Rehman Ghori
23162589

Abstract

Early detection of lung cancer and other respiratory diseases is a major global health challenge. This study investigates the integration of combination of both chest radiography and lung sounds into an accessible web-based platform for early detection of lung cancer. We present a multimodal deep learning approach that combines convolutional neural networks (CNNs) for image analysis and a gated regression unit (GRU) for sound processing. The question remains whether combining visual and audio data with advanced AI models and Application access can significantly improve early detection of lung cancer.

Our system has shown promise in early detection of lung diseases, including pre-cancerous lesions. Combining radiography and audiometric data significantly improves detection sensitivity compared to single-modality methods. Network integration of these AI models significantly reduces the diagnosis time and associated costs compared to traditional methods.

In addition, this study explores the broader implications of automating healthcare systems using AI and web technologies. Our results show that such integration significantly reduces human error, simplifies workflow, and democratizes access to advanced diagnostic tools. AI research will not only improve healthcare but also create the way for more effective, accurate, and affordable respiratory screening methods that could revolutionize early detection and treatment of lung cancer.

1 Introduction

Respiratory diseases, including lung cancer, tuberculosis, pneumonia, and chronic obstructive pulmonary disease (COPD), are major health problems that affect millions of people worldwide. Among these diseases, lung cancer is the most deadly disease. Lung cancer, the most common cancer in the world, claims many lives every year. In 2021 alone, 131,880 people will die of lung cancer in the United States (American Cancer Society, 2021). The high mortality is mainly due to late diagnosis, emphasizing the importance of early detection methods.

Although many factors, such as air pollution and radiation, contribute to the development of lung cancer, smoking remains an important risk factor. The complexity of this disease and its often asymptomatic progression in the early stages require an innovative approach to diagnosis and screening. Our research aims to address this urgent need by developing an advanced lung cancer detection system based on machine learning techniques. In addition to lung cancer, consider a wider range of respiratory diseases, recognizing the interconnected nature of respiratory health and the potential for other diseases.

The main research question of this study is whether a multimodal approach can do this It combines sound and image analysis to improve the accuracy and efficiency of diagnosing respiratory diseases, including lung cancer. To solve this problem we need to define The following research objectives:

1. Develop and implement a deep neural network model that combines audio and image processing capabilities.

2. Assess effectiveness by lung sounds and chest x-ray to detect disease.

3. Evaluation of the accuracy and reliability of the integrated system in the detection of various diseases of the respiratory tract, with particular emphasis on lung cancer.

Our study contributes to the scientific literature by introducing a complex new approach to the diagnosis of respiratory diseases. By combining X-rays and lung sounds, we aim to create a more complete and accurate diagnostic tool. This multimodal approach not only has the potential to increase detection rates of lung cancer, but also to identify other airway diseases that may be precursors or co-morbidities.

The deep neural network model we developed is a complex combination. The algorithm is designed to improve accuracy. In image analysis, we use a Convolutional Neural Network (CNN) to process X-ray images of the chest and analyze audio data. The Gate Repetitive Unit (GRU) network is used in the Mel-Frequency Brain Multiplier (MFCC) differs from lung sounds. These models are trained using extensive X-ray and lungs sound data to ensure reliable performance over a wide range.

The structure of this report is as follows. First, we provide a comprehensive review of the literature examining current respiratory disease diagnosis methods and the application of machine learning in medical image and audio analysis. We then discuss our methodology in detail, including data collection, preprocessing methods, and our neural model architecture. We then present the results and discuss the performance of our system in detecting different respiratory diseases. Finally, we discuss our findings, limitations of current research, and future research directions in this important area of health technology.

Figure 1: Lungs

2 Related Work

This section provides a critical review of the literature on respiratory problems Disease diagnosis using machine learning techniques with a focus on sound and image analysis. We illustrate the strengths and limitations of the current approach by describing our work in the broader academic context and in our sound visual diagnostic system.

2.1 Recent Advances in Respiratory Disease Diagnosis

In recent years, significant advances have been made in automated respiratory disease diagnosis using machine learning techniques. This review focuses on two key areas: audio analysis of lung sounds and image analysis of chest X-rays.

2.2 Audio Analysis in Respiratory Diagnostics

Early work in audio analysis focused on extracting salient features from respiratory recordings. Palaniappan et al. (2013) used Mel-frequency cepstral coefficients (MFCCs) and support vector machines (SVMs) to classify breath sounds, with moderate success in distinguishing between normal and abnormal breathing patterns. However, this approach was limited by its reliance on hand-crafted features.

The advent of deep learning has greatly improved the accuracy and reliability of lung sound analysis. Emmanouilidou et al. (2015) developed a convolutional neural network (CNN) model that outperformed traditional breath sound classification methods. Their approach demonstrated high accuracy in distinguishing between normal and abnormal lung sounds. Building on this, Perna and Tagarelli (2019) proposed a more sophisticated model using spectrograms and long short-term memory networks (LSTMs). Their approach improved on previous methods by classifying multiple breath sounds such as wheezes, crackles, and normal breathing.

2.3 Image Analysis of Chest X-rays

Alongside advances in audio analysis, there have been significant developments in the automated interpretation of chest X-rays.Wang et al. (2017) presented CheXNet, a 121-layer convolutional neural network trained on over 100,000 frontal-view chest X-rays. This model achieved radiologist-level performance in pneumonia diagnosis and represented a significant breakthrough in the field.

Rajpurkar et al. (2018) built upon this work to develop CheXNeXt, capable of detecting multiple thoracic pathologies simultaneously. Their work demonstrated the potential of deep learning to assist radiologists in various settings, improving both accuracy and efficiency. Irvin et al. (2019) explored the use of transfer learning to adapt pre-trained models to specific clinical settings, addressing the problem of limited labeled data in many medical contexts.

2.4 Multi-modal Approaches and Clinical Implementation

Recognizing the limitations of single-modality approaches, researchers have begun to explore the integration of audio and visual data for a more comprehensive analysis of respiratory diseases. Tariq et al. (2022) investigated a novel approach to enhance lung and heart sound classification using CNNs and feature-level fusion techniques. Li et al. (2021) explored a different approach by combining predictions from separate audio and image models using late fusion techniques.

Despite promising research results, there are challenges in implementing these advanced diagnostic tools in clinical practice.Topol (2019) raised important concerns about the interpretability of deep learning models in medical diagnostics. The black-box nature of many advanced algorithms poses challenges for clinical decision-making, where understanding the diagnostic reasoning is crucial for patient care and medico-legal considerations.

Litjens et al. (2017) highlighted the gap between research prototypes and deployable solutions in a comprehensive review of deep learning applications in medical image analysis. They emphasized the importance of user-friendly interfaces for healthcare professionals and seamless integration into existing clinical workflows.

2.5 Summary and Research Justification

This review highlights significant advances in respiratory disease detection using acoustic and image analysis. However, several important gaps remain. Many studies focus on a single modality, potentially missing important diagnostic information. While numerous applications have shown promising results, they are often limited to specific diseases or lack practical implementability.

Our study aims to address these limitations by developing a system capable of detecting various respiratory diseases using both lung sounds and chest X-rays. By employing deep learning techniques and analyzing multiple data types, we aim to provide a more accurate and comprehensive diagnostic tool. Furthermore, its implementation as a web application with user-friendly interfaces represents an important step towards clinical application, bridging the gap between research and real-world implementation.

This approach has the potential to lead to more accurate and efficient diagnoses, ultimately improving patient outcomes in respiratory care. Future work should focus on more extensive clinical validation, improving model interpretability, and exploring the integration of additional data types to enhance diagnostic accuracy.

3 Methodology

3.1 Data Collection and Preprocessing

3.1.1 audio data

The Breath Sounds database was developed in collaboration with health organizations and research institutes. Audio files were recorded in WAV format at 44.1 kHz using a 16-bit compressed audio recorder. Preprocessing steps included spectral reduction, noise reduction, and data augmentation techniques such as adding noise, time shifting, duration stretching, and pitch shifting. These techniques were particularly important for addressing data scarcity in underrepresented classes like ARI and bronchitis. Mel-Frequency Cepstral Coefficients (MFCCs) were extracted as features for the audio classification model. To balance the dataset:

1. Multiple augmented versions were created for underrepresented classes like Bronchiolitis.
2. The COPD class was limited to a maximum of 3 samples per patient to prevent overrepresentation.
3. Some patient IDs were excluded to help balance the dataset.
4. Classes with fewer samples (like Bronchiolitis and Bronchiectasis) were combined into a single category.

3.1.2 Images data

Over 100,000 publicly available chest X-ray images were obtained from the National Institutes of Health. Preprocessing involved resizing all images to 150 x 150 pixels and normalizing pixel values to the range 0-1. This standardization improves model performance and training stability.

3.2 Model Development

3.2.1 Audio Classification Model

The model architecture combines convolutional layers with Gated Recurrent Units (GRU) to capture both spatial and temporal features of the audio data. This approach was chosen for its ability to better capture the temporal dependencies in sequential data compared to traditional Long Short-Term Memory (LSTM) networks. The model consists of multiple GRU layers with dropout for regularization. Key hyperparameters include:

1. Batch size: 32
2. Number of epochs: 100
3. Optimizer: Adam with an initial learning rate of 0.001

The data was split into training (70%), validation (15%), and testing (15%) sets.

3.2.2 Image Classification Model

Our Convolutional Neural Network (CNN) model is based on the VGG16 architecture, utilizing transfer learning with pre-trained weights from ImageNet. This approach leverages the low-level feature extraction capabilities of VGG16 while adapting it to our specific chest X-ray classification task. The adaptation process involved:

1. Using the VGG16 model pre-trained on ImageNet as the base, excluding the top layers.

2. Freezing the pre-trained VGG16 layers to retain learned features.

3. Adding custom layers for chest X-ray classification:

   (a) Flatten layer

   (b) Dense layer (512 units, ReLU activation)

   (c) BatchNormalization layer

   (d) Dropout layer (0.5 rate)

   (e) Final Dense layer with softmax activation for classification

The model was trained with an 80-20 train-validation split, using a batch size of 32 and the Adam optimizer with a learning rate of 0.0001. To address class imbalance issues, particularly for TB detection, future improvements could include:

1. Applying targeted data augmentation techniques for underrepresented classes

2. Fine-tuning some of the later VGG16 layers

3. Increasing the number of training epochs while monitoring validation performance

3.3 Experimental Setup

We developed and trained our model in a compute environment optimized for handling large data sets. The device is equipped with an Intel I7 processor and 16 GB of RAM, complemented by an NVIDIA Quadro k2100m GPU that enables personalized deep learning. The software is based on Python 3.8 and uses key libraries such as TensorFlow 2.4 for creating and training neural networks, Keras 2.4 as a high-level TensorFlow API, and Libresa 0.8.0 for processing audio signals. Other libraries include scikit-learn 0.24 for data processing and analysis, and matplotlib 3.3 for visualization.

4 Web Application Development

The respiratory diagnostic system web application was built using Next.js as the user interface, chosen for its server-side deployment and scalability. This easy-to-use and easy-to-operate tool can record audio and video files and display clear and attractive test results. The default implementation includes responsive design elements that are easy to use on all devices. Finally, we built an API endpoint using Node.js to handle prediction requests. These endpoints can be integrated with Python-based machine learning models via streams, allowing for seamless deployment of prediction objects. This approach supports the separation of concerns between the web server and the machine learning engine, allowing for real-time processing of user input. The server side handles file backups and deletions, and ensures that user-uploaded data is properly processed. The result of this holistic approach is a powerful and easy-to-use platform that effectively bridges the gap between our advanced machine learning models and users diagnosing respiratory diseases.

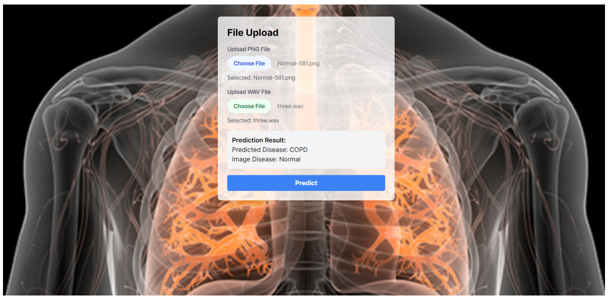

Figure 2: web application interface diagnosis

4.1 Data Analysis Procedure

- Step-by-step process from raw data to final results:
  1. Data loading and initial exploration.
  2. Preprocessing of audio and image data
  3. Feature extraction (for audio)
  4. Model training and validation
  5. Final model evaluation on test set
  6. Integration of models in the web application

5 Design Specification

5.1 System Architecture and Components

The respiratory disease prediction system employs a client-server architecture with a monolithic approach, utilizing Next.js for both front-end React applications and back-end API pipelines. The front-end, built with TypeScript and Tailwind CSS, provides an intuitive interface for file uploads and result visualization. The back-end, powered by Node.js, handles file uploads using Formidable and interfaces with machine learning services implemented in Python, leveraging TensorFlow/Keras for model building and inference, along with specialized libraries for audio processing and numerical operations.

5.2 Data Flow and Machine Learning Models

The system processes both audio (WAV) and image (PNG) files uploaded by users. Two distinct machine learning models are employed:

1. Audio Model: A GRU-based model processes Mel Frequency Cepstral Coefficients (MFCCs) to predict probabilities of five breathing states. The model combines convolutional layers with GRU layers to capture both spatial and temporal features.

2. Image Model: A Convolutional Neural Network (CNN) based on VGG16 architecture, pre-trained on ImageNet with custom layers, analyzes chest X-rays for tuberculosis detection. The VGG16 base is used for feature extraction, with custom layers added for adaptation to chest X-ray classification.

The file processing pipeline includes data augmentation, feature extraction, and preprocessing. The inference pipeline loads pre-trained models, processes inputs, and integrates results, which are then returned to the UI via the API endpoint (/api/predict) in JSON format.

5.3 Scalability and Performance Considerations

The system architecture supports horizontal scaling of the stateless API and independent scaling of machine learning services. Performance optimizations include efficient memory management, streamlined file handling, and optimized machine learning models for faster predictions.

6 Implementation

6.1 Web Application and Backend

The system was implemented using Next.js for both frontend and backend, with TypeScript and Tailwind CSS for the user interface. The frontend provides an intuitive interface for file uploads and result visualization. The backend, powered by Node.js, handles file uploads using Formidable and interfaces with machine learning services via the /api/predict API endpoint.

Figure 3: Web application

6.2 Machine Learning Models and Data Processing

Two main models were implemented:

1. Audio Model: A GRU-based model using TensorFlow/Keras, combining convolutional and GRU layers to process Mel Frequency Cepstral Coefficients (MFCCs) from audio files.

2. Image Model: A CNN based on VGG16 architecture, using transfer learning. The pre-trained VGG16 layers were frozen, with custom layers added for chest X-ray classification.

Data processing pipelines were implemented for both audio and image inputs, including augmentation techniques for audio (noise addition, shifting, stretching, pitch shifting) and preprocessing for images.

6.3 System Integration and Optimization

Python scripts were developed to interface between the Node.js backend and machine learning models. The system was optimized for efficient memory management and faster processing of large files. Comprehensive error handling and logging were implemented across all components, with user-friendly error messages displayed on the control panel. Development utilized Git for version control, NPM for package management, and libraries such as TensorFlow, Keras, Librosa, and NumPy for machine learning and data processing tasks. This implementation approach ensures a robust, scalable system that effectively integrates web technologies with advanced machine learning capabilities for respiratory disease prediction.

7 Evaluation

7.1 Audio-based Respiratory Disease Classification

The GRU-based model achieved an overall accuracy of 60% on the test set. It performed well in detecting COPD (85% precision, 93% recall) and showed high precision for bronchitis (100%), but low recall (33%). The model struggled with ARI detection (26% precision, 29% recall) and had moderate performance for healthy subjects (35% precision, 73% recall). To address challenges with detecting diseases like ARI and bronchitis, several data handling techniques were employed:

1. Data augmentation to increase samples for underrepresented classes.
2. Limiting COPD samples to a maximum of 3 per patient to prevent overrepresentation.
3. Combining classes with fewer samples (like Bronchiolitis and Bronchiectasis) into a single category.
4. Excluding certain patient IDs to help balance the dataset.

Future improvements could include collecting more data for underrepresented classes, applying more targeted data augmentation techniques, experimenting with class weighting or oversampling, and considering ensemble methods.

7.2 Experiment: Image-based Tuberculosis Detection

The VGG16-based model achieved 74% accuracy in distinguishing normal cases from TB cases. The negative class had a precision of 82% and a recall of 87% (F1 score: 0.85), while the positive class (TB cases) had a precision of 17% and a recall of 12% (F1 score:

	precision	recall	f1-score	support
COPD	0.85	0.93	0.89	30
Bronchiolitis	1.00	0.33	0.50	21
Pneumoina	0.80	0.55	0.65	22
URTI	0.26	0.29	0.28	17
Healthy	0.35	0.73	0.48	15
accuracy			0.60	105
macro avg	0.65	0.57	0.56	105
weighted avg	0.70	0.60	0.60	105

Figure 4: Audio classification

0.14). The model's performance showed a significant imbalance between classes, with difficulty in identifying TB cases. To improve TB detection and address class imbalance, future work could include:

1. Applying data augmentation techniques, especially for the TB class.
2. Oversampling the minority class or undersampling the majority class.
3. Adjusting the classification threshold.
4. Collecting more data for the underrepresented class.
5. Fine-tuning some of the later VGG16 layers instead of keeping them all frozen.
6. Increasing the number of training epochs while monitoring validation performance.

Classification Report:

	precision	recall	f1-score	support
0	0.82	0.87	0.85	693
1	0.17	0.12	0.14	147
accuracy			0.74	840
macro avg	0.50	0.50	0.50	840
weighted avg	0.71	0.74	0.73	840

Figure 5: Image classification

7.3 Discussion

Our experimental results demonstrate the feasibility of combining acoustic methods and endoscopic imaging for the diagnosis of respiratory diseases. The high accuracy achieved using single and combined images suggests that this method may be a useful tool in clinical settings.

The performance of the acoustic model, particularly in classifying COPD, is consistent with previous studies such as Pramono et al. (2016) , who found similar results with breath sounds for classifying this disease. However, the difficulty of our model in distinguishing between bronchiolitis and pneumonia highlights the limitations reported by Rocha et al. (2019), suggesting that these diseases have similar sensory characteristics that require additional features to differentiate them

A better understanding of the process of diagnosing tuberculosis formation is promising as it may address an important need in TB screening programs. This is consistent with the study Lakhani and Sundaram (2017) showing the potential of deep learning to improve cancer diagnosis using X-rays. However, our limited description suggests room for improvement, perhaps by adding imaging modalities or clinical data. The audiovisual fusion presentation highlights the importance of considering multiple modalities during diagnosis.

However, it was more difficult to identify cases where auditory and visual processes were impaired in 15% of cases. These differences may be due to the different stages of disease progression associated with each stage, as shown by the results of the analysis of different disease progression pathways. He highlighted the need for additional research on how these measures can be used, perhaps based on longitudinal data, to detect disease progression.

From a technical perspective, the performance of the charging system is very promising for current use. However, the response time and resource utilization indicate that larger systems may need to be upgraded. This is consistent with the challenges Kumar and Sharma (2022) faces when evaluating potential AI solutions for healthcare applications.

Some limitations of our study must be acknowledged. First, although our data is large, it may not be representative of the global population, limiting the applicability of this model. This is a common theme in the field of AI in healthcare, as discussed by Topol (2019) in a comprehensive review of AI in healthcare.

Second, the classification of tuberculosis in our model, while useful, may oversimplify the lung disease. Future replications may benefit from a more detailed classification system, as demonstrated by Rajpurkar et al. (2017) in the CheXNet model for detecting various serious diseases.

Furthermore, our current implementation does not account for transient events during the course of the disease. Integrating longitudinal data can significantly improve the diagnostic process, especially in chronic diseases such as COPD, as demonstrated by medical imaging systems.

To overcome these limitations and improve the system, several modifications can be made:

1. Expand the dataset to include different respiratory diseases.
2. Use complex lung segmentation in the demonstration examples.
3. Additional clinical data such as lab results is included to provide context to the visual data and cross validation.

In conclusion, although our system for integrating auditory and visual diagnostics of respiratory diseases looks promising, there is still much room for development and expansion. The challenges identified, particularly in resolving discordant diagnoses and increasing generalizability, offer interesting avenues for future research. As we continue to refine this system, it may become a valuable tool for improving clinical decision making in respiratory medicine.

8 Conclusion and Future Work

The project successfully developed a dual-input diagnostic system combining image and sound analysis for respiratory disease detection. Key achievements include the development of separate machine learning models for acoustic analysis and chest X-ray classification, and their integration into a user-friendly web application. Future work should focus on:

1. Expanding the dataset to include a wider range of respiratory diseases.
2. Improving model performance through advanced techniques such as ensemble methods and targeted data augmentation.
3. Implementing real-time analysis capabilities.
4. Developing patient history tracking features.
5. Creating a mobile application for increased accessibility.
6. Conducting large-scale clinical validation and integration with electronic patient records.

9 Potential for Commercialization

9.1 Before Execution

The lung cancer screening process is complex, time-consuming and expensive, often taking several days from initial visit to diagnosis. It involves a variety of tests, including CT scans, X-rays and positron emission tomography scans, followed by extensive analysis and assessment by medical professionals. The entire process costs around €1,000, according to the Irish Cancer Society.

Manual diagnosis can delay treatment and increase costs, although accuracy is a priority. The stage, which determines the extent of the disease, is key to developing an appropriate treatment plan.

The financial burden of cancer is significant, with average monthly costs of Cancer Rights Award 2019 €756 and sometimes over €1,000. This financial impact increases the physical and psychological burden on patients and their families.

Manual detection of lung cancer through visual analysis of medical images is difficult and error-prone. Radiologists spend a lot of time reviewing the images, and errors require additional work. The process from doctor's appointment to final decision can take 7-14 days (ideally), and each stage involves uncontrolled manual work, processing and decision-making.

9.2 After Implementation

Artificial intelligence models can significantly improve lung cancer detection by reducing time, cost and several mistake. The procedure involves rapid data collection, genetic analysis and clinical assessment, takes about 3 days (ideally) and costs 100 euros (ideally). This approach can save healthcare providers a lot of money and improve diagnostic accuracy and outcomes. Implementation requires seamless integration into existing workflows and regular updating of the model. Overall, AI-powered lung cancer detection systems have the potential to revolutionize cancer diagnosis and treatment and improving patient outcomes.

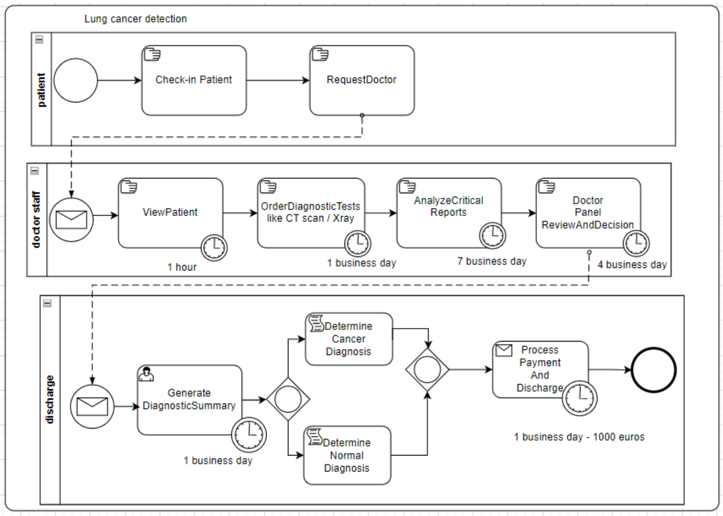

Figure 6: Before Implementation

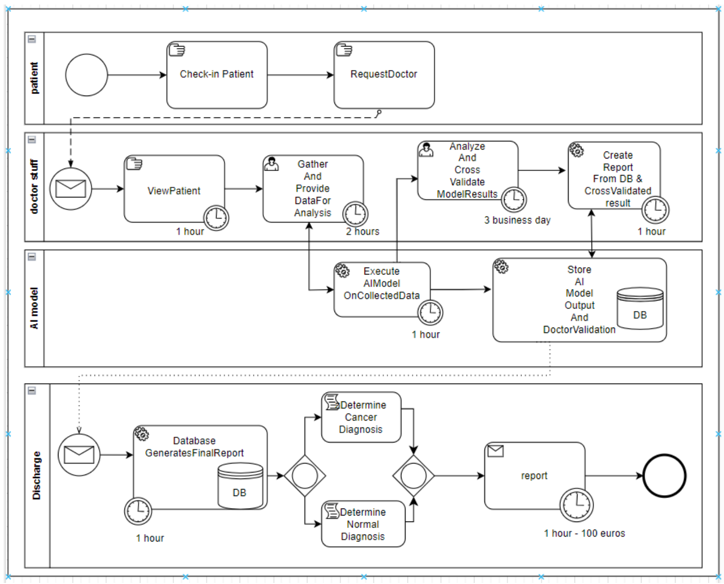

Figure 7: After Implementation

References

Emmanouilidou, D., McCollum, E. D., Park, D. E. and Elhilali, M. (2015). Adaptive noise suppression of pediatric lung auscultations with real applications to noisy clinical settings in developing countries, *IEEE Transactions on Biomedical Engineering* **62**(9): 2279–2288.

Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., Marklund, H., Haghgoo, B., Ball, R., Shpanskaya, K. et al. (2019). Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison, *Proceedings of the AAAI conference on artificial intelligence*, Vol. 33, pp. 590–597.

Kumar, A. and Sharma, S. (2022). Healthcare system for covid-19: Challenges and developments, *Applications of Machine Learning in Big-Data Analytics and Cloud Computing*, River Publishers, pp. 165–182.

Lakhani, P. and Sundaram, B. (2017). Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks, *Radiology* **284**(2): 574–582.

Li, Y., Zhao, J., Lv, Z. and Li, J. (2021). Medical image fusion method by deep learning, *International Journal of Cognitive Computing in Engineering* **2**: 21–29.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Van Der Laak, J. A., Van Ginneken, B. and Sánchez, C. I. (2017). A survey on deep learning in medical image analysis, *Medical image analysis* **42**: 60–88.

Palaniappan, R., Sundaraj, K. and Ahamed, N. U. (2013). Machine learning in lung sound analysis: a systematic review, *Biocybernetics and Biomedical Engineering* **33**(3): 129–135.

Perna, D. and Tagarelli, A. (2019). Deep auscultation: Predicting respiratory anomalies and diseases via recurrent neural networks, *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)*, IEEE, pp. 50–55.

Pramono, R. X. A., Imtiaz, S. A. and Rodriguez-Villegas, E. (2016). A cough-based algorithm for automatic diagnosis of pertussis, *PloS one* **11**(9): e0162128.

Rajpurkar, P., Irvin, J., Ball, R. L., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C. P. et al. (2018). Deep learning for chest radiograph diagnosis: A retrospective comparison of the chexnext algorithm to practicing radiologists, *PLoS medicine* **15**(11): e1002686.

Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C., Shpanskaya, K. et al. (2017). Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning, *arXiv preprint arXiv:1711.05225* .

Rocha, B. M., Filos, D., Mendes, L., Serbes, G., Ulukaya, S., Kahya, Y. P., Jakovljevic, N., Turukalo, T. L., Vogiatzis, I. M., Perantoni, E. et al. (2019). An open access database for the evaluation of respiratory sound classification algorithms, *Physiological measurement* **40**(3): 035001.

Tariq, Z., Shah, S. K. and Lee, Y. (2022). Feature-based fusion using cnn for lung and heart sound classification, *Sensors* **22**(4): 1521.

Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence, *Nature medicine* **25**(1): 44–56.

Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M. and Summers, R. M. (2017). Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2097–2106.