Hazard Analysis
Chest Scan

Team 16, Ace
Harrison Chiu
Hamza Issa
Ahmad Hamadi
Jared Paul
Gurnoor Bal

Table 1: Revision History

Date	**Developer(s)**	**Change**
25 October 2024	Harrison	Section 5, 6
25 October 2024	Hamza	Section 6
25 October 2024	Jared	Section 1, 3
25 October 2024	Gurnoor	Section 2
25 October 2024	Ahmad	Section 7

1 Introduction

This document outlines the potential hazards and corresponding controls for the "AI for Chest X-ray" project, focusing on convolutional neural networks (CNNs) to automatically detect lung and heart conditions. A hazard, in this context, refers to any condition, event, or malfunction that could compromise system accuracy, reliability, security, or patient safety. Conducting a thorough hazard analysis is crucial to mitigate risks, ensure legal compliance (e.g., HIPAA), and improve the system's safety and reliability for clinical decision support, while safeguarding patient data integrity and confidentiality.

In developing machine learning systems for medical diagnostics, hazard analysis is essential to identify potential risks, whether from system failures, unintended consequences, or improper functionality. These risks can lead to harm, mislead users with erroneous outputs, or negatively impact the system's performance. For this project, hazards encompass both software issues and interactions between the AI system and users, such as radiologists and medical institutions. The goal of this hazard analysis is to assess these risks, evaluate their potential impacts, and implement measures to ensure the system functions safely and effectively in a real-world medical environment.

2 Scope and Purpose of Hazard Analysis

The purpose of this hazard analysis is to identify and mitigate risks associated with the "AI for Chest X-ray" system, particularly in clinical environments. Hazards may arise from various issues, including misdiagnosis (false positives/negatives), loss of data integrity, unauthorized access to patient records, and system downtime. The goal is to implement technical, procedural, and security controls that ensure the system operates safely and reliably while supporting radiologists in their diagnoses. This analysis covers the entire AI system's lifecycle—from data acquisition, image analysis, and diagnosis reporting to user interactions and system maintenance.

By focusing on these areas, the hazard analysis aims to safeguard the accuracy of the diagnostic process, protect patient privacy, and prevent errors that could lead to misdiagnoses or delays in care. The potential losses associated with these hazards include not only misdiagnosis and delayed diagnoses but also security breaches and a loss of trust among medical professionals. Given the high stakes involved in medical decision-making, addressing these hazards comprehensively is critical to ensuring patient safety and system reliability.

3 System Boundaries and Components

The system under analysis is designed to facilitate the detection of lung and heart conditions from chest X-ray images using advanced technologies. Its key components include the Machine Learning Model, specifically Convolutional Neural Net-

works (CNNs), which processes the X-rays to identify potential health issues. Complementing this is the Diffusion Model, which generates realistic chest X-ray images that enhance the training of the CNN. The Web Application serves as the interface for radiologists and researchers, allowing them to upload images and review the outputs generated by the AI.

Additionally, the system relies on publicly available datasets like CheXpert and MIMIC for training the AI model, raising concerns about dataset accuracy and bias.

In terms of functionality, the system includes:

Frontend

1. A web interface for user interaction and secure login authentication.
2. Features for image upload, retrieval, and display of diagnostic findings.
3. Access to diagnostic notes and reports restricted to authorized personnel.

Backend

1. An AI model endpoint for detecting anomalies in chest X-rays.
2. Databases that store patient records, X-ray images, and diagnostic reports.
3. Connections to external systems, including hospital IT infrastructure and Radiology Information Systems (PACS).

The system boundary is defined by these components and their interactions, focusing on the direct relationship between healthcare professionals and the AI while excluding external factors like hospital server management and potential network outages that are beyond our control. This ensures that the hazard analysis remains centered on elements that directly impact the system's safety, reliability, and effectiveness.

4 Critical Assumptions

In conducting the hazard analysis for the CNN-based chest X-ray analysis system, several critical assumptions are made regarding the functionality, environment, and usage of the system. These assumptions are necessary to define the scope of potential hazards and their mitigation strategies. While assumptions help simplify the analysis, they also highlight areas where potential risks may need to be revisited as the system evolves.

1. **Assumption 1: Model Performance**
   It is assumed that the convolutional neural network (CNN) model will perform within the expected accuracy levels (90%) as defined by the system requirements. While the system is tuned for high performance, this assump-

tion leaves open the risk of false positives or false negatives in diagnosis, which could lead to misdiagnosis if not managed properly.

2. **Assumption 2: Dataset Quality and Relevance**
   It is assumed that the datasets used for training (CheXpert, MIMIC) are of high quality and representative of real-world chest X-ray images, both in terms of diversity and annotation accuracy. This implies that the data contains no significant biases or labeling errors; however, if these assumptions are incorrect, the model may generate inaccurate or biased results.

3. **Assumption 3: Web Application Stability**
   The web application is assumed to operate reliably under typical internet conditions, with minimal risk of downtime or interruptions. Any disruption in the service (e.g., server outages, high latency) could impact user experience and limit accessibility, especially in time-sensitive clinical situations.

4. **Assumption 4: Data Privacy and Security**
   It is assumed that the system does not collect or store any personally identifiable information (PII) or sensitive health data from users. This assumption is crucial for minimizing privacy risks; however, failure to anonymize or secure data properly could lead to breaches, exposing users to legal and ethical issues related to healthcare privacy.

5. **Assumption 5: User Competency**
   It is assumed that the primary users (radiologists, researchers) possess adequate medical knowledge to interpret the system's outputs. The system does not explain its internal decision-making processes ("black box" issue), which could lead to misinterpretation of results by inexperienced users. If this assumption is invalid, there is a heightened risk of incorrect diagnoses.

6. **Assumption 6: Integration with Hospital IT Infrastructure**
   The system is assumed to integrate seamlessly with secure hospital IT infrastructure, ensuring that patient data (X-rays and records) is stored and accessed securely in compliance with HIPAA regulations. This integration is vital for the effective and safe operation of the AI system in clinical environments.

7. **Assumption 7: Availability of Server Resources**
   It is assumed that the server resources, both frontend and backend, will be continuously available, except during planned maintenance or unexpected disruptions. Any unanticipated downtime could hinder the system's functionality and impact clinical workflows.

5 Failure Mode and Effect Analysis

Table 2: FMEA Worksheet Part 1

Component	Failure Modes	Effects of Failure	Causes of Failure	Detection	Controls	Risk	Recommended Action	Req.	Ref.
Web Application: User Access Authentication	Fail to authenticate user	Medical professional cannot login	Authentication error in web application	Manual testing		Low	Allow users to have alternative methods to login and be authenticated	AR1, SR2	H1.1
		Unauthorized third party attempting to login					Include security safeguards to prevent unauthorized parties from logging in such as passwords, or only making the web application accessible locally.	SR2, SR3, AR0, AR2	H1.2
Web Application: Loading Images from Backend	Fails to load chest x-ray image	Users cannot see chest x-ray images	Failed to encode image or failed to send data to frontend from backend	Manual testing		Low	Include failsafe if backend could not send image to user. And try sending it again if possible.		H2
Web Application: Uploading Images to Backend	Fails to upload chest x-ray image	Backend server which runs the model cannot receive input in order to run disease detection	Failed to encode image or failed to send data to frontend from backend	Manual testing		Low	Add redundant data transfer when sending image data. Ensure network connection to backend is setup correctly		H3
Web Application: Display Findings	Fails to show diagnostic reports of an x-ray image	User cannot read the diagnostic summary report of finding diseases in the images	Failed to send results from backend or failed to parse output from the model into a summary	Manual testing		Low	Include other ways for users to read the reports generated from the model		H4.1
	Shows an incorrect disease report	User reads a diagnostic report unrelated to the image at hand					Double check if the output of the model is correct	SR0	H4.2
Model: Disease detection in the chest x-ray image	False positive detection	Healthy patient could be diagnosed; waste of treatment and healthcare time	Model is too sensitive to certain patterns and shapes, causing it to detect diseases in normal x-ray images	Automated validation testing during model training		High	Optimize the neural network model to minimize false positives	SR0, SR1	H5.1
	False negative detection	Patient does not correctly diagnosed and goes untreated; could worsen its state	Model is too insensitive to certain shapes and patterns, so it fails to detect disease in x-ray images with the diseases				Optimize the neural network model to minimize false negatives	SR0, SR1	H5.2
Model: Training	Model overfitting	Model performs very well on training data with high accuracy but has poor accuracy on validation data (unseen images). Causes inaccurate detection	Overfitting due to high number of training epochs or overly complex model architecture layers	Validation testing after training	Use early stoppage, large training dataset, and regularization	High	Use techniques mentioned in "Controls" column to detect and prevent overfitting	SR0, SR1	H6

Table 3: FMEA Worksheet Part 2

Component	Failure Modes	Effects of Failure	Causes of Failure	Detection	Controls	Risk	Recommended Action	Req.	Ref.
Data Security	Cyberattacks	Unauthorized access database which stores patient health records	Weak security controls. This can include unsecured systems and login data unencrypted	Unauthorized access detection	Good security policies	Risk of leaking patient health records in a data breach	Improve cybersecurity	AR0, AR2, SR3	H7.1
	Unauthorized access	Patient privacy at risk and possible data leak	Weak access control measures		Minimize unnecessary access with with access groups		Improve access control		H7.2
Access data from Database	Data transfer failed	Failed to retrieve patient data	Network communication issues	Automated data transfer checks with network handshakes	Redundant data transfer paths	Data retrieval risk	Add data transfer redundancy when sending data	SR3	H8
Backend Server	Network failure	Connection to database is disrupted	Network connectivity issues	Monitoring server performance	Redundant network connections	Operational disruption	Use network connection redundancy		H9.1
	Server downtime	Unable to access database or run model	Server overloaded with tasks blocking new tasks		Distributed systems		Use distributed systems to ensure server is always on		H9.2
Data Storage	Data loss	Patient images and data is lost	Database accidentally deleted data	Regular data backups	Data redundancy by storing data in multiple places at once	Data loss risk	Regular backup and have a robust data storage system	SR3, AR2	H10.1
	Data corruption	Patient images and data is lost	Database server corruption or data got corrupted during transmission to database	Data integrity checks	Regular data backups	Data loss leak	Check integrity of data in database regularly		H10.2

6 Safety and Security Requirements

6.1 Access Requirements

AR0 The x-ray images used as training data should not be made available publicly; rather, they must be stored on a secure database system to prevent unauthorized access.

Rationale: This is to ensure that private patient data remains secure and is not accessible by unauthorized individuals or bad actors.

Fit Criterion: Only users with appropriate authorization credentials (e.g. researchers or system administrators) should be able to access the training data via the secure database.

AR1 The website should be made publicly accessible to anyone, allowing users to experiment with the associated research findings, including generating synthetic chest X-ray images.

Rationale: This is to allow wide dissemination and experimentation with the research findings, ensuring that researchers and the public can interact with the model without restrictions.

Fit Criterion: Any user, without needing credentials, should be able to access the website and experiment with generating synthetic chest X-ray images.

6.2 Integrity Requirements

AR2 The system will encrypt all chest-x ray training data within the secure database.

Rational: This data should not be available to any website user

Fit criterion: All data is not stored in any discernible language

6.3 Safety Requirements

SR0 The system will indicate the accuracy of the generated chest X-ray image relative to real-world test data.

Rationale: Accurate generation is crucial to ensure the validity of research results and their practical application in the medical field.

Fit Criterion: The system will compare generated images with real test data and display a precision and accuracy score.

SR1 The model will routinely experiment with different training data sets and model parameters to ensure optimal accuracy and precision of generated results.

Rationale: Continuous experimentation and optimization are needed to refine the diffusion model and improve the quality of generated data.

Fit Criterion: Results of the experiments will show improvement in precision and accuracy over time, as tested with various data sets.

SR2 There will be no requirement to log in to the platform, as users are not required to input any personal details. The platform serves purely as an opportunity to experiment with the findings of the research and model.

Rationale: Removing the need for authentication simplifies access and ensures that no personal data is collected, enhancing privacy.

Fit Criterion: Users will have unrestricted access to the platform without the need for authentication.

SR3 Patient's data will be encrypted during data transfers.

Rationale: Encryption is needed to secure and ensure the privacy of patient's health data. It also prevents unauthorized access.

Fit Criterion: The system and database will use a secure encryption algorithm like AES.

7 Roadmap

Given the project's goals and requirements, our focus will be on implementing core access, integrity, and safety requirements. However, due to the research-oriented nature of the project and limited development time, we will prioritize essential functionalities initially, while some advanced features and security enhancements will be deferred for future phases.

To be implemented during the capstone timeline:

- **AR0, AR1:** Secure storage of training data and public website access, scheduled for implementation after the proof-of-concept demo
- **SR0, SR1:** Displaying accuracy indicators for synthetic chest X-ray images in relation to real-world data, scheduled for implementation by January 2025.

To be implemented in the future:

- **Algorithm Optimization:** Continuous refinement of model parameters and training data to improve precision and realism of synthetic X-ray images.
- **Audit Log Maintenance:** Developing a logging system to track access and interactions with the secure database for research tracking purposes.
- **Regular Security Audits:** Periodic reviews and vulnerability assessments to ensure data integrity and model accuracy, preventing unauthorized access or misuse of training data.

- **Encryption Updates:** Continuously update encryption standards and protocols to align with best practices in data protection and cybersecurity, ensuring ongoing compliance and security of stored data.

Appendix — Reflection

The purpose of reflection questions is to give you a chance to assess your own learning and that of your group as a whole, and to find ways to improve in the future. Reflection is an important part of the learning process. Reflection is also an essential component of a successful software development process.

Reflections are most interesting and useful when they're honest, even if the stories they tell are imperfect. You will be marked based on your depth of thought and analysis, and not based on the content of the reflections themselves. Thus, for full marks we encourage you to answer openly and honestly and to avoid simply writing "what you think the evaluator wants to hear."

Please answer the following questions. Some questions can be answered on the team level, but where appropriate, each team member should write their own response:

1. What went well while writing this deliverable?

   One aspect that went well was defining the safety and security requirements for the project. We had already defined non functional requirements in the SRS document, which provided us with a clear idea of how to execute this task and format it. Making the assumptions went smoothly since we were all on the same page and had a solid understanding of what assumptions to make. Our previous work on the project contributed to this clarity, making the process more efficient. Team collaboration went smoothly, as we were able to brainstorm effectively and divide the tasks efficiently.

2. What pain points did you experience during this deliverable, and how did you resolve them?

   One of the main challenges we faced was determining how to limit the system boundaries and what components to include. Given the complexity of our product and its integration with external systems (such as hospital IT infrastructure and datasets), it was difficult to know which elements to focus on within the scope of the hazard analysis. Another complicating factor is that much of our project is research-based, meaning that as we make new discoveries or refine our approach, the system boundaries and components may need to be changed. This uncertainty made it challenging to confidently establish a stable set of boundaries for the analysis. We addressed this by keeping the analysis flexible, acknowledging that components and boundaries may shift over time as we progress in our research. This way, we can revisit and adjust the hazard analysis as necessary to accommodate any new directions the project may take.

3. Which of your listed risks had your team thought of before this deliverable, and which did you think of while doing this deliverable? For the latter ones (ones you thought of while doing the Hazard Analysis), how did they come about?

Risks Identified Before Deliverable

- Failure to Authenticate User (Web Application: User Access Authentication)
- Failure to Load Chest X-ray Images (Web Application: Loading Images from Backend)
- Failure to Display Report of an X-ray Image (Web Application: Display)
- False Positive Detection (Model: Disease Detection in Chest X-ray Image)
- False Negative Detection (Model: Disease Detection in Chest X-ray Image)
- Model Overfitting (Model: Training)
- Cyberattacks (Data Security)
- Data Transfer Failed(Access Data from Database)
- Network Failure (Backend Server)
- Server Downtime (Backend Server)
- Data Loss (Data Storage: Data Loss)
- Data Corruption

Risks Identified During Deliverable

- Failure to Upload Chest X-ray Image (Web Application: Uploading Images to Backend)

  This risk came about during the hazard analysis, when we conducted an assessment of the data flow from the web application to the backend. This assessment was part of our process for the system boundaries and components section which provided a better understanding of our application. This led us to recognize potential failures in image uploads due to connectivity issues or encoding errors. This insight emerged from systematically analyzing how user interactions with the frontend could encounter issues during backend integration.

- Unauthorized access (Data Security)

  We identified this risk during the safety and security requirements section. Specifically, we consider scenarios where data might be breached and accessed by unknown/unauthorized members. Training data incorporates real-life patient data which should be protected, and so the risk of cyberattacks is a crucial consideration.

- Incorrect Display of Disease Report (Web Application: Display of Findings)

As part of the hazard analysis, we looked into how diagnostic findings are presented to users (User interface characteristics). We identified that incorrect parsing or summarization of data could lead to inaccurate outputs on the interface. By examining each layer of interaction between the backend processing and frontend display, we discovered that failures in interpreting the model's output could cause misleading reports.

4. Other than the risk of physical harm (some projects may not have any appreciable risks of this form), list at least 2 other types of risk in software products. Why are they important to consider?

   Security Risks: Unauthorized access or data breaches could expose sensitive patient information. Security risks are critical to consider, especially in healthcare, because they can lead to legal liabilities and the patient's to lose trust in our process.

   Reliability Risks: Software downtime or malfunction, especially in critical environments like healthcare, can delay diagnosis, leading to unsatisfied patients and negative outcomes. Ensuring high system reliability is essential to maintain efficiency and avoid any negative impact on patient care.