Hazard Analysis
CXR

Team 27, Neuralyzers
Ayman Akhras
Nathan Luong
Patrick Zhou
Kelly Deng
Reza Jodeiri

Date	Developer(s)	Change
October 15, 2024	All	Scope and Introduction completed
October 17, 2024	All	Critical Assumptions and FMEA documented
October 20, 2024	All	FMEA tables developed and Safety and Security Requirements updated
October 25, 2024	All	Roadmap and References added

Table 1: Revision History

Contents

1 Introduction

Following the principles outlined by Leveson [2011], a hazard in our AI-based chest X-ray analysis system is any condition or event that could negatively affect the system's operation, integrity, or safety. This includes software bugs, hardware failures, user mistakes, or external disruptions. A hazard becomes significant when it might interfere with the system's ability to provide accurate diagnostic support, protect patient confidentiality, or operate reliably in clinical settings.
This document presents a detailed hazard analysis of our AI-powered chest X-ray analysis application. Throughout this document, we identify potential hazards and suggest ways to eliminate or reduce them. By examining different types of hazards, assessing how likely they are to happen and how severe they might be, and proposing actions to address them, we aim to improve the safety and security requirements for our project.

2 Scope and Purpose

The purpose of this analysis is to identify potential hazards in our AI-based chest X-ray analysis system and to propose ways to mitigate them. We focus on specific system components and boundaries to enhance the safety and reliability of the system. While we recognize that certain external factors, like differences in input image quality or user environments, are beyond our control, our system is designed to handle standard digital chest X-ray images provided by qualified healthcare professionals.
We assume that all system functions—especially those related to image processing, AI analysis, and reporting results—are working as intended. With this in mind, we concentrate on strengthening key components, such as the user interface, backend servers, machine learning models, and data storage systems, to prevent potential failures. Our goal is to ensure that the system provides an accurate, secure, and reliable experience that supports healthcare professionals in diagnosing and managing chest diseases.

3 System Boundaries

3.1 System Components

The system consists of the following main components:

- User Interface (Web Application)
- Backend Servers and APIs
- Machine Learning Model and Inference Engine
- Data Storage and Management Systems
- Security and Authentication Modules

These components are essential to our system's functionality. They handle user interactions, data processing, AI-based analysis, and secure data management.

3.2 Environment Components

The system interacts with the following external components:

- External Medical Imaging Systems (e.g., Picture Archiving and Communication Systems)
- Standard Digital Chest X-ray Images
- Network Infrastructure (Internet Connectivity)

These environment components influence how our system operates. While they are outside our system, they play a critical role in ensuring a smooth and effective user experience in clinical settings.

4 Critical Assumptions

To keep the hazard assessment focused and effective, we make the following critical assumptions:

1. We expect users to upload only legitimate and appropriate medical images. However, we recognize that irrelevant or corrupted data may sometimes be uploaded, and measures should be in place to detect and handle such inputs.
2. We assume that users are qualified healthcare professionals with the necessary training to interpret the system's outputs correctly. The system is primarily intended for professional use in clinical settings.
3. Our system is designed to handle unintentional user errors and common misuse scenarios, but it may not be fully protected against deliberate malicious activities aimed at exploiting vulnerabilities or deceiving the AI algorithms.
4. Our hazard analysis for external environment components, like network infrastructure and third-party systems, considers typical use cases and does not cover extreme conditions outside the intended operation of the system.

5 Failure Modes and Effects Analysis

The Failure Modes and Effects Analysis (FMEA) was selected as the hazard analysis tool to help identify, assess, and propose solutions to the risks and hazards associated with our AI-based chest X-ray analysis system.

5.1 Hazards Out of Scope

- Failures related to external AI libraries (e.g., TorchXRayVision)
- Issues due to external data sources (e.g., incorrect image formats or corrupted data files)
- Failures of the external hardware used by healthcare professionals

Our project is not responsible for the hazards listed above, as they are controlled by third-party systems or external users. While we will take steps to minimize the impact of these hazards, complete mitigation cannot be guaranteed.

Component	Failure Mode	Effect	Cause	Recommended Action	SR	Ref
Input System	Failure to upload chest X-ray images to the system.	Delays in diagnosis and analysis, affecting clinical workflow.	a. Incompatible file formats or corrupted images cause upload failures. b. Insufficient server storage space leads to failed uploads.	a. Validate image files during upload by checking formats and integrity, and provide immediate error messages to guide users in resolving issues. b. Monitor server storage capacity regularly, and employ scalable storage solutions like AWS S3 to ensure adequate space is available.	none	H1-1
	System fails to process or analyze uploaded images due to poor image quality.	Inaccurate AI analysis leading to misdiagnosis.	a. System lacks mechanisms to detect and handle poor-quality images. b. Insufficient preprocessing capabilities to enhance or correct image quality issues. c. Absence of feedback to users when images are unsuitable for analysis. d. Limitations in the AI model to handle variations in image quality.	a. Implement image quality assessment during upload, providing warnings or errors if images do not meet quality thresholds. b. Enhance preprocessing algorithms to improve image quality where possible, such as noise reduction. c. Provide users with guidelines on acceptable image quality and instructions for obtaining better images if necessary. d. Train the AI model on a diverse dataset that includes variations in image quality to improve robustness.	SR2 HS2	H1-2

Table 2: FMEA for Image Input Component (H1-1 and H1-2)

Component	Failure Mode	Effect	Cause	Recommended Action	SR	Ref
User Interface	Interface is non-intuitive or difficult to navigate for healthcare professionals.	Decreased efficiency and user frustration	a. Inconsistent use of medical terminology and symbols leads to confusion. b. Lack of user training or insufficient documentation. c. Interface not optimized for different devices or screen resolutions.	a. Standardize terminology and symbols according to medical standards, and ensure consistency throughout the interface. b. Provide comprehensive training materials, including user manuals and tutorials, and offer ongoing support. c. Test the interface across various devices and screen sizes, optimizing responsive design to ensure accessibility.	HS3	H2-1
	Displays incorrect or misleading analysis results.	Risk of misdiagnosis due to wrong information.	a. Bugs in UI logic lead to incorrect data display. b. Data mismatch between frontend and backend systems causes inconsistencies. c. Network issues result in incomplete data retrieval, leading to partial displays.	a. Fix UI logic errors by conducting code reviews focusing on data binding and state management, and implementing unit tests for UI components. b. Ensure data synchronization by using consistent data formats, implementing version checks, and validating data integrity between frontend and backend. c. Implement reliable data retrieval methods using robust APIs with error handling and providing user feedback during data loading.	HS3 HS4	H2-2

Table 3: FMEA for User Interface Component (H2-1 and H2-2)

Component	Failure Mode	Effect	Cause	Recommended Action	SR	Ref
Data Preprocessing	Incorrect preprocessing of chest X-ray images before AI analysis.	AI model receives improperly formatted data, leading to reduced accuracy or errors.	a. Misalignment in image resizing results in images not scaled to required dimensions. b. Incorrect normalization or standardization of pixel values distorts image data.	a. Implement strict preprocessing protocols and validate image dimensions using automated checks in the data pipeline. b. Utilize standardized libraries (e.g., PyTorch transforms) for image normalization, ensuring consistent preprocessing steps.	SR2	H3-1
	Data augmentation introduces artifacts that mislead the AI model.	AI model learns from distorted data, leading to poor generalization and increased errors.	a. Introduction of noise that do not represent real-world variations. b. Mislabeling augmented data due to incorrect augmentation metadata handling. c. Lack of validation on the quality and clinical relevance of augmented datasets.	a. Monitor the quality of augmented images by reviewing samples and ensuring they maintain anatomical correctness. b. Verify and maintain accurate labels and metadata post-augmentation by automating checks and utilizing data integrity tools. c. Incorporate validation steps to assess the impact of augmentation on model performance, adjusting techniques based on results.	none	H3-2

Table 4: FMEA for Data Preprocessing Component (H3-1 and H3-2)

Component	Failure Mode	Effect	Cause	Recommended Action	SR	Ref
AI Module	AI model misclassifies chest X-ray images, providing incorrect diagnoses.	Potential misdiagnosis leading to inappropriate treatment plans and patient harm.	a. Training data includes poor-quality images or lacks diversity, leading to biased model performance. b. Software bugs in the CNN architecture implementation or data preprocessing pipeline.	a. Levrage a high-quality, diverse training dataset, ensuring representation across demographics and conditions. b. Conduct thorough code reviews, unit tests, and integration tests on the AI model and preprocessing code.	HS1 HS5 HS6	H4-1
	Model fails to improve over time despite updates.	Inability to detect new or rare conditions, reducing clinical effectiveness.	a. Insufficient incorporation of new training data or lack of ongoing data collection. b. Failure to integrate feedback from radiologists and healthcare professionals.	a. Establish a continuous data collection pipeline. b. Create a feedback loop with clinicians to gather real-world performance data and insights.	HS2	H4-2
Backend Services	Experiences slow response times or timeouts during image upload and analysis requests.	Users face delays, reducing efficiency in clinical workflow .	a. High server load due to multiple concurrent image processing requests overwhelms resources. b. Insufficient server resources (CPU, memory) lead to lower performance.	a. Implement load balancing mechanisms to distribute requests, and optimize server configurations for concurrency. b. Scale server resources dynamically using AWS Auto Scaling groups, and monitor resource utilization to adjust as needed.	SR6 HS1	H5-1
	Experiences server downtime or crashes.	System is unavailable, affecting patient care.	a. The server cannot handle peak loads due to lack of scalability. b. Unplanned maintenance or deployment errors result in service interruptions.	a. Adjust capacity based on demand, ensuring sufficient server instances are running during peak times. b. Schedule maintenance during off-peak hours with advance notifications to users, and test updates in staging environments before production.	SR6	H5-2

Table 5: FMEA for AI Module Component (H4-1-H5-1)

Component	Failure Mode	Effect	Cause	Recommended Action	SR	Ref
Data Storage	Experiences data loss or corruption of chest X-ray images and patient records.	Loss of critical medical data, disrupting patient care and violating data retention policies.	a. Software bugs cause data corruption during processing or storage operations. b. Accidental deletion of data by users or administrators due to inadequate safeguards. c. Inadequate backup procedures or failure to test data recovery processes.	a. Implement data validation checks during processing and use database transactions to maintain data integrity. b. Enforce strict permissions, and provide training to prevent accidental deletions. c. Establish regular automated backups, and periodically test data recovery procedures to ensure they function correctly.	SR3 HR4 HR5	H6-1
Security Modules	Suffers security issues leading to unauthorized access to personal information.	Exposure of patient data, violating HIPAA and PIPEDA regulations, leading to legal consequences and loss of trust.	a. Weak authentication mechanisms allow unauthorized access. b. Insufficient data encryption at rest and in transit exposes sensitive information. c. Inadequate access controls and monitoring fail to detect or prevent unauthorized activities.	a. Strengthen authentication using multi-factor authentication and enforce strong password policies. b. Encrypt data at rest using AWS KMS and enforce SSL/TLS protocols for data in transit to secure communications. c. Implement role-based access control (RBAC), and set up alerts for suspicious activities.	SR1 SR4 SR5 HS6	H7-1

Table 6: FMEA for Data Storage Component (H6-1 and H7-1)

6 Safety and Security Requirements

Using the results of the FMEA, we can derive the following safety and security requirements for our system to mitigate the identified hazards.

6.1 Security and Privacy Requirements

SR1:The system shall anonymize all chest X-ray images and associated patient data before processing or storage, ensuring that no personal information is retained. *Rationale*: Anonymization protects patient privacy and complies with healthcare privacy regulations, preventing unauthorized access to sensitive information.

Fit Criterion: Security audits confirm that all stored and processed data is anonymized; attempts to retrieve personal information from the system yield no results.

Traceability: H7-1

SR2: The system shall validate the quality of images upon upload, ensuring that only diagnostically valid images enter the system. *Rationale*: Ensuring image quality prevents inaccurate diagnoses and maintains the system's reliability and effectiveness.

Fit Criterion: Automated quality checks must validate image resolution, contrast, and clarity before processing.

Traceability: H1-2, H3-1

SR3:The system shall define a data preservation policy that specifies how long patient data is stored and ensures secure deletion procedures after the retention period expires.

Rationale: Limiting data retention minimizes the risk of data leaks and complies with legal requirements for data protection.

Fit Criterion: Policy documents outline data retention periods; system logs verify that data is securely deleted after expiration.

Traceability: H6-1

SR4:The system shall implement a consent management system that allows patients to control the use and sharing of their data, in accordance with relevant privacy regulations.

Rationale: Obtaining and managing patient consent ensures ethical use of data and compliance with laws like HIPAA of Health and Services [2024] and GDPR Commission [2016].

Fit Criterion: Records show that patient consent is given; audits confirm that data sharing aligns with patient preferences.

Traceability: H7-1

SR5:The system shall create detailed audit trails that track all access to and modifications of patient data, enabling detection and investigation of unauthorized activities.

Rationale: Audit trails enhance security by providing accountability and facilitating incident response in case of security issues.

Fit Criterion: Audit logs are comprehensive; security reviews confirm that all access and changes are properly recorded.

Traceability: H7-1

SR6:The system shall implement security measures for API endpoints and service communications to protect against security vulnerabilities

Rationale: Secure APIs and communications prevent unauthorized access and data breaches, ensuring the system's integrity and confidentiality.

Fit Criterion: Implementation of security best practices for API endpoints; secure service communications are verified.

Traceability: H5-1, H5-2

6.2 Health and Safety Requirements

HS1:The system shall ensure that all AI-generated diagnoses are reviewed and confirmed by qualified medical professionals before being used in patient care decisions.

Rationale: To prevent misdiagnoses and ensure patient safety by involving human oversight in the diagnostic process.

Fit Criterion: System workflow requires medical professional validation before finalizing reports; logs confirm this process.

Traceability: H4-1, H5-1

HS2:The system shall provide clear warnings and notifications if the input data quality is insufficient for reliable analysis, prompting users to provide better data.

Rationale: Processing poor-quality images can lead to inaccurate diagnoses, potentially harming patients; users should be aware of data limitations.

Fit Criterion: The system detects low-quality inputs and displays warnings; user acknowledgment is required before proceeding.

Traceability: H1-2, H4-2

HS3:The system shall manage user fatigue to reduce errors from extended use.

Rationale: Fatigue management improves focus, reducing mistakes during long-term system use by helping users maintain optimal physical and cognitive conditions.

Fit Criterion: The system provides break reminders, eye-strain reduction settings, and session timeouts for inactivity.

Traceability: H2-1, H2-2

HS4:The system shall maintain essential medical information by displaying relevant patient history alongside current diagnostic outcomes.

Rationale: Ensuring diagnostic background supports accurate decision-making by providing essential background for interpreting results.

Fit Criterion: Relevant patient history and prior analyses are displayed with current results, and missing data is clearly indicated.

Traceability: H2-2, H6-1

HS5:The system shall include disclaimers indicating that AI analysis is a diagnostic aid and not a replacement for professional medical use.

Rationale:Users must understand the limitations of AI to prevent overreliance and potential errors in patient care.

Fit Criterion: Disclaimers are prominently displayed on analysis results; users acknowledge understanding upon first use.

Traceability: H4-1, H6-1

HS6:The system shall monitor patient safety by conducting ongoing assessments and collecting user feedback.

Rationale: Continuous monitoring ensures patient safety by identifying potential risks.

Fit Criterion: The system undergoes regular audits, collects user feedback on safety, and supports incident reporting mechanisms.

Traceability: H4-1, H7-1

7 Roadmap

In the hazard analysis and FMEA documentation for our chest X-ray diagnostic system, we have identified and prioritized a comprehensive set of safety and security requirements to mitigate potential hazards. These requirements aim to ensure the accuracy, reliability, and integrity of the system, while addressing essential privacy, safety, and usability concerns in a clinical environment.

Requirements focusing on fundamental safety and security priorities, such as ensuring the correctness of AI-generated diagnoses, preserving user privacy, and minimizing patient risk, are given the highest priority. These include SR1, SR2, SR5, HS1, HS2, and HS5. Implementing these will ensure that the core functionalities, such as diagnosis review by medical professionals and anonymization of sensitive data, are in place by the end of the capstone project to reduce the likelihood of critical errors and privacy breaches.

High-priority requirements focused on operational continuity, such as system availability during peak loads, responsive user interfaces, and the management of backend downtime, will also be addressed within the project timeline. These include HS3, SR6, SR4, and HS6, which are essential for sustaining a seamless clinical workflow and ensuring minimal disruptions during diagnosis.

Some requirements aimed at further enhancing the system's long-term usability and performance will be postponed for future implementation. These include advanced quality monitoring of uploaded images (SR2), integration of patient feedback mechanisms (HS4), and additional fatigue management improvements for clinicians (HS3). While these features are beneficial for improving efficiency and user satisfaction, they are considered medium priority and will be explored after the initial system is operational.

Given the scope and time constraints of the capstone project, the roadmap prioritizes essential safety, security, and operational requirements to be implemented during the course. The remaining non-critical requirements will be reserved for future iterations. This roadmap will ensure that our chest X-ray diagnostic system evolves towards meeting high standards of safety, usability, and compliance, aligning with clinical needs and regulatory requirements.

References

European Commission. Regulation (eu) 2016/679 on the protection of natural persons with regard to the processing of personal data (general data protection regulation). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679, 2016.

Nancy G. Leveson. *Engineering a Safer World: Systems Thinking Applied to Safety*. MIT Press, Cambridge, MA, 2011.

U.S. Department of Health and Human Services. Summary of the hipaa security rule. https://www.hhs.gov/hipaa/for-professionals/security/laws-regulations/index.html, 2024.

Appendix — Reflection

1. **What went well while writing this deliverable?**

   Writing this hazard analysis allowed us to think more critically and in specific layers about our AI-based chest X-ray analysis system. By breaking down the project into major components like the user interface, backend servers, machine learning models, and data storage, we were able to systematically identify and analyze potential hazards associated with each part. This detailed examination enabled us to anticipate risks ahead of the implementation phase and develop effective mitigation strategies. The process not only deepened our understanding of the system's nuances but also ensured that we considered a wide range of failure modes and their effects through a comprehensive FMEA analysis. Additionally, aligning our safety and security requirements with industry standards and regulations such as HIPAA and GDPR helped strengthen the overall quality and compliance of the deliverable.

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

   One of the main challenges was ensuring that we captured all relevant hazards, especially those that are not immediately obvious, such as ethical biases in AI models or regulatory compliance issues. Initially, our focus was mostly on technical failures and user interface problems. To overcome this, we broadened our perspective by consulting additional resources and engaging with our supervisor Dr. Mehdi Mordadi who is an experts in the field of medical imaging. This helped us identify and incorporate hazards related to data privacy, AI bias, and user training into our analysis. Another pain point was managing the complexity of interrelated system components. We resolved this by clearly defining system boundaries and assumptions, which streamlined our analysis process.

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?**

   Before this deliverable, we had already considered risks such as incorrect AI diagnoses due to model inaccuracies (H4-1), data loss or corruption (H6-1), and security breaches leading to unauthorized access of patient data (H7-1). These risks were apparent due to the critical nature of healthcare data and diagnostics.

   While conducting the hazard analysis, we identified additional risks, including:

   - *AI model failing to improve over time* (H4-2): We identified that without continuous learning and updates, the AI model might become outdated, especially as new medical findings are discovered.
   - *Failure to handle poor-quality images* (H1-2): We realized that image quality significantly impacts AI analysis accuracy. This hazard emerged when examining the image input component and considering real-world scenarios where image quality might vary.
   - *User interface challenges leading to misinterpretation of results* (H2-2): During the FMEA, we recognized that even if the AI model is accurate, a confusing UI could mislead users, causing misdiagnosis.

These hazards came to light as we methodically assessed each component's failure modes and their potential effects, encouraging us to think beyond initial assumptions.

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? software products. Why are they important to consider?**

   Neuralanalyzer is a complex system with many components. However our team has considered the following two as ones of the **highest risks**:

   (a) Unreliable prediction models lead to misdiagnoses: This risk was identified as the highest risk since the product primary selling point is the accuracy of the AI models. There are many reasons for this risk to occur:

   - *Model's ability to recognize chest-embedded medical device*: Devices such as chest tubes, pacemakers, or defibrillators can be misinterpreted as medical conditions by the model.
   - *Biased training data*: When the chest X-Ray data is only from a specific demographic, scanning machine, or hospital, the model will tend to over-fit to that specific data bias. This will lead to unreliable results on other demographics, machines, or hospitals.
   - *Insufficient training data*: When the model is trained on a small dataset, it will not be able to generalize well on new data, leading to unreliable results.
   - *Developer's operation errors*: When developers make mistakes with the model's architecture, training, or deployment, it will affect the downtime of the AI service.

   Given the severity of this risk, our team has came up with a strategy to minimize this risk on production:

   - *Gather big and diverse Chest X-ray images*, from multiple sources: Under the supervision of Dr. Mehdi, we will be training our model with Imagenome and ChexPert datasets, which are well-known and massive datasets.
   - *Transfer Learning and tuning pre-trained models*: As mentioned within SRS, we will utilize the TorchXrayVision model as our back-up mechanism. Since this model is pre-trained on ChexPert, our team will be performing additional training with Imagenome on top of TorchXrayVision.
   - *Automate Developers workflow with CI/CD and automation scripts*: Developers operation such as data handling, version control, model deployment and model retraining should be automated to reduce human errors.

   (b) System's SLA is not met due to unexpected downtime: Hospitals when using Neuralanalyzer system will under the assumption that our system will be up, as agreed in the SLA, to serve patients diagnosis. However, there are multiple reasons for the system to go down unexpectedly:

   - *Unreliable CI/CD pipeline*: When the CI/CD pipeline is not developed and tested thoroughly, buggy code can be deployed to production, leading to unexpected crashes and downtime.

- *Unmonitored logs and resources*: Some operational issues might run through CI/CD undetected until very later on, such as spikes in traffic, high CPU usage, or memory leaks. These issues will lead to performance degradation and eventually system crashes.
- *Deployment of major versions*: Minor versions changes are not considered as risky as major versions changes. Major versions changes usually introduce breaking changes that will affect production reliability if not tested properly. When major versions are introduce, the team must ensure egregious testing and provides backwards-compatibility analysis.

The risk of unexpected downtime can be handled in many ways, with many tools available on the markets, such as AWS CloudWatch, Datadog, PagerDuty, etc. In the context of Neuralanalyzer's Capstone implementation, we will negate downtime via the following methods:

- *Implementing log monitoring*: automatically notify developers when a certain level of errors had occurred via email/SMS. Developers will then come online to fix the problem.
- *Implementing a manual rollback mechanism*: When a bug takes too long to trace and debug, developers should be able to rollback their changes to the previous stable working state.