TED UNIVERSITY

Faculty of Engineering

Department of Software Engineering

Department of Computer Engineering

CMPE 492 / SENG 492

XAI Healthcare Bot

Detailed Design Report

by

Doğukan Yetgin
Hilal Yurtoğlu
Ruşen Deniz Kaplan
Zeynep Bölükbaşı

Table of Contents

1. Introduction

1.1 Purpose of the system

The XAI Healthcare Bot is intended to help patients and healthcare professionals by providing explainable, AI-powered diagnostic support. This system ensures openness and trust by offering interpretable insights into its decision-making process, in contrast to conventional black-box AI systems. The system's main goals are as follows:

Supporting Healthcare Providers: Giving physicians thorough justifications for AI-generated diagnoses and assisting them in more efficient medical data analysis.

Improving Patient Understanding: Giving patients clear explanations of diagnosis can assist them comprehend their medical issues.

Improving Diagnostic Accuracy: Making use of explainable AI models to guarantee that medical forecasts are both accurate and comprehensible.

Ensuring Compliance: Upholding data security and privacy by following healthcare laws like HIPAA and GDPR.

1.2 Design goals

Performance:

The system should be able to respond to user queries in real time and manage interactions with multiple users simultaneously without delay.

Accuracy:

AI recommendations and explanations must have a high accuracy rate in line with the latest medical guidelines.

Reliability:

During times of heavy use, the system must operate without interruption and must be able to quickly return to its previous state.

Usability:

The user interface should be intuitive, enabling seamless navigation with clear instructions for all user profiles.

Security:

Patient information must be protected with strong encryption and secure access controls and comply with regulations such as HIPAA and GDPR.

Scalability:

The system must support an increasing number of users and features without compromising performance.

Maintainability:
The codebase should be modular and well documented, with regular updates that can be made with minimal user interaction.

Interoperability:
Must be compatible with existing healthcare systems and electronic health records (EHR).

Accessibility:
The system should adhere to accessibility standards, ensuring that users with disabilities can effectively interact with the bot. This includes support for screen readers and adaptable interface options.

Cost-Effectiveness:
The system must provide high value while keeping both development and operating costs low.

Adaptability:
It should be able to adapt to different environmental conditions and be updated according to changing health standards.

Responsiveness:
It should be able to adapt to different environmental conditions and be updated according to changing health standards.

Aesthetics:

Designing a clean and professional user interface that enhances user experience.

1.3 Definitions, acronyms, and abbreviations

- AI (Artificial Intelligence): A technology enabling machines to perform tasks that typically require human intelligence.
- XAI (Explainable Artificial Intelligence): AI systems designed to provide interpretable and transparent decision-making.
- HIPAA (Health Insurance Portability and Accountability Act): A US law regulating healthcare data privacy.
- GDPR (General Data Protection Regulation): A European Union regulation governing data protection and privacy.
- SHAP (Shapley Additive Explanations): A method for interpreting machine learning models.
- Grad-CAM (Gradient-weighted Class Activation Mapping): A technique for visualizing deep learning model decision-making.
- DICOM (Digital Imaging and Communications in Medicine): A standard for storing and transmitting medical images.

1.4 Overview

The XAI Healthcare Bot's architecture, security protocols, and technological limitations are described in depth in this study. In order to guarantee a successful implementation, the next sections will go over the system's design, limitations, security standards, and teamwork techniques.

2. Proposed software architecture

2.1 Overview

The XAI Healthcare Bot project is designed to revolutionize diagnostic processes by integrating Explainable Artificial Intelligence (XAI) techniques within the healthcare domain. The system’s core objective is to assist medical professionals and patients by not only delivering highly accurate diagnoses but also by providing clear, interpretable explanations of its reasoning. Leveraging advanced algorithms such as YOLO for real-time image processing, along with SHAP and Grad-CAM for interpretability the system processes complex medical images (e.g., brain scans for Alzheimer’s detection) and associated clinical data. The solution is built upon a modular and scalable architecture that integrates robust data management, hardware acceleration via cloud-based GPU servers, and a user-friendly interface accessible on multiple devices. This combination of technologies ensures that diagnostic outcomes are both reliable and transparent, fostering trust and facilitating improved healthcare outcomes.

2.2 Subsystem decomposition

The overall architecture is decomposed into several interconnected subsystems, each responsible for a specific set of functions:

- **Data Preprocessing and Feature Extraction:**
  This subsystem handles the initial intake of raw medical data and images. It performs tasks such as data cleaning, normalization, and feature extraction using advanced algorithms (e.g., YOLO) to ensure that the data is in a suitable format for further analysis.

- **AI-Based Diagnosis Subsystem:**
  At the heart of the system, this component leverages deep learning models combined with XAI techniques (SHAP, Grad-CAM) to diagnose diseases. It outputs both a diagnosis and detailed explanations of the contributing factors, supporting clinical decision-making.

- **Notification and Alert Subsystem:**
  Designed for real-time communication, this module sends alerts, reminders, and notifications to both patients and healthcare providers regarding diagnostic updates

and medication schedules.

- **Profile Management Subsystem:**
  This part manages user profiles separating patients from healthcare professionals and ensures that data is customized and securely accessible based on user roles.

- **Persistent Data Management Subsystem:**
  Critical for storing large volumes of data, this subsystem utilizes a hybrid database approach (combining relational and NoSQL databases) to maintain both structured records (e.g., patient histories) and unstructured data (e.g., high-resolution images and AI outputs).

- **Security and Compliance Subsystem:**
  Enforcing regulatory compliance (such as HIPAA and GDPR), this module implements robust access control, encryption, and auditing mechanisms to protect sensitive medical information.

- **User Interface Subsystem:**
  A responsive, web-based interface designed for ease of use, ensuring that users with varying levels of technical expertise can navigate the system and interpret AI outputs effortlessly.

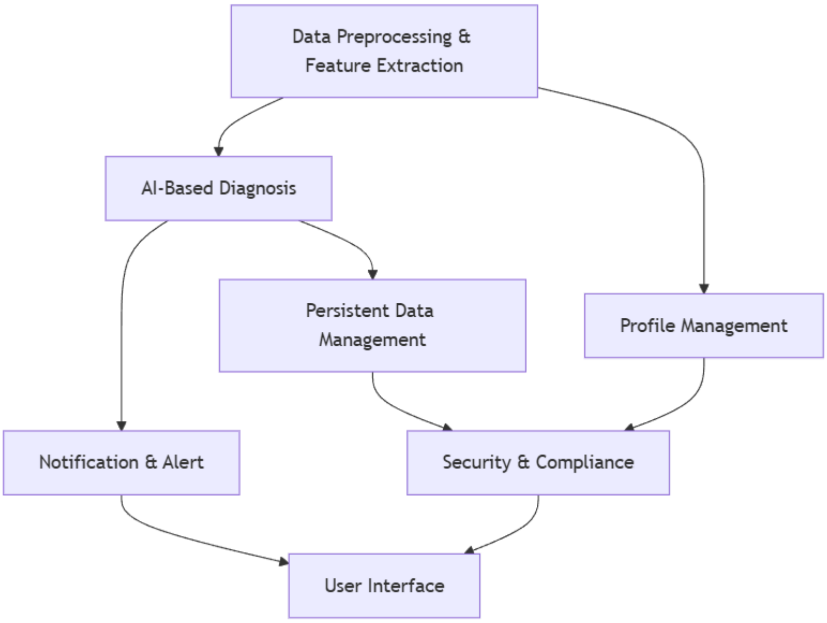

Figure 1. Subsystem Decomposition Diagram

2.3 Hardware/software mapping

The efficient mapping between hardware and software components is essential for achieving high performance and scalability:

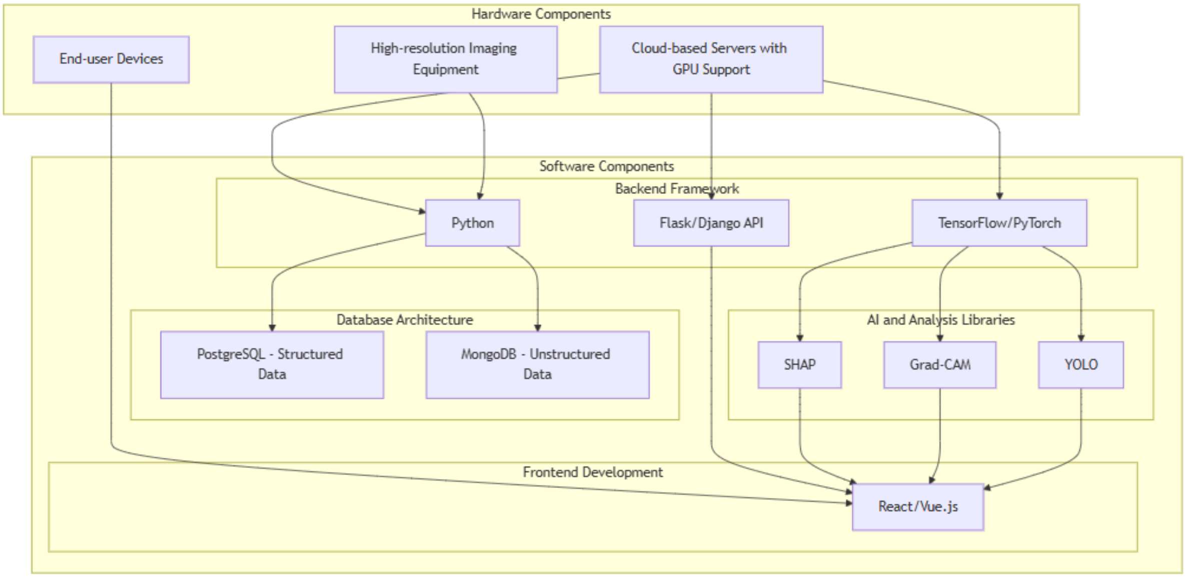

Figure 2. Hardware/Software Mapping Diagram

- **Hardware Components:**
  - **Cloud-based Servers with GPU Support:** These servers provide the computational power required for training and running deep learning models. GPUs accelerate the processing of intensive AI tasks.
  - **End-user Devices:** Patients and healthcare professionals interact with the system through a variety of devices such as smartphones, tablets, and PCs. These devices must be compatible with modern web browsers.
  - **High-resolution Imaging Equipment:** Specialized imaging tools capture high-quality medical images (e.g., MRI, CT scans) that are critical for accurate diagnosis.

- **Software Components:**
  - **Backend Framework:** Implemented in Python, the backend uses deep learning libraries (TensorFlow, PyTorch) for model development and frameworks like Flask or Django to manage API endpoints.

  - **Frontend Development:**
    Modern web technologies (React or Vue.js) deliver a responsive, interactive interface for users.
  - **Database Architecture:**
    A hybrid database solution combining PostgreSQL (for structured data such as patient records) and MongoDB (for unstructured data such as medical images and AI analysis outputs).
  - **AI and Analysis Libraries:**
    Tools like SHAP and Grad-CAM provide transparency by visualizing the factors that influence AI decisions, while YOLO ensures rapid image processing.

2.4 Persistent data management

Given the high volume and diverse nature of healthcare data, the system utilizes a robust, hybrid persistent data management strategy:

- **Data Storage Strategy:**
  The system divides data into two primary categories:

  - **Structured Data:**
    Patient records, diagnostic reports, medication histories, and treatment outcomes are stored in PostgreSQL. This ensures relational integrity and facilitates complex queries.
  - **Unstructured Data:**
    Medical images (e.g., high-resolution brain scans) and AI analysis outputs (e.g., SHAP values, Grad-CAM visualizations) are stored in MongoDB. This flexibility allows for efficient handling of large volumes of multimedia data.
- **Data Security Measures:**
  Both data storage systems incorporate robust encryption mechanisms:

  - **Encryption at Rest:**
    Data stored in databases and cloud storage is encrypted using industry-standard protocols (such as AES-256) to prevent unauthorized access.
  - **Encryption in Transit:**
    Data exchanges between system components are secured with TLS 1.3 to ensure that sensitive information remains confidential.
- **Data Integrity and Backup:**
  Regular backups and audit trails are implemented to safeguard against data loss and ensure compliance with healthcare regulations (e.g., HIPAA). The system maintains a detailed log of user accesses and data modifications, enabling thorough audits and real-time recovery in case of failures.

2.5 Access control and security

Security is a cornerstone of the XAI Healthcare Bot design, ensuring that sensitive patient data is protected at every level of interaction:

- **Authentication:**
  The system employs multiple layers of user verification:
  - **Secure Session Management & Password Policies:**
    Enforcing strong password requirements and secure session handling.
  - **Multi-Factor Authentication (MFA):**
    Providing an additional layer of security to verify user identities.
  - **OAuth 2.0 Protocols:**
    Ensuring that third-party integrations adhere to secure authentication standards.
- **Authorization:**
  Access rights are managed through Role-Based Access Control (RBAC), where:
  - **Role Definitions:**
    Different roles (e.g., healthcare provider, patient, administrator) have tailored access privileges.
  - **Audit Trails:**
    Every access and modification is logged for accountability and regulatory compliance.
- **Encryption and Data Protection:**
  - **Data in Transit:**
    Secure connections via TLS 1.3 protect data during transmission.
  - **Data at Rest:**
    Storage encryption using AES-256 safeguards the stored data.
- **Compliance and Continuous Monitoring:**
  The system adheres to industry standards and regulatory requirements (such as HIPAA and GDPR) by:
  - **Regular Security Audits:**
    Conducting periodic assessments and vulnerability scans.
  - **Intrusion Detection Systems (IDS):**
    Monitoring for suspicious activities and potential breaches.
  - **Backup and Recovery:**
    Implementing robust backup procedures to ensure system resilience and rapid recovery.

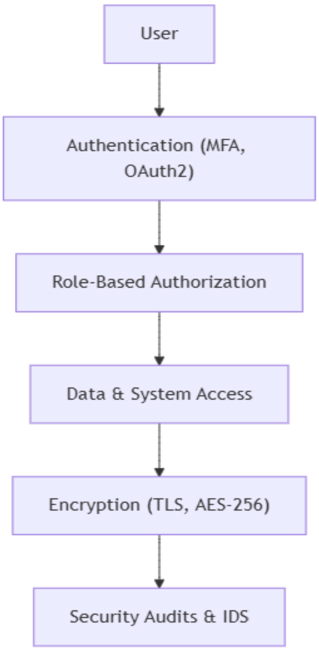

Figure 3. Access Control and Security Diagram

4. Subsystem services

4.1 Data Collection Subsystem

Data Collection Subsystem is responsible for acquiring, validating, and organizing data from various sources. This system ensures the integrity and accessibility of data for subsequent processing and analysis stages.

Services

● Data Acquisition: Collects data from imaging devices such as MRI, CT and X-ray using DICOM standard protocols.

● Data Validation: Performs quality control to ensure data completeness, format compliance, and accuracy. Validated data and metadata are forwarded to the Image Processing Subsystem for analysis.

● Metadata Extraction: Relevant metadata such as patient ID is extracted from DICOM headers.

● Data Storage: Raw imaging data and metadata are stored in a hybrid database architecture for retrieval by the Backend Service Subsystem (PostgreSQL for structured data, MongoDB for unstructured data).

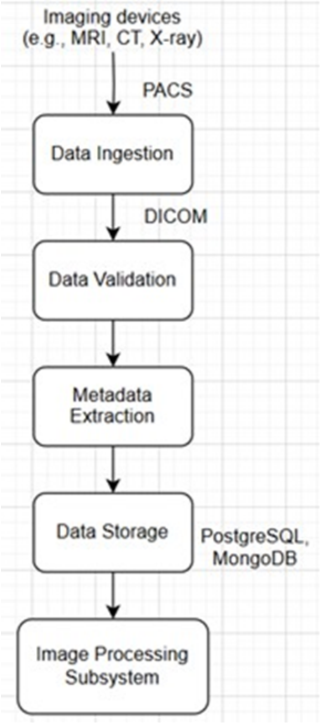

Figure 4. Data collection Subsystem

4.2 Image Processing Subsystem

Image Processing Subsystem is responsible for analyzing medical images and deriving meaningful insights. This subsystem utilizes machine learning models to detect abnormalities and generate explainable outputs.

Services

● Preprocessing: Normalizes the dimensions, resolutions, and formats of raw images and metadata received from the Data Collection Subsystem to ensure model compatibility.

● Feature Extraction: Applies convolutional neural networks like YOLO to extract key features from the images.

● Explainability Analysis: Generates SHAP values and Grad-CAM heatmaps to visualize model decisions.

● Result Packaging: Prepares analysis results such as abnormality detection scores and visual explanations for backend integration and sends them to the Backend Service Subsystem for delivery to the end user.

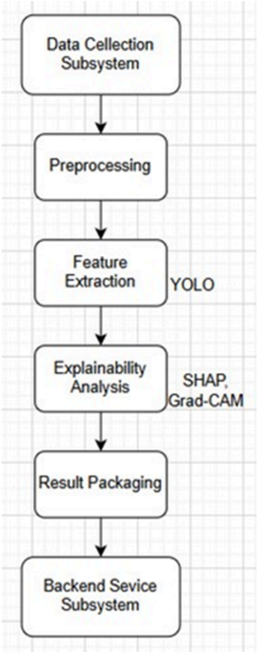

Figure 5. Image Processing Subsystem

4.3 Backend Service Subsystem

Backend Service Subsystem acts as the central application server that manages user interactions, data flow, and integration of subsystems.

Services

● API Gateway: Provides RESTful endpoints to facilitate communication between the frontend and subsystems.

● Data Aggregation: Combines results from the Image Processing Subsystem with metadata received from the Data Collection Subsystem.

● Access Control: Performs user authentication and authorization in compliance with HIPAA and GDPR standards.

● Result Delivery: Transmits processed results and explanations to the user interface.

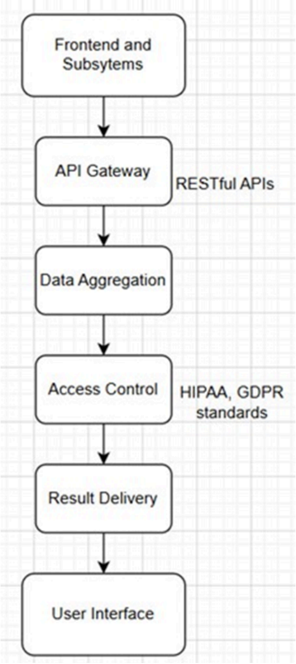

Figure 6. Backend Service Subsystem

5. Test Cases

Currently we are performing user roles and authorization tests. Here is a table of cases.

Test Name	Test Step	Expected Result
Login with correct password and username (Doctor)	A user with doctor role tries to login with correct username and password	Login should be successful
Login with correct password and username (Patient)	A user with patient role tries to login with correct username and password	Login should be successful
Login with incorrect role (Doctor)	A user with doctor role tries to login with correct username and password for the patient panel	Invalid email or password for selected role
Login with incorrect role (Patient)	A user with patient role tries to login with correct username and password for the doctor panel	Invalid email or password for selected role
Login with incorrect password and/or username (Doctor)	A user with doctor role tries to login with incorrect username and/or password	Invalid email or password for selected role
Login with incorrect password and/or username	A user with patient role tries to login with incorrect	Invalid email or password for selected role

(Patient)	username and/or password	

- **Test Items**
  - Test items have a crucial role to provide a reliable and high performance system.
  - Accuracy of ML model
  - Notification system
  - Data handling and security

Accuracy of ML Model

Since our bot is related to health, the accuracy is even more important. While testing this feature we want to see the accuracy of diagnosis. While doing it we want to split our dataset into train (%70), validation (%20) and testing (%10).

Notification System

When we test this feature, we want to be sure that correct notifications are sent to correct users.

Data Handling and Security

We will be working with a large amount of data, and this data should be stored properly and safely.

6. Consideration of Various Factors in Engineering Design

6.1 Constraints

Several constraints must be considered in the development of the XAI Healthcare Bot:

Technical Constraints

- Real-Time Processing: The system must process medical images and patient data with minimal latency.
- Data Quality: The accuracy of AI-driven diagnoses depends on the quality and diversity of training datasets.

Regulatory Constraints

- HIPAA & GDPR Compliance: The system must adhere to strict data privacy regulations to protect patient information.
- Medical Standards: The system must align with healthcare industry standards, including DICOM for image handling.

Ethical Constraints

- Bias Mitigation: The AI model must be trained on diverse datasets to prevent biased recommendations.
- Transparency: The system should provide explainable insights to users rather than just predictive results.

Environmental Constraints

- Infrastructure Limitations: The AI models require significant computational power, necessitating optimized cloud or on-premise deployment.
- Energy Consumption: Efficient resource allocation is needed to ensure sustainable operation.

6.2 Standards

The XAI Healthcare Bot adheres to several industry standards to ensure compliance and effectiveness:

- HIPAA (Health Insurance Portability and Accountability Act) – Ensuring secure handling of patient data.
- GDPR (General Data Protection Regulation) – Protecting user privacy and enforcing strict data access policies.
- DICOM (Digital Imaging and Communications in Medicine) – Standardizing medical image processing.
- ISO 27001 – Implementing best practices in information security management.
- IEEE / ACM Code of Ethics – Following ethical AI development guidelines.

7. Teamwork Details

7.1 Contributing and functioning effectively on the team

The XAI Healthcare Bot's success depends on efficient teamwork. Important elements consist of:

Task Distribution: Every team member has a designated role, such as frontend development, data security, or training AI models.

Code Reviews: To guarantee code quality and maintainability, regular peer reviews are conducted.

Agile development is the process of incorporating input and changes through an iterative development cycle.

7.2 Helping creating a collaborative and inclusive environment

In order to create a productive team atmosphere:

Encouraging team members to openly exchange ideas and criticism is known as "open communication."

Diverse Expertise: Making use of team members' talents in fields such as software engineering, artificial intelligence, and healthcare.

Frequent Stand-ups: Having weekly meetings to talk about achievements and difficulties.

Mentoring and Learning: Organizing sessions for the exchange of information to enhance team proficiency.

The XAI Healthcare Bot development team guarantees effectiveness, inclusiveness, and creativity in providing a top-notch healthcare solution by adhering to these collaborative principles.

7.3 Taking lead role and sharing leadership on the team

To maintain a well-rounded and efficient workflow, the XAI Healthcare Bot team shares leadership among its members. Key leadership responsibilities include:

Project Coordination: A lead project manager plans meetings, leads discussions, and makes sure deadlines are fulfilled.

Technical Leadership: Skilled team members manage system architecture, security implementations, and AI development.

Task Delegation: To maximize effectiveness and productivity, responsibilities are distributed according to competence.

Making Decisions: All team members contribute to the collaborative decision-making process when making important technical and strategic choices.

Mentoring and Support: In order to promote a culture of ongoing learning, senior team members offer advice to younger team members.

The team guarantees that each member makes a significant contribution and gains leadership experience inside the project by using a dispersed leadership style.

8. Glossary

Artificial Intelligence (AI): Technology that enables machines to mimic human intelligence, process data, and make decisions.

Explainable Artificial Intelligence (XAI): AI that provides transparent explanations of its reasoning processes, enhancing trust and understanding.

SHapley Additive exPlanations (SHAP): A method for explaining individual predictions of AI models, breaking down feature importance.

Gradient-weighted Class Activation Mapping (Grad-CAM): Visualization technique for highlighting areas in medical images influencing AI decisions.

You Only Look Once (YOLO): A real-time object detection algorithm used for analyzing medical images efficiently.

Convolutional Neural Network (CNN): A type of neural network optimized for analyzing visual data like medical images.

Digital Imaging and Communications in Medicine (DICOM): Standard for storing and transmitting medical imaging data across healthcare systems.

Application Programming Interface (API): A set of protocols allowing communication between different software systems or components.

HIPAA: Health Insurance Portability and Accountability Act, a US regulation ensuring patient data privacy.

GDPR: General Data Protection Regulation, an EU regulation safeguarding personal data privacy.

Electronic Health Records (EHR): Digital records of patient health information shared across healthcare systems.

PostgreSQL: A relational database system used to manage structured data like patient records.

MongoDB: A NoSQL database used to manage unstructured data like medical images and AI results.

9. References

- ACM Code of Ethics and Professional Conduct

- Object-Oriented Software Engineering, Using UML, Patterns, and Java, 2nd Edition, by Bernd Bruegge and Allen H. Dutoit, Prentice-Hall, 2004, ISBN: 0-13-047110-0.

- Computer Engineering Department, CS492 Senior Design Project II, 2025, https://www.cs.bilkent.edu.tr/~cs4912/current/CS492_deliverables.html