TED UNIVERSITY

Faculty of Engineering

Department of Software Engineering

Department of Computer Engineering

CMPE 491 / SENG 491

XAI Healthcare Bot

High Level Design Report

by

Doğukan Yetgin
Hilal Yurtoğlu
Ruşen Deniz Kaplan
Zeynep Bölükbaşı

Table of Content

1. Introduction

1.1 Purpose of the system

The XAI Healthcare Bot project aims to create a helpful AI system that makes medical diagnosis easier and clearer for both doctors and patients. Our system uses explainable AI technology to look at medical images and patient information to catch illnesses early. What makes our system special is that it clearly explains how it makes its decisions. The system helps doctors make better decisions about diagnosis while also helping patients understand their medical conditions in simple terms. By making the AI's decision-making process clear and easy to understand, we build trust between the system and its users. Our system is designed to save time and resources while making medical diagnosis more accurate. The XAI Healthcare Bot helps improve communication between doctors and patients, making healthcare better and more accessible for everyone.

1.2 Design Goals

1.2.1 Accuracy and Reliability

Our primary goal is to achieve high diagnostic accuracy through comprehensive model training using medical dataset. The system is designed to minimize false diagnoses by implementing robust validation processes and continuous model refinement.

1.2.2 Explainability and Transparency

At the core of our XAI Healthcare Bot is the commitment to transparency in medical decision-making. The system produces clear, easy-to-understand explanations for its diagnostic suggestions, breaking down complex medical terms into simple language. We achieve this by implementing visualization tools and detailed reasoning paths that help both healthcare providers and patients understand how the system reaches its conclusions.

1.2.3 Scalability and Performance

The system is built to efficiently handle large volumes of medical data and multiple user requests simultaneously. Our architecture ensures smooth performance even as the user base grows and more

medical data is processed. The design incorporates efficient data processing techniques and optimized algorithms to maintain fast response times while managing increasing workloads.

1.2.4 Security and Data Protection

We prioritize the protection of sensitive medical information through robust security measures. The system implements strong encryption, secure access controls, and regular security audits to maintain data confidentiality. All data handling processes comply with healthcare privacy regulations, ensuring patient information remains protected at all times.

1.3 Definitions, Acronyms, and Abbreviations

Artificial Intelligence (AI): The core technology that enables computers to mimic human intelligence and make decisions. In our healthcare system, AI processes medical data to assist in diagnosis.

Explainable Artificial Intelligence (XAI): An advanced form of AI that not only makes decisions but also provides clear explanations for its reasoning. This is crucial for building trust in medical diagnosis.

SHapley Additive exPlanations (SHAP): A method we use to explain how our AI model makes decisions. It helps break down which factors were most important in reaching a diagnostic conclusion.

Gradient-weighted Class Activation Mapping (Grad-CAM): A visualization technique that helps highlight which parts of medical images the AI focuses on when making diagnoses, making the process more transparent.

You Only Look Once (YOLO): A fast and efficient algorithm we use for detecting and analyzing medical images in real-time, helping to speed up the diagnostic process.

Convolutional Neural Network (CNN): A specialized type of AI network that we use to process and analyze medical images, particularly good at identifying patterns and features in visual data.

Digital Imaging and Communications in Medicine (DICOM): The standard format for storing and transmitting medical images, ensuring compatibility across different healthcare systems.

Application Programming Interface (API): A set of rules that allows our healthcare bot to communicate with other medical systems and databases, enabling seamless data exchange.

1.4 Overview

The XAI Healthcare Bot project combines artificial intelligence with transparent decision-making to improve healthcare delivery. Our system uses advanced image analysis and patient data processing to help with medical diagnoses while clearly explaining how it reaches its conclusions.The system uses AI algorithms like YOLO and CNN to process medical images and detect potential health issues. These work together with explainable AI techniques like SHAP and Grad-CAM to show doctors and patients how decisions are made. We built the system using Python and various medical data processing libraries.

This report details our system's architecture, including our design goals, software components, and how different parts work together. We include diagrams that show how data flows through the system and how users interact with it. We also explain our choices of technology and how we maintain security and scalability while keeping the AI's decision-making process clear and understandable.

2. Current software architecture

Our project’s current architecture is like followings:

User Interface: Different interfaces for patients and doctors to interact with each other, follow appointments and see the other components.

Explainable AI Module: Core component of the system, real time diagnostic support with clear explanations.

Image Processing: Tries to apply to proper diagnosis to the provided image. Planning it to support common medical imaging format DICOM.

3. Proposed software architecture

3.1 Overview

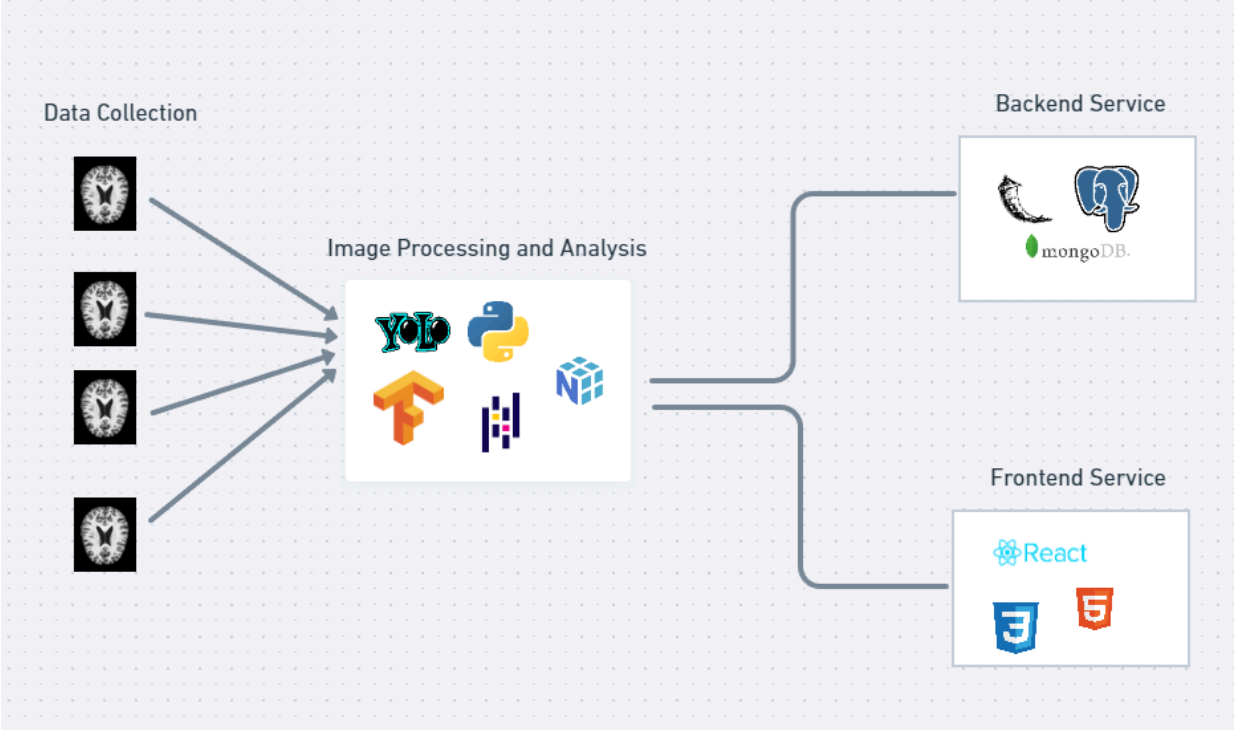

Image 1: General Architecture of the XAI Healthcare

Our XAI Healthcare Bot project aims to develop the project using Explainable Artificial Intelligence (XAI) techniques to ensure transparency and explainability in diagnoses in the field of healthcare. For Alzheimer's disease, which is a more specific area of the health domain, the system will use SHAP (SHapley Additive exPlanations) or Grad-CAM (Gradient-weighted Class Activation Mapping) for the detection and explanation of this disease using the Alzheimer's dataset. Additionally, YOLO (You Only Look Once) will be used for image processing and feature extraction. This project is designed to assist healthcare professionals and patients by providing real-time notifications, diagnostic assistance, medication tracking, and customizable profiles.

3.2 Subsystem decomposition

The functional subsystems of the system are as follows:

Data Preprocessing and Feature Extraction:

- Alzheimer dataset containing images and/or tables is used as input
- Data cleaning, normalization and image analysis is performed using YOLO.
- Ready processed features for classification and interpretability tasks are taken as output.

AI-based Diagnosis Subsystem:

- As an algorithm, SHAP/Grad-CAM are used to obtain interpretable outputs and YOLO is used for image processing.
- Functionally, the system diagnoses diseases and provides explainable results.
- Connected to the diagnostic support module and patient records to provide real-time insights.

Notification and Alert Subsystem:

- The system sends real-time notifications for medication reminders, health indicators, and diagnostic updates.

Profile Management Subsystem:

- Users are divided into two groups: patients and healthcare professionals.
- Created profiles can be customized and managed.
- There is access control to profiles for security.

Persistent Data Management Subsystem:

- Patient records, medication charts and diagnostic reports are stored in the database.
- Interoperability with existing Electronic Health Records (EHR) systems.

Security and Compliance Subsystem:

- The standards comply with HIPAA and GDPR compliance requirements.
- Secure authentication and encryption mechanisms to protect sensitive data are features of the system.

User Interface Subsystem:

- The system is designed to support users with different levels of technical expertise.
- In terms of accessibility, it offers an intuitive interface with screen reader-friendly and adaptable layouts.
- It uses the Alzheimer dataset as input, including images and/or tables.

3.3 Hardware/software mapping

3.3.1 Hardware Components

• Cloud-based Servers with GPU Support:

Training and implementing AI models for Alzheimer's detection requires high-performance computing infrastructure, such as servers that offer the computational capacity and parallel processing capabilities required by GPU acceleration.

• End-user Devices:

Users can engage with the AI diagnostic tools and access the system interface using a variety of devices, such as smartphones, tablets, and personal computers for best performance, these devices need to be compatible with contemporary web browsers.

• High-resolution Imaging Equipment:

High-quality brain scans and other important medical imaging data are collected by specialized medical imaging machines in order to diagnose Alzheimer's disease. These tools ensure accurate and comprehensive datasets are collected for training AI models.

3.3.2 Software Components

• Backend Framework:

Python-based implementation for training and developing AI models that makes use of deep learning frameworks (TensorFlow/PyTorch) for smooth data flow, web frameworks such as Flask or Django manage server-side logic and API endpoints.

• Frontend Development:

React or Vue.js-built modern web-based user interface that offers healthcare workers a responsive and easy-to-use experience. Makes it possible to organize patient data and visualize AI analysis results effectively.

• Database Architecture:

Hybrid database approach that uses PostgreSQL to manage structured patient data and diagnostic outcomes with MongoDB to handle unstructured medical imaging data.

• AI and Analysis Libraries:

- SHAP: Implements explainable AI features for transparent decision-making

- Grad-CAM: Provides visualization of AI model focus areas in medical images

- YOLO: Enables rapid object detection in medical imaging analysis

- NumPy/Pandas: Support data manipulation and numerical computations

3.3.3 Mapping

• AI Models and Processing:

GPU-enabled cloud servers power deep learning models, which use TensorFlow/PyTorch for inference and training.

• Web Application:

Frontend components (React/Vue.js) deployed on web servers, communicating with backend APIs for data exchange and processing.

• Database Systems:

MongoDB and PostgreSQL databases hosted on dedicated database servers, ensuring data integrity and efficient retrieval.

• Analysis Tools:

SHAP, Grad-CAM, and other analytical libraries integrated within the backend infrastructure, running on the same servers as the AI models.

3.4 Persistent data management

Using a hybrid database architecture (PostgreSQL and MongoDB), the system employs a thorough and secure data management technique to efficiently handle a variety of medical and analytical data types. While MongoDB manages AI model interpretability data, such as SHAP analysis results, Grad-CAM visualizations, and model metadata, PostgreSQL is the main relational database used to store structured patient data, such as diagnostic reports, medication histories, patient records, and treatment outcomes. Encrypted high-resolution brain scans and medical pictures are stored on a secure cloud storage platform with effective metadata connections to patient records for medical imaging data. HIPAA compliance and security depend on the system's thorough audit trail, which tracks user access, data changes, and AI model usage. To guarantee safe and effective handling of all system data while supporting Alzheimer's diagnosis and monitoring goals, the entire infrastructure includes frequent backups, stringent access controls, data encryption at rest and in transit, healthcare data protection compliance measures, and a scalable architecture.

3.5 Access control and security

Authentication, authorization, encryption, and regulatory compliance are the four main pillars that support the system’s implementation of a thorough security architecture to safeguard private medical information. The system employs secure session management, password rules, multi-factor authentication (MFA), and OAuth 2.0 protocols for authentication. Role-based access control (RBAC), which offers audit recording capabilities and specific permissions for various medical staff jobs, is used to manage authorization. Secure key management is used together with TLS 1.3 for data in transit and AES-256 encryption for data at rest to guarantee data safety. The system uses intrusion detection systems, backup plans, and regular security assessments to ensure compliance with HIPAA and GDPR regulations. In order to defend against new threats and preserve essential healthcare functions, this security architecture is regularly reviewed and updated.

3.6 Global software control

A strong centralized architecture is implemented via the global software control system of the XAI Healthcare Bot, effectively managing and coordinating all system components. In order to guarantee regular updates and real-time performance monitoring of all AI models, the system is based on a centralized AI model management system that is housed on dedicated servers. A complex notification control system that manages all time-sensitive notifications, such as important health updates and prescription reminders, via a priority-based queue is integrated with this architecture. Along with extensive error handling features, automated recovery processes, and real-time monitoring, the system also has a scalable architecture that uses cloud services to dynamically modify resources in response to user demand. Standardized medical data processing and synchronization with external healthcare systems are managed by the data integration hub, while role-based access control and HIPAA compliance are enforced by a centralized security framework. By using multiple systems and failover methods, this integrated solution maintains high system availability while guaranteeing the safe, effective, and dependable delivery of healthcare insights to patients and healthcare practitioners.

3.7 Boundary Conditions

The XAI Healthcare Bot system has clearly defined boundary conditions that handle all aspects of system operation, from startup to shutdown:

Initialization:

The system starts by installing the basic components and establishing connections. Users access the platform via the web interface using their credentials. Patient accounts require standard verification, while healthcare provider accounts go through additional authentication steps. Upon successful login, users are directed to role-specific dashboards where they can access health records, diagnostic support, and notifications based on their permissions.

Normal Operation:

During regular operation, users can actively access various system features. Healthcare providers can access patient records, get AI-powered diagnostic support, and manage appointments. Patients can view

their medical history, receive medication reminders, and access simplified explanations of their diagnoses.

Failure Handling:

The system implements comprehensive failure recovery mechanisms for different scenarios:

Database Failure: Although preserving data consistency, the system immediately moves to a redundant backup database. Since the system keeps running on the backup infrastructure, users encounter no disruptions.

AI Model Failure: When the primary AI model encounters issues, the system smoothly transitions to a stable baseline model. This ensures continuous service while logging the failure for investigation.

Notification System Failure: If the notification service becomes unavailable, the system queues all pending alerts in a secure buffer. Once service is restored, notifications are delivered in order of priority.

Network Connectivity Issues: The system maintains session state and unsaved data, allowing users to resume their activities once connection is restored.

Termination

Users can safely exit the system through multiple methods:

Using the logout option, which properly closes their session and secures any sensitive data

Closing the browser or app, which triggers automatic session cleanup

System-initiated timeout after extended inactivity, protecting user privacy

During shutdown, the system ensures all active processes are completed, data is properly saved, and logs are archived for future analysis.

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).

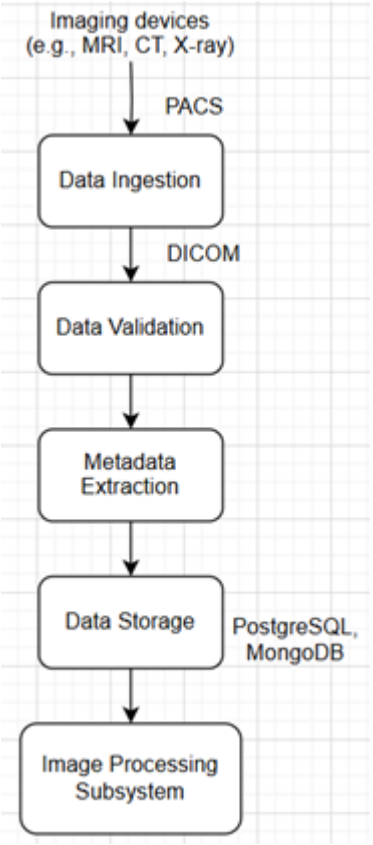

Figure 2: Data collection Subsystem

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.

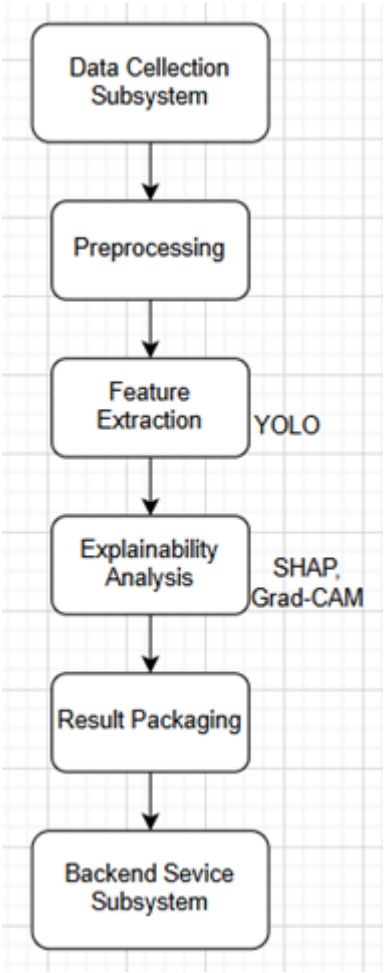

Figure 3: Image Processing Subsystem

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.

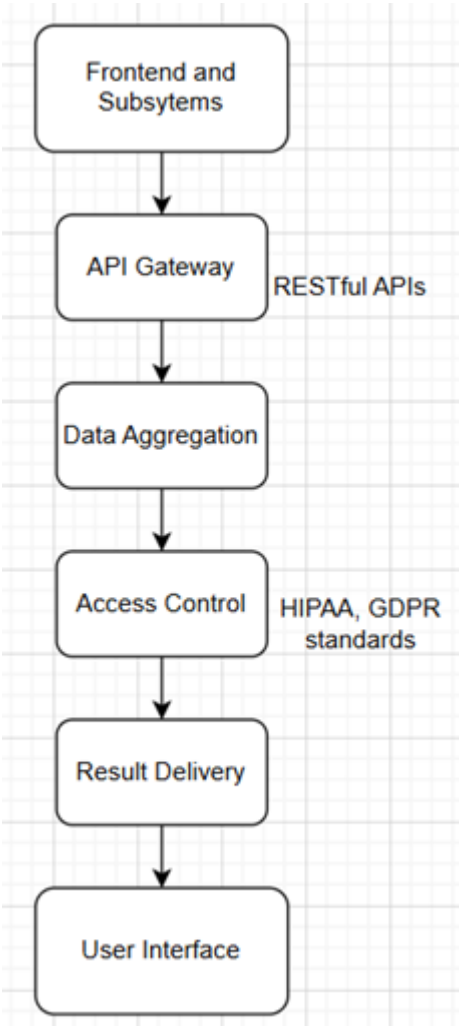

Figure 4: Backend Service Subsystem

5. 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.

6. References

[1]: Rahim, N., Abuhmed, T., Mirjalili, S., El-Sappagh, S., & Muhammad, K. (2023). Time-series visual explainability for Alzheimer's disease progression detection for smart healthcare. Alexandria Engineering Journal, 82, 484-502.

[2]: He, K., Zhang, X., Ren, S., & Sun, J. (2016). "Deep Residual Learning for Image Recognition." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[3]: Selvaraju, R. R., Cogswell, M., Das, A., et al. (2017). "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization." Proceedings of the IEEE International Conference on Computer Vision (ICCV).

[4]: Healthcare IT Standards: DICOM and EHR Interoperability Guides (2023). Accessed from healthcareitstandards.org.

[5]: U.S. Department of Health and Human Services. "Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule." Accessed from hhs.gov.

[6]: The Iterative Workspace for Product Teams. Whimsical. (n.d.). https://whimsical.com/

[7]:Nelson, J. (2024, October 15). *What is Yolo? the ultimate guide [2024]*. Roboflow Blog. https://blog-roboflow-com.translate.goog/guide-to-yolo-models/?_x_tr_sl=en&_x_tr_tl=tr&_x_tr_hl=tr&_x_tr_pto=wa