

 **COMSATS University Islamabad,**

**COMSATS Road, off GT Road, Sahiwal, Pakistan**

**Project Proposal**

**(SCOPE DOCUMENT)**

**for**

**DiagnoXpert**

Version 1.0

***By***

**Eman Khadim CIIT/FA22-BCS-066/SWL**

**Saira Bano CIIT/FA22-BCS-107/SWL**

**Wasib Mehmood CIIT/FA22-BCS-130/SWL**

***Supervisor***

**Dr. Yawar Abbas Abid**

***Co-Supervisor***

**Muhammad Jamil**

***Bachelor of Science in Computer Science (2022-2026)***

**SCOPE DOCUMENT REVISION HISTORY**

|  |  |  |
| --- | --- | --- |
| **No.** | **Comment** | **Action** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**Supervisor Signatures \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Date: \_\_\_\_\_\_\_\_\_\_\_\_\_**

Table of Contents

[1. Introduction 1](#_Toc165323149)

[2. Problem Statement 1](#_Toc165323150)

[3. Problem Solution for Proposed System 1](#_Toc165323151)

[4. Related System Analysis/Literature Review 2](#_Toc165323152)

[5. Advantages/Benefits of Proposed System 2](#_Toc165323153)

[6. Scope 2](#_Toc165323154)

[7. Modules 3](#_Toc165323155)

[7.1 Disease Prediction and Risk Analysis 3](#_Toc165323156)

[7.2 Health Score & Reccomendation 3](#_Toc165323157)

[7.3 Chatbot Integration 3](#_Toc165323158)

[7.4 Highlight abnormalities in Report 3](#_Toc165323159)

[7.5 Context-Based Medical Insights 3](#_Toc165323160)

[7.6 Severity Level Indication 3](#_Toc165323161)

[7.7 Different Report Type support 3](#_Toc165323162)

[7.8 Voice Input For Queries 3](#_Toc165323163)

[7.9 Multi-Format Report Processing 3](#_Toc165323164)

[7.10 Rating and Review System 3](#_Toc165323165)

[8. System Limitations/Constraints 4](#_Toc165323166)

[9. Software Process Methodology 4](#_Toc165323167)

[10. Tools and Technologies 5](#_Toc165323168)

[11. Project Stakeholders and Roles 5](#_Toc165323169)

[12. Team Members Individual Tasks/Work Division 6](#_Toc165323170)

[13. Data Gathering Approach 6](#_Toc165323171)

[14. Concepts 6](#_Toc165323172)

[14.1 Machine learning 6](#_Toc165323173)

[14.2 Database 6](#_Toc165323174)

[14.3 Data Science&Analytics. 6](#_Toc165323175)

[14.4 Cloud & Depolyment 6](#_Toc165323176)

[14.5 Frontend and Backend 6](#_Toc165323177)

[14.6 Version Control 6](#_Toc165323177)

[15. Gantt chart 7](#_Toc165323178)

[16. Mockups 9](#_Toc165323178)

[17. Conclusion 12](#_Toc165323180)

[References 12](#_Toc165323181)

**List of Figures**

[Figure 1 Gantt Chart 8](#_Toc165319585)

[Figure 2 Web Application Interface 10](#_Toc165319585)

[Figure 3 Diagnostic Insights Visualization 10](#_Toc165319585)

[Figure 4 Insight Summary 11](#_Toc165319585)

[Figure 5 Chatbot Interface 11](#_Toc165319585)

**List of Tables**

[*Table 1 Related system analysis* 2](#_Toc165319595)

[*Table 2 Tools and Technologies* 5](#_Toc165319596)

[*Table 3 Stakeholder* 5](#_Toc165319597)

[*Table 4 Team Members Tasks* 6](#_Toc165319598)

**Abstract**

DiagnoXpert is an innovative platform that harnesses AI to revolutionize healthcare by accurately analyzing medical documents and offering intelligent, clear diagnostic insights. Utilizing advanced natural language processing (NLP) and optical character recognition (OCR), the platform pulls key data from multiple reports (such as blood tests) and provides clear interpretations, possible diagnoses and customization of health trends analysis. DiagnoXpert with its specially-created chatbot for smooth query resolution and voice-input support to make it more user friendly, ensures that patients are able to confidently engage in complex medical conversations. By using technology to increase access, understanding, and personalization, DiagnoXpert sets a new standard for modern, patient-focused healthcare system.

**Project Category:**

1. Web Development
2. App Development
3. Artificial Intelligence

# Introduction

Artificial intelligence (AI) innovations including natural language processing (NLP) and optical character recognition (OCR) are revolutionizing how we view and deliver solutions in healthcare. If blood work and other medical notes were generated, the sometimes-complex information is not explained to individuals in a way they can understand without professional help. With DiagnoXpert, an AI-based tool utilizing these technologies, patients can directly access their health-related data and obtain clear, personalized diagnostic conclusions. Patients feel dissonance and cannot comprehend the technical aspects of medical data. That uncertainty leads them to make misguided decisions, and access limitations prevent them from engaging intuitively with technology. To facilitate fast diagnoses and analysis of health trends, DiagnoXpert relies on medical record analysis through natural language processing (NLP). Its voice-driven chatbot ensures user convenience and accessibility. DiagnoXpert empowers patients, reduces the need for medical experts to provide basic explanations, and establishes a new benchmark for the addition and clarity in digital healthcare by combining advanced AI with **user-centric design.**

# Problem Statement

In health care, getting the correct diagnosis for a patient is critically important, but reading medical reports is very time-consuming and requires expertise. Medical records are typically difficult for patients to comprehend because of technical wording and data. And this ambiguity can contribute to confusion, anxiety, and suboptimal health choices. For a lot of people doctors are their only source of explanation, not to mention not everybody has a doctor nearby, or at all.While some digital tools exist, most are designed for healthcare professionals and fail to offer **patient-friendly interpretations** of diagnostic data. There's a growing need for an AI-powered solution that simplifies medical information for patients and helps them take control of their health.

# Problem Solution for Proposed System

DiagnoXpert solves the problem of patients not being able to comprehend complex medical test reports through an integrated, AI-based solution on easy to-use mobile and web platform. Leveraging AI, natural language processing (NLP), OCR and more, the system automates the analysis of different types of medical reports (text PDFs, scanned blood tests…), extracts key values, compares them to health standards, and provides clear diagnostic insights, trend analysis and personal recommendations. Additionally, an AI-driven chatbot provides real-time, patient-friendly explanations, enhancing accessibility. DiagnoXpert revolutionizes healthcare by **simplifying medical data**, boosting patient engagement, and reducing **the manual workload** for professionals, ultimately fostering better health outcomes.

# Related System Analysis/Literature Review

We studied some existing systems, like IBM Watson Health, which is a popular website used by hospitals and doctors**. IBM Watson Health** helps doctors analyze medical data, but it has a complex interface that can be hard for regular people to use. Our proposed project is designed specifically for patients. It has a simple interface that’s easy to use, making it more convenient for patients to understand their medical reports and get health insights.

*Table 1 Related system analysis*

|  |  |  |
| --- | --- | --- |
| **Application Name** | **Weakness** | **Proposed Project Solution** |
| IBM Watson Health | * Complex interface, hard for patients to use. * Not customized for local health needs. * Designed for hospitals and doctors, not for individual patients. | Our proposed project offers a simple, easy-to-use interface for patients. It adapts to local health needs and helps patients understand their medical reports with personalized insights, making healthcare more accessible and user friendly. |

# Advantages/Benefits of Proposed System

Here are the following advantages of the systems.

* 1. **Enlightens and empowers patients** by interpreting their complex blood test results into easy to read and understand the language
  2. **Saving time and reducing anxiety** by instantly delivering clear diagnostic insights.
  3. **Minimize interpretation errors** through consistent, AI-driven report analysis.
  4. **Improve accessibility** for users in remote or underserved areas without frequent access to specialists.
  5. **Supports early detection** by flagging critical values and prompting users to seek timely medical attention.
  6. **Continuously evolves** through adaptive learning based on feedback and new medical data.

# Scope

DiagnoXpert specifically works in analyzing Blood Test report (as a start point) through AI, NLP and OCR to extract critical health indicators and to create easily understood report summaries for patients. It accommodates structured pdf and scanned documents, and has a user-friendly web-interface that is designed for patients. The project is focused only on blood test reports for now, but the scope is huge. **Future enhancements** may include the analysis of radiology data such as X-rays and CT scans, along with integration with hospital management systems and **Electronic Medical Records** **(**EMRs**)**. This would allow healthcare professionals to use the platform alongside the patients. DiagnoXpert is designed to **support medical decisions**, not replace them, with the aim of enhancing accuracy, efficiency, and health awareness.

# Modules

The following modules form the core of the DiagnoXpert system, ensuring robust functionality and user-centric healthcare solutions.

**7.1 Disease Prediction and Risk Analysis**

The system uses machine learning algorithms to predict potential illnesses from medical report data. It analyzes trends in the test results and patient history to determine risk levels.

**7.2 Health Score & Recommendation**

DiagnoXpert generates a personalized health score based on medical data for comprehensive health tracking. It will provide tailored lifestyle or nutritional advice driven by AI insights.

**7.3 Chatbot Integration**

The platform will incorporate an NLP-powered chatbot for seamless user interaction with medical insights. It will handle queries and deliver context-aware explanations of report details.

**7.4 Highlight Abnormalities in Report**

The system will employ AI analysis to identify and flag abnormal data in medical reports. It will visually highlight critical metrics for easy user recognition and follow-up.

**7.5 Context-Based Medical Insights**

DiagnoXpert will deliver insights by correlating report data with patient-specific information. It will leverage NLP to produce meaningful interpretations tailored to user health profiles.

**7.6 Severity Level Indication**

The system will assign severity levels to detected abnormalities to prioritize medical attention. It will use rule-based or ML models to classify issues as mild, moderate, or severe.

**7.7 Different Report Type Support**

The platform will process various medical report types, including blood tests and imaging reports. It will utilize OCR and NLP to extract data from diverse formats and structures.

**7.8 Voice Input for Queries**

DiagnoXpert will enable voice-based querying through automated speech recognition (ASR) technology. It will enhance accessibility for users with disabilities or limited technical skills.

**7.9 Multi-Format Report Processing**

The system will support multiple report formats, such as PDFs, scanned images, and digital text. It will standardize data extraction using advanced OCR for consistent analysis.

**7.10 Rating and Review System**

The platform will include a user feedback mechanism to rate and review insights. It will improve AI models and user experience through iterative feedback analysis.

# System Limitations/Constraints

The DiagnoXpert AI-Powered Medical Report Analysis System has certain limitations that need to be addressed to ensure its effectiveness and usability. These include:

* **Dependent on report quality (OCR may fail on poor scans)**: The system’s OCR technology struggles with blurry or low-quality scans, leading to errors in data extraction. This can affect the accuracy of the analysis and insights provided.
* **Limited to diagnostic report types (not prescriptions or handwritten notes)**: DiagnoXpert only processes structured diagnostic reports like blood tests, excluding prescriptions or handwritten notes. This restricts its ability to analyze a wider range of medical documents.
* **Requires internet connectivity for cloud-based models**: The system needs a stable internet connection to access its cloud-based AI models for processing. Without internet, users in remote or offline areas cannot use the system.
* **Not a substitute for actual medical consultation**: While DiagnoXpert provides helpful insights, it cannot replace professional medical advice from a doctor. It should only be used as a supplementary tool for understanding health reports.

# Software Process Methodology

We will use the **Object-Oriented Methodology (OOM)**, which is ideal for modular design and code reusability. This approach allows for clear separation of concerns across different modules like user management, report processing, and results generation. It also aligns well with Python-based frameworks such as Django or Flask, enabling rapid development and easy maintenance. Since the system currently focuses only on **structured blood test reports** (PDF/text format), OOM ensures clean handling of input processing, analysis, and output generation without complexity.

# Tools and Technologies

These are all the tools and technologies which we’re going to use throughout the project. Example:

*Table 2 Tools and Technologies*

|  |  |  |  |
| --- | --- | --- | --- |
| **Tools And**  **Technologies** | **Tools** | **Version** | **Rationale** |
| Visual Studio | Latest | Web based project |
| Android Studio | Latest | App based project |
| Figma | Latest | Mockups |
| Overleaf | 365 | Documentation |
| MS PowerPoint | 365 | Presentation |
| **Technology** | **Version** | **Rationale** |
| JavaScript | Latest | Programming language |
| Python | Latest | Programming language |
| React Native |  | Framework |
| Chatbot Integration |  | Gemini |
| MongoDB |  | Database |
|  | Node Express |  | For API |

# Project Stakeholders and Roles

The following are the stakeholders of our projects.

|  |  |
| --- | --- |
| **Project Sponsor** | COMSATS University Islamabad, Sahiwal Campus |
| **Stakeholder** | * Eman Khadim * Saira Bano * Wasib Mehmood * Dr. Yawar Abbas Abid * Final Year Project Committee: Evolution of project |

*Table 3 Stakeholder*

# Team Members Individual Tasks/Work Division

The following are the stakeholders of our projects.

|  |  |  |
| --- | --- | --- |
| **Name** | **Registration no** | **Role and tasks** |
| Eman Khadim | FA22-BCS-066 | * Frontend/Backend * Application design * Model Training   Model Integration |
| Saira Bano | FA22-BCS-107 | * Documentation * Requirements Gathering * Data Gathering * Front end |
| Wasib Mehmood | FA22-BCS-130 | * Data Analysis * ETL Process * Model Training/ Model Integration * Development of OCR |

*Table 4 Team Members Tasks*

# Data Gathering Approach

Following are the techniques we used to gather the requirements.

a.Public and Open-source Datasets

b. Collaboration with Diagnostic Labs

c. Crowdsourcing and User Contributions

d. Synthetic Data Generation

# 14. Concepts

**14.1 Machine Learning:**

Machine learning makes it possible to anticipate diseases and analyze health trends. TensorFlow models categorize laboratory data for disease prediction and risk analysis.

**Example:** Using glucose levels to predict diabetes.

**14.2 Database:**

Database store and manage medical report data, user profiles, and feedback for DiagnoXpert’s operations. Indexing and query optimization enable quick retrieval of Health Score and Recommendation calculations.

**Example:** retrieving blood test results for chatbot inquiries.

**14.3 Data Science & Analytics**

Data science uses statistical analysis to extract insights from report data. Analytics models track health trends, enabling personalized insights for Context-Based Medical Insights.

**Example:** Analyzing glucose trends to suggest lifestyle changes.

**14.4 Cloud & Deployment**

Cloud infrastructure ensures DiagnoXpert’s scalability, security, and accessibility across modules. AWS or Google Cloud hosts ML models, OCR pipelines, and web applications, supporting Different Report Type Support.

**Example:** Running the chatbot for Voice Input for Queries.

**14.5 Frontend and Backend:**

React with Tailwind. CSS creates an accessible UI, whereas Express manages APIs.

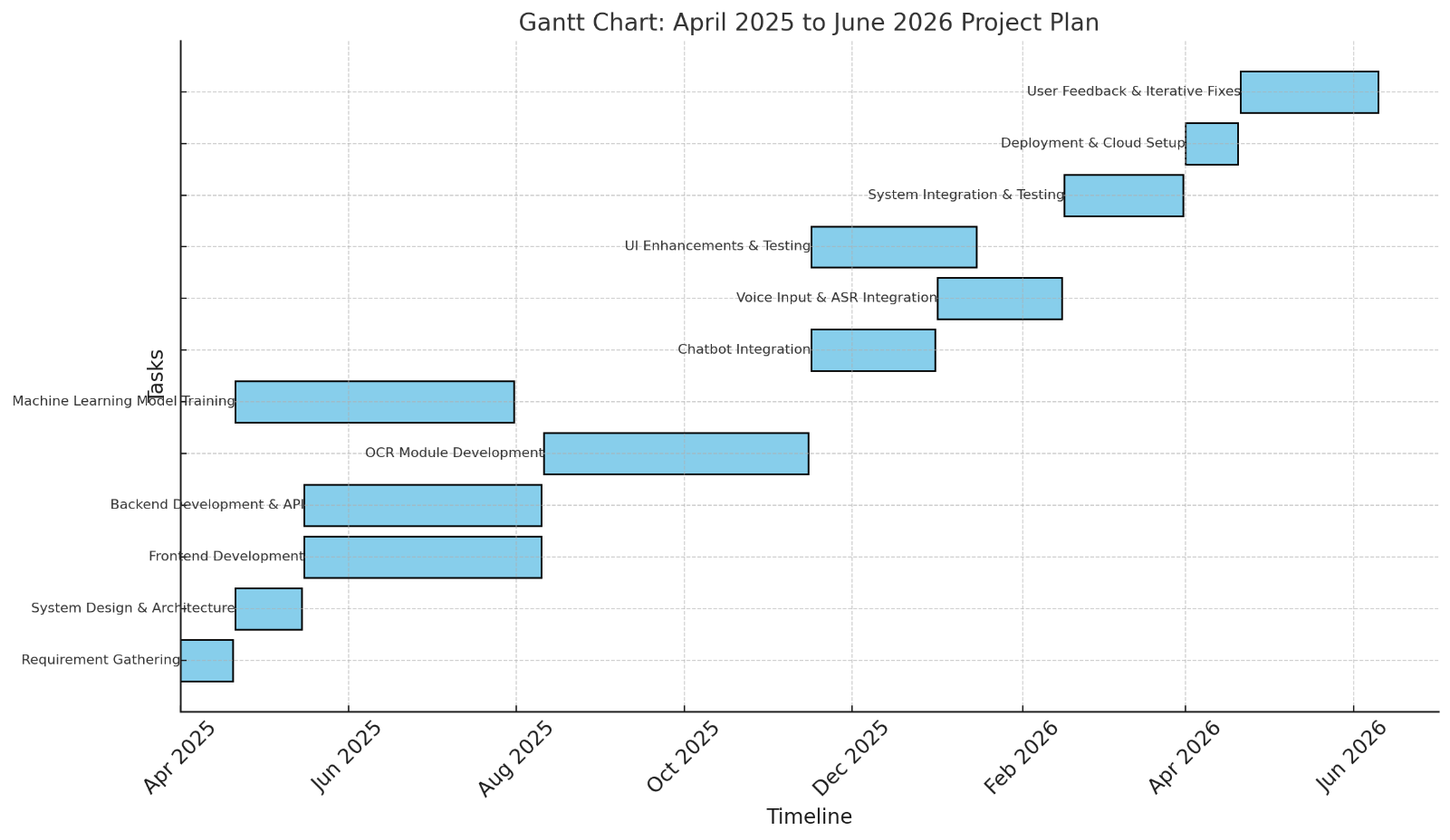
**Example:** Displaying flagged lab values on a dashboard.

**14.6 Version Control**

Git manages code for Frontend and Backend, and ML modules collaboratively. Branching ensures stable Multi-Format Report Processing updates.

**Example:** Committing a Tesseract OCR update via GitHub.

# 15. Gantt chart

Figure 1 Gantt Chart

**16. Mockups**

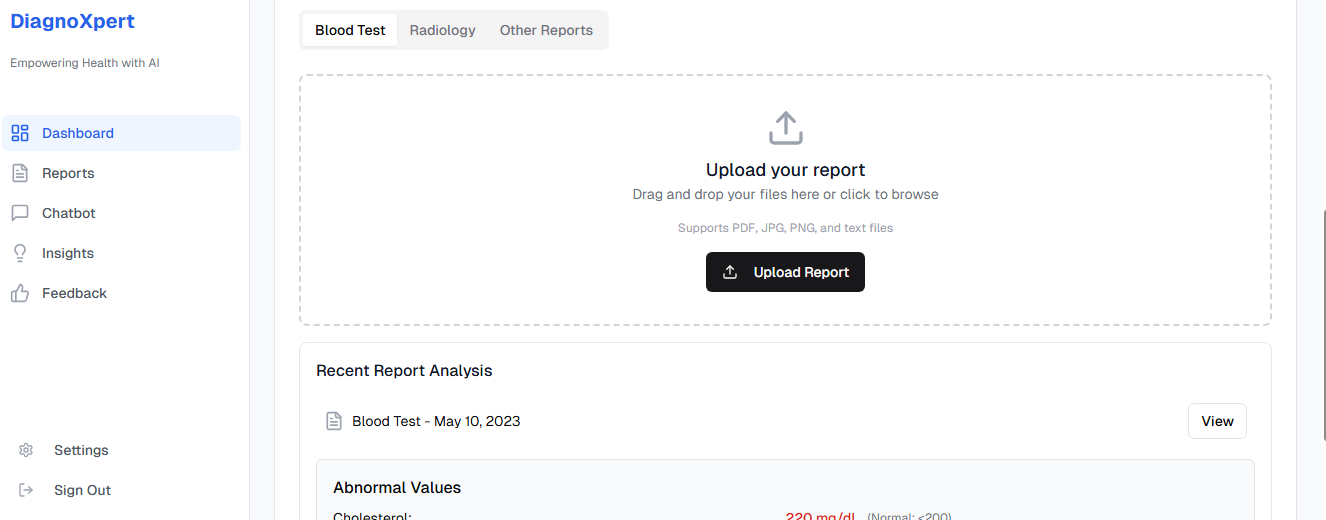


Figure 2 Web Application Interface

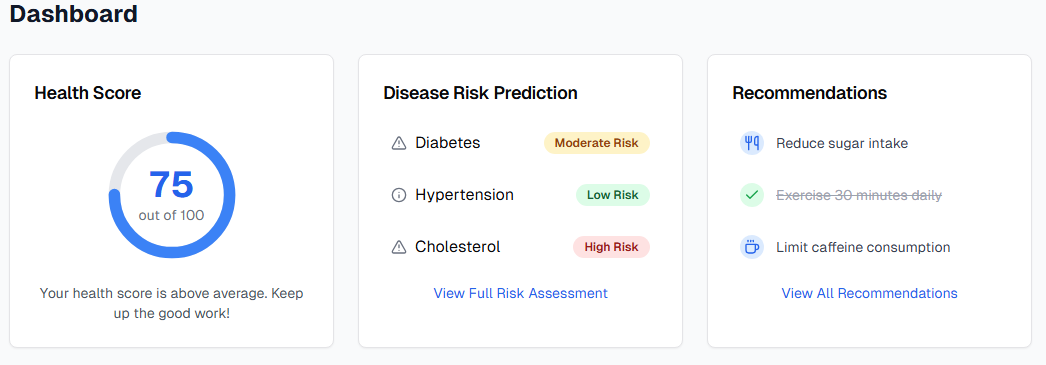
****

Figure 3 Diagnostic Insights Visualization

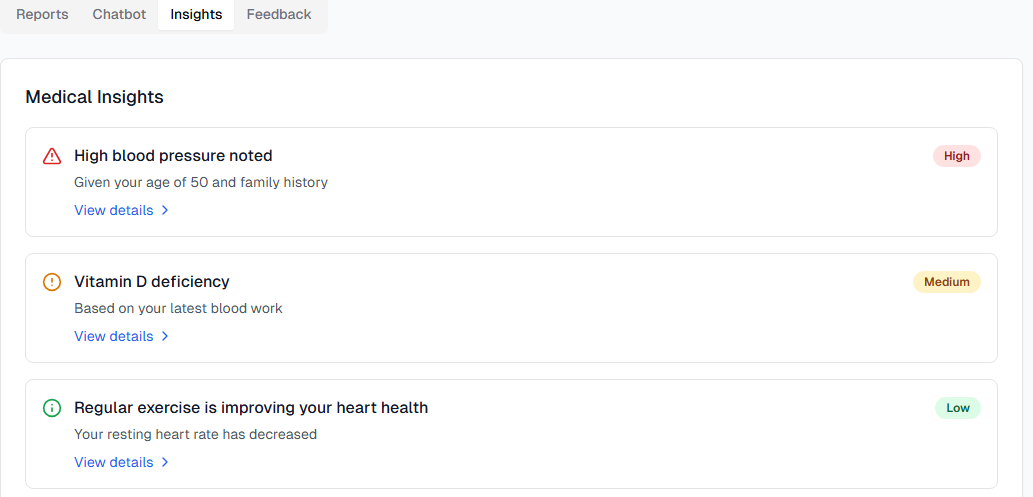
****

Figure 4 Insight Summary

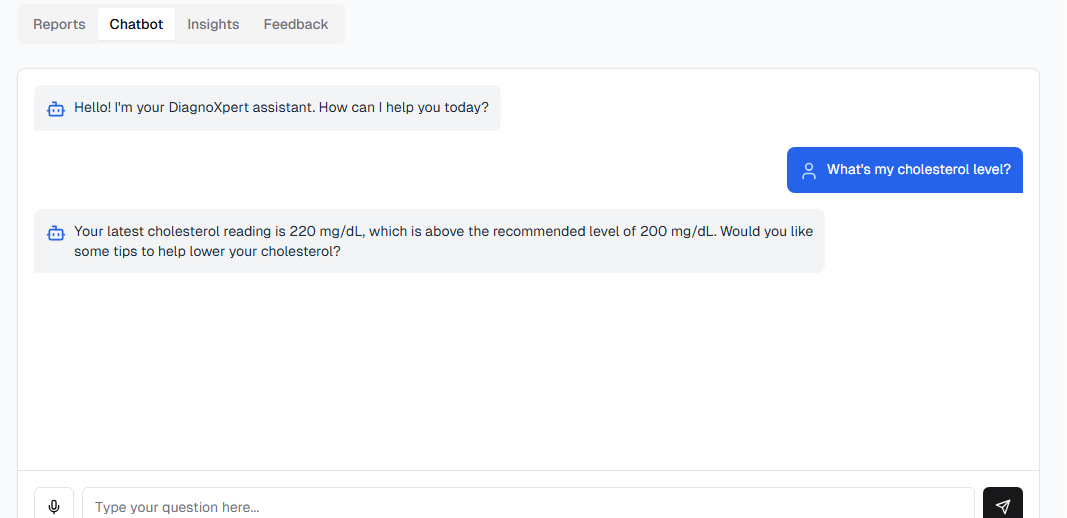
****

Figure 5 Chatbot Interface

# 17. Conclusion

DiagnoXpert is a game changer in **patient-centered** healthcare, employing artificial intelligence to provide clear, accessible, and tailored diagnostic insights from complicated medical information. The platform bridges the gap between technical medical data and patient comprehension by providing powerful modules such as **illness prediction** and **health score recommendations**, as well as voice-enabled chatbot interactions and multi-format report processing. DiagnoXpert establishes a new standard for digital healthcare solutions by stressing accessibility, clarity, and inclusion. Future developments in real-time monitoring and broader report compatibility have the potential to further empower patients globally.

# References

|  |  |
| --- | --- |
| [1] | Abhishek Kumar, Ashutosh Kumar Dubey, Surbhi Bhatia, Swarn Avinash Kumar, Dac-Nhuong Le, Evolving Predictive Analytics in Healthcare: New AI Techniques for Real-time Interventions, The Institution of Engineering and Technology, 2022. |
| [2] | C Friedman, G Hripcsak , 1999 - columbia.edu, "Natural language processing and its future in medicine," columbia.edu, 1999. |
| [3] | OD Kazakov, AV Averchenkov,N Yu Kulagina, "Machine learning in the analysis of medical data," in *Journal of Physics*, 2019. |
| [4] | Gregor Gunčar, Matjaž Kukar, Mateja Notar, Miran Brvar, Peter Černelč, Manca Notar & Marko Notar, "An application of machine learning to haematological diagnosis," *Scientific Reports ,* no. 11 January , p. 12, 2018. |