

FINE TUNING LLM BASED ON LEUKEMIA THE BLOOD CANCER

INTELLIGENCE OF BIOLOGICAL SYSTEMS - 1

Team - D06

NAREN SUNDAR L	-	CB.SC.U4AIE23322
VATTURU PARDHEEV	-	CB.SC.U4AIE23344
VIBHU SANCHANA	-	CB.SC.U4AIE23347
ASMI K	-	CB.SC.U4AIE23351

INTRODUCTION

Leukemia Overview:

- A type of blood cancer affecting normal blood cell production, often requires early detection and personalized care.

Challenges for Patients:

- Limited access to reliable guidance for managing symptoms, diet, and treatment, along with challenges in interpreting lab reports.

Project Goal:

- To create an AI-powered assistant that supports leukemia patients with query handling, personalized nutritional guides, and lab report analysis.



PROBLEM STATEMENT

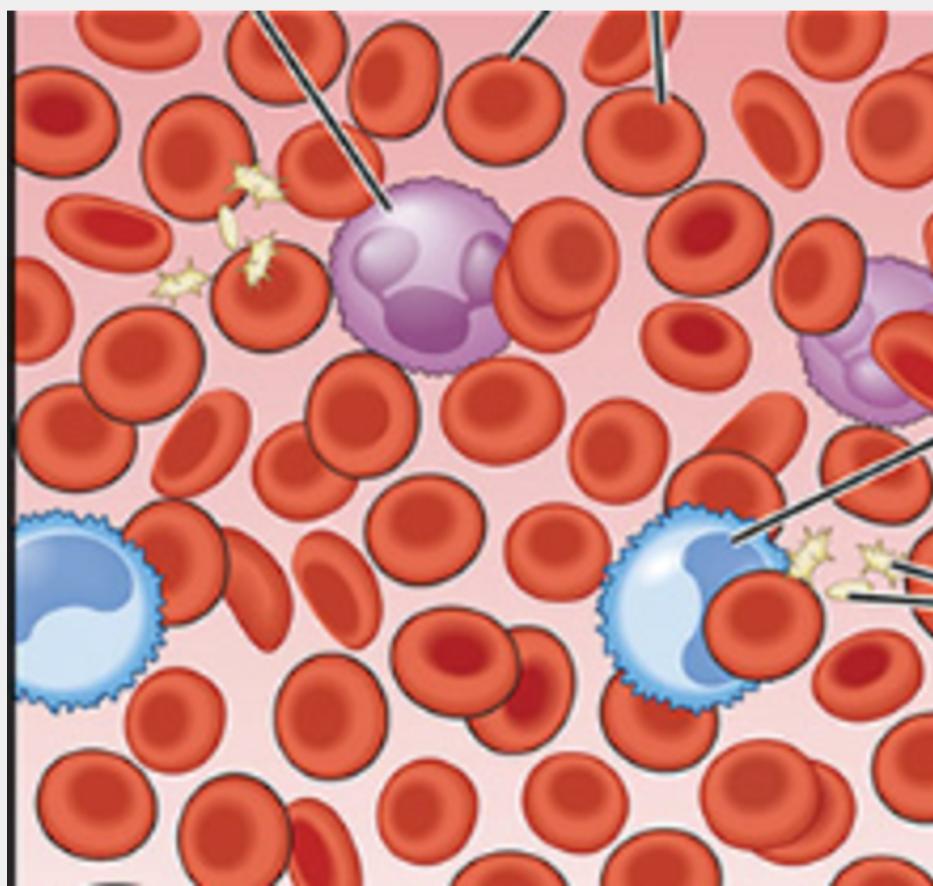
- Limited access to reliable, personalized tools for blood cancer assistance.
- Inefficient methods for providing tailored nutritional guidance.
- Challenges in analyzing and interpreting lab reports for potential leukemia indicators.

Solution:

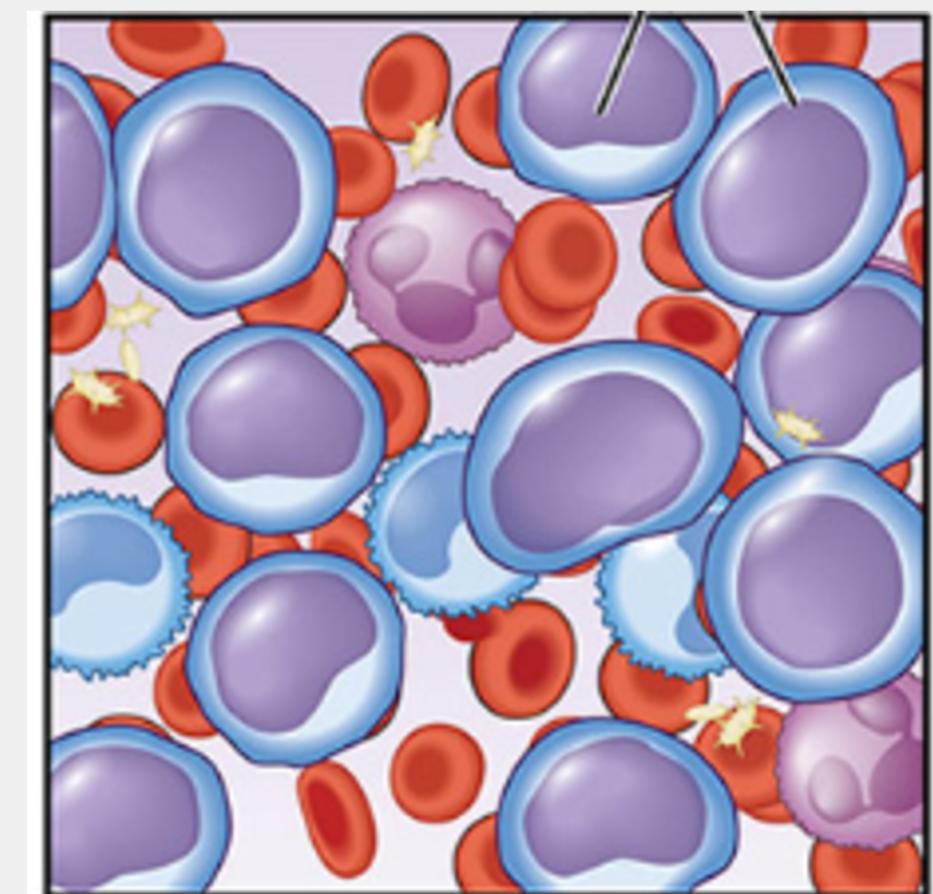
- An AI-powered assistant to address blood cancer-related queries.
- Personalized nutritional guide generation based on patient-specific data.
- Automated lab report analysis for preliminary leukemia diagnosis and recommendations.

LEUKEMIA

- Leukemia is a type of cancer found in your blood and bone marrow and is caused by the rapid production of abnormal white blood cells.
- These abnormal white blood cells are not able to fight infection and impair the ability of the bone marrow to produce red blood cells and platelets.
- Unlike other cancers, leukemia doesn't generally form a mass that shows up in imaging tests, such as X-rays or CT scans.
- There are many types of leukemia. Some are more common in children, while others are more common in adults. Treatment depends on the type of leukemia and other factors.



Normal blood



Leukemia

LEUKEMIA

Types:

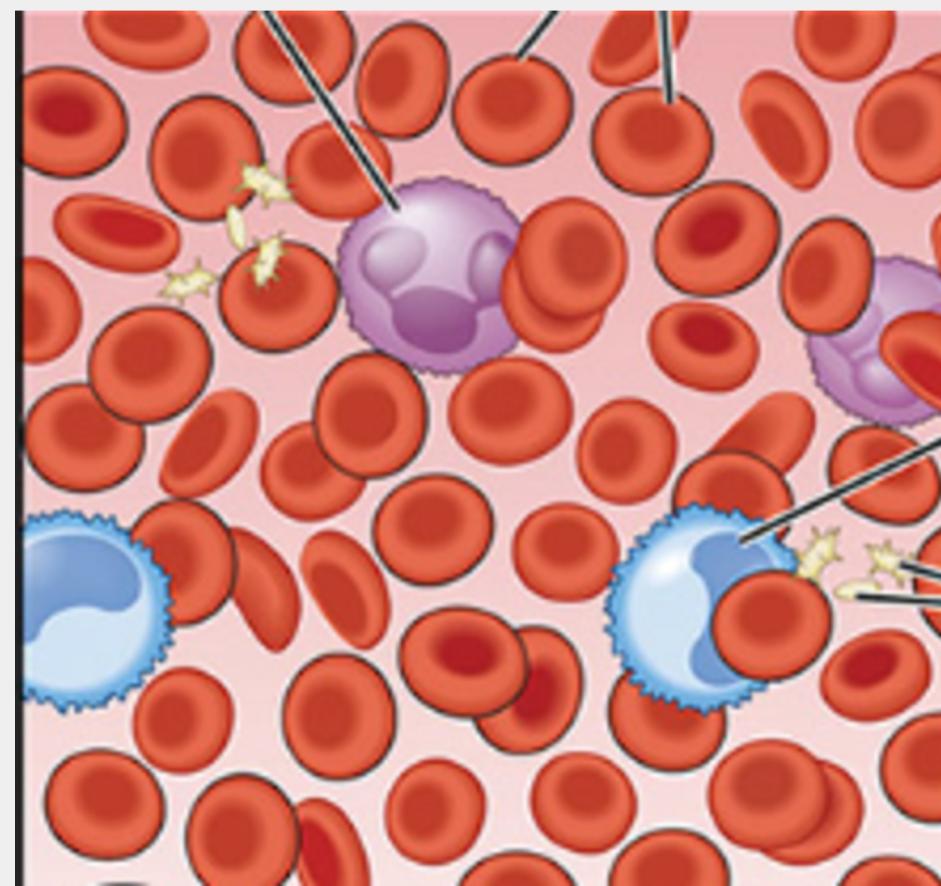
- Acute: Rapid progression (e.g., Acute Lymphoblastic Leukemia - ALL, Acute Myeloid Leukemia - AML).
- Chronic: Slower progression (e.g., Chronic Lymphocytic Leukemia - CLL, Chronic Myeloid Leukemia - CML).

Causes:

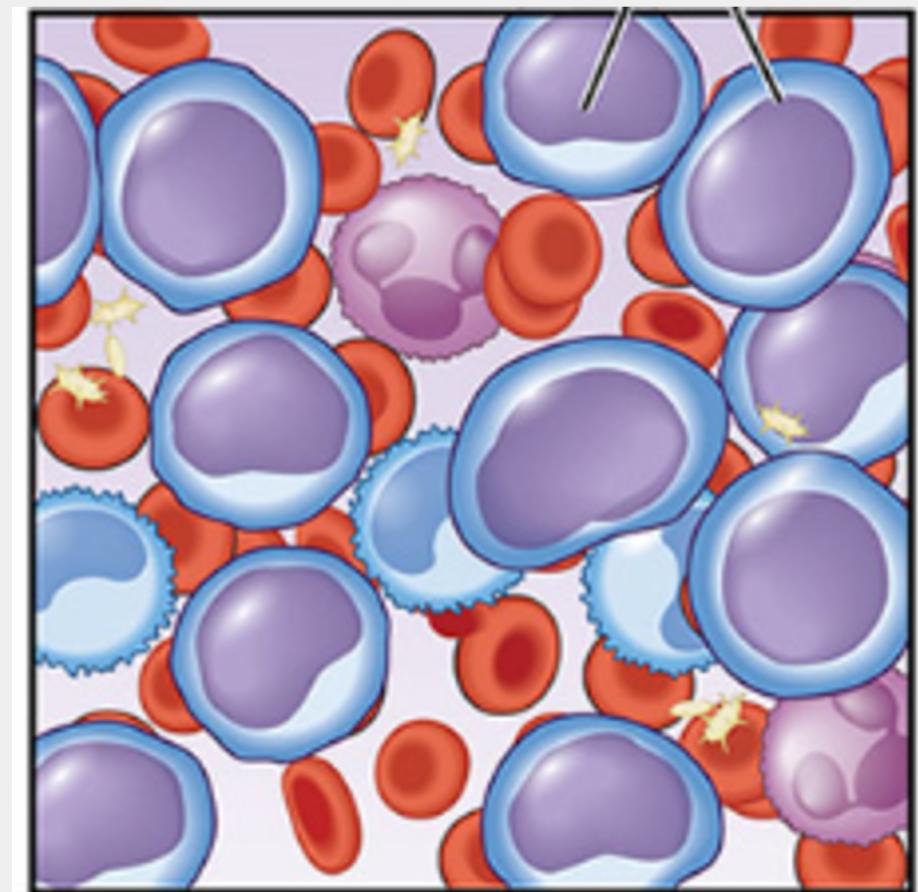
- Genetic mutations and chromosomal abnormalities.
- Environmental factors like radiation exposure and chemicals.
- Symptoms:
- Fatigue, frequent infections, bruising, bleeding, weight loss, and swollen lymph nodes.

Treatment:

- Chemotherapy, radiation therapy, targeted therapy, stem cell transplantation, and supportive care.

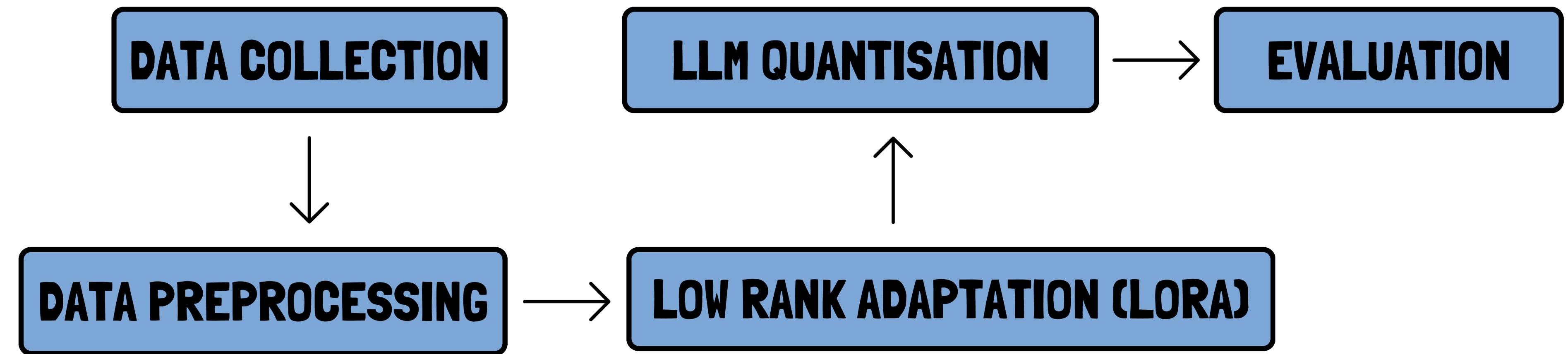


Normal blood



Leukemia

METHODOLOGY



In the process of fine-tuning our LLaMA 3.1 model, the methodology involves collecting and preprocessing leukemia-related data, leveraging LoRA for efficient adaptation to leukemia-specific tasks, applying LLaMA quantization to optimize performance in resource-constrained environments, and ultimately evaluating the model's effectiveness in leukemia-related predictions and analyses.

Data Collection

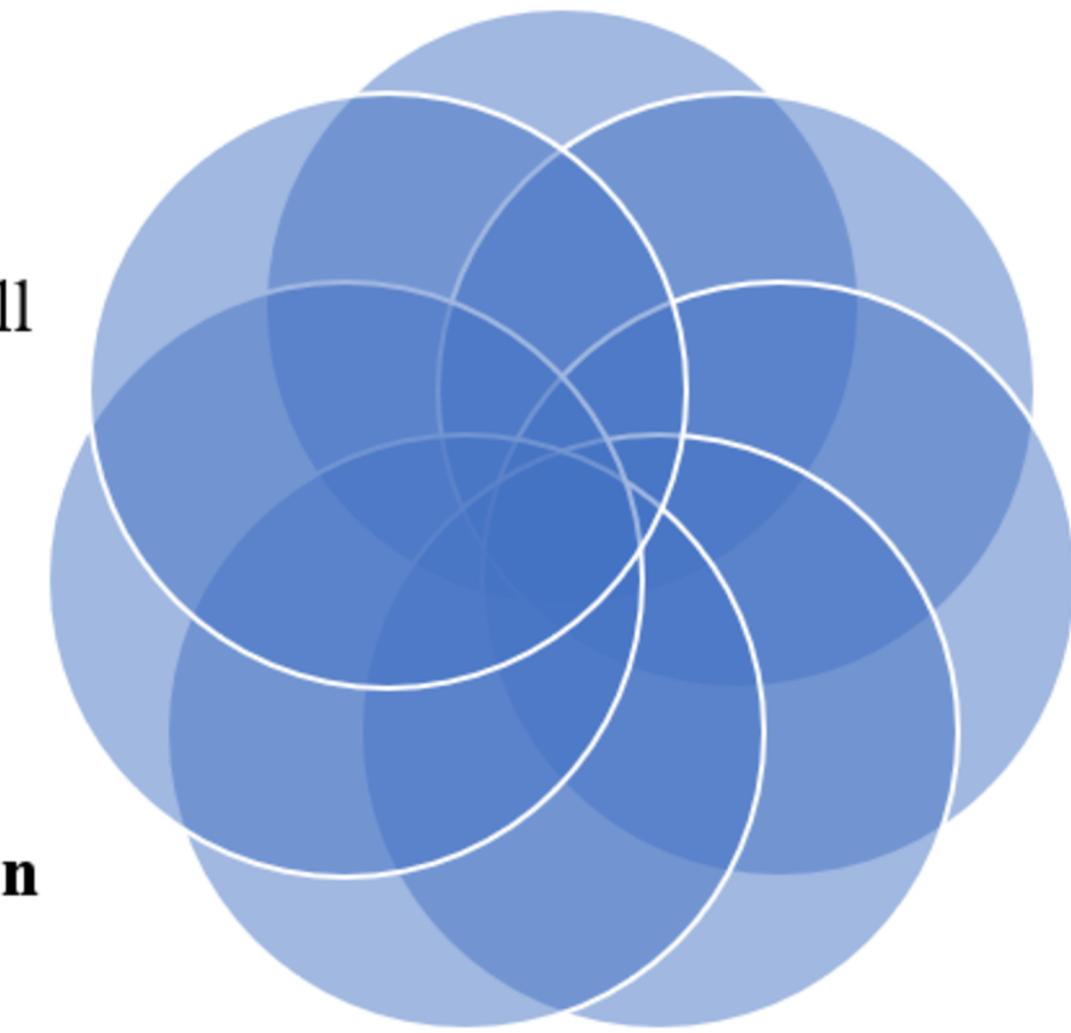
- Leukemia related PDFs
- Hugging face: Title and Abstract format

Evaluation

- Accuracy
- Precision and Recall
- Medical Relevance
- Validation

LLM Quantization

32 –bit to 8-bit



Data Preprocessing

- Data Cleaning
- Normalization
- Tokenization
- Annotation

Low-Rank Adaptation (LoRA)

- Low-rank updates to the LLM's weights

LoRA

- LoRA (Low-Rank Adaptation) is a technique that fine-tunes large language models by using low-rank matrices to approximate the model's original weight matrices, significantly reducing the number of parameters that need updating. The process involves initializing and integrating these low-rank layers into a pretrained model, then fine-tuning the model on specific data, like leukemia-related content.
- After fine-tuning, the model's performance is evaluated, and hyperparameters related to LoRA may be adjusted to optimize results. Finally, the adapted model is deployed, with LoRA making it efficient enough for use in resource-limited environments.

Nutritional Assistant

Inputs:

- Patient-specific details: age, weight, dietary restrictions, and preferred foods.
- Symptoms: anemia, unintended weight loss, and fatigue.

Process:

- Collect user inputs through an interactive form.
- Use LLAMA3.1 LLM to analyze nutritional needs and align with patient goals.

Output:

- Generates a tailored nutritional guide to improve energy levels, boost immunity, and manage chemotherapy side effects like nausea or appetite loss.
- Includes personalized food recommendations based on restrictions and preferences.

Diagnose Assistant

Input:

- Users upload lab report images in formats like PNG, JPG, or JPEG.

Process:

- Extract text from the uploaded images using EasyOCR for accurate text recognition.
- Analyze the extracted text using LLAMA to interpret medical information.

Output:

- Delivers a preliminary diagnosis for leukemia or related conditions.
- Provides actionable next-step recommendations for further testing or treatment.

LITERATURE REVIEW

S.No.	Authors and Year	Title	Observation
1	Biao Zhang + Zhongtao Liu + Colin Cherry + Orhan Firat (2024)	When Scaling Meets LLM Finetuning: The Effect Of Data, Model And Finetuning Method	examining the intersection of scaling large language models and fine-tuning methods, focusing on how model size, data, and fine-tuning techniques influence outcomes. Unlike typical studies that focus on one variable, this paper provides a holistic analysis of these factors together.
2	AMANDA S. DAVIS, MD, Anderson, ANTHONY J. VIERA, MONICA D. MEAD, MD (2014)	Leukemia: An Overview for Primary Care	a concise yet comprehensive guide for primary care physicians on recognizing, diagnosing, and managing leukemia. It distinguishes itself by offering practical, clinical insights specifically tailored to primary care settings, making it particularly useful for early detection and referral processes
3	Giulio Genovese, Ph.D., Anna K. Kähler, Ph.D., Robert E. Handsaker, B.S., Johan Lindberg, Ph.D., Samuel A. Rose (2014)	Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence	In this study, clonal hematopoiesis with somatic mutations was found in 10% of otherwise healthy people older than 65. The risk of hematologic cancer was substantially increased among these persons; in two cases, the subsequent cancer was related to the clone that predated the cancer

FUTURE SCOPE

Expand to Other Cancers:

- Extend the assistant's capabilities to address other types of cancer, providing personalized guidance and diagnosis support across a broader range of conditions.

Voice Assistance and Multilingual Support:

- Integrate voice interaction for improved accessibility and add support for multiple languages to reach diverse user populations.

Enhanced Diagnosis and Insights:

- Improve diagnostic accuracy with larger datasets and integrate predictive analytics for early detection and treatment optimization.