HepaBot: AI-Based Voice Assistant for Patient-Doctor Conversations in Liver Diseases

I. Business Problem

In the modern healthcare landscape, doctors face significant challenges in documenting patient interactions while maintaining an efficient workflow. Manual notetaking is time-consuming, prone to errors, and often reduces direct engagement with patients. Additionally, many critical details from conversations may be lost due to lack of proper documentation, leading to suboptimal patient care.

To address these issues, we propose an AI-based voice assistant that records patient-doctor conversations, performs disease prediction based on the discussion, and generates two reports: a transcription summary for patient records and

an insightful report for doctors for future references.

This solution will improve clinical efficiency, enhance patient outcomes, and streamline medical documentation. Our focus will be on liver diseases, utilizing data from official hepatic disease guidelines available online and patient data from the Pakistan Kidney and Liver Institute and Research Center (PKLI&RC).

II. Introduction

Artificial Intelligence (AI) and Large Language Models (LLMs) have revolutionized healthcare by automating documentation, enhancing decision-making, and improving efficiency. Our project aims to develop an AI-powered voice assistant that listens to patient-doctor conversations, extracts key medical insights, predicts possible liver diseases, and generates structured reports. By using advanced natural language processing (NLP) and machine learning techniques, this solution will enhance diagnostic accuracy and facilitate seamless medical documentation.

III. Project Modules

a. Module 1: Audio Data Acquisition

The project begins with audio data acquisition, where high-quality recordings of clinical interactions between doctors and patients are collected. This stage ensures clear voice capture in controlled clinical settings using appropriate recording equipment. Ethical guidelines and regulatory standards will be followed, ensuring informed consent from all participants and secure data storage.

b. Module 2: Transcription Process and Speaker Diarization

Once the audio data is acquired, it will be converted into text using A**utomated Speech Recognition (ASR)** systems such as **OpenAI Whisper or Wav2Vec 2.0,** fine-tuned on medical dialogues. These models, avail self-supervised learning, and will accurately transcribe complex medical terminology, particularly related to liver diseases.

There are two possible ways to separate doctor and patient speech. First, **unsupervised speaker diarization** automatically splits the recording into two distinct speaker tracks by clustering voice characteristics—no prior voice samples needed. Second, an **LLM analyzes each transcript segment's** content (medical terminology, question‑answer patterns) to assign labels of "Doctor" or "Patient." Combining both ensures accurate speaker separation and meaningful role identification.

For enhanced accuracy, **Whisper (fine-tuned for medical ASR)** and **Wav2Vec 2.0 with domain adaptation** will be integrated, ensuring efficient transcription even in noisy environments. Additionally, **LLaMA 2, fine-tuned on Aboonaji Wiki medical terms**, will refine the transcribed text and improve medical terminology recognition. Parameter-efficient fine-tuning (PEFT) and 4-bit quantization (AutoGPTQ, bitsandbytes) will enable LLaMA 2 to run efficiently while maintaining high accuracy with a low memory overhead.

c. Module 3: Text Processing, Feature Extraction, and Disease Prediction

Following transcription and speaker diarization, the text will be cleaned and normalized. This includes removing filler words, correcting errors, and ensuring standardized formatting for a precise conversation record. Natural Language Processing (NLP) techniques such as rule-based text cleaning, regex-based filtering, and spell correction (Levenshtein Distance or contextual embeddings) will be applied to improve accuracy.

The refined text will be converted into numerical representations using **TF-IDF, Word2Vec, FastText**, or transformer-based embeddings like B**ioBERT or ClinicalBERT** to capture domain-specific medical phrases. Additionally, LLaMA 2 embeddings, fine-tuned on medical data, will be leveraged to enhance feature extraction.

d. Module 4: Disease Prediction

This module will utilize machine learning and deep learning models to analyze extracted features and predict potential liver disease outcomes. By analyzing structured clinical conversations, these models will estimate the likelihood of specific conditions and generate data-driven diagnostic insights. **Naïve Bayes** approach will most probably be applied here.

For disease classification, transformer-based models with domain-specific adaptation will recognize complex medical patterns. Additionally, few-shot and prompt-based learning will allow the system to adapt to evolving medical knowledge without requiring extensive retraining.

The system will correlate key medical indicators with known liver disease symptoms to improve diagnostic precision. To ensure transparency in model predictions, **Explainable AI (XAI)** techniques will be incorporated, enabling clinicians to interpret and validate recommendations with confidence.

e. Module 5: Reporting and Final Diagnosis

The final module will generate structured and concise reports based on the disease prediction results. **LLaMA 2**, fine-tuned on medical terminology and hepatic disease guidelines, will be used for extractive and abstractive summarization, ensuring accurate and clinically relevant documentation.

The system will summarize patient-doctor conversations, highlighting critical symptoms, diagnostic cues, and predicted disease outcomes to support clinical decision-making.

To optimize efficiency, **dynamic quantization** techniques such as bitsandbytes (4-bit/8-bit quantization) and GPTQ (GPT Quantization) will be applied to reduce memory usage while maintaining high-quality medical text generation. FP16 mixed precision will further optimize inference speed without compromising accuracy.

The system will produce structured reports tailored for physicians and patient records, streamlining medical documentation and ensuring actionable insights.

f. Module 6: Multi-Agent System for Predictive Questioning (Additional Feature)

This module integrates LLaMA 2 with **LangChain** to predict follow-up medical questions based on doctor-patient conversations. A Conversation Context Agent tracks dialogue history, while a Medical Question Generator Agent formulates relevant follow-up queries using prompt engineering and few-shot learning. LangChain memory ensures context retention, enabling adaptive questioning based on prior exchanges. Optimized with 4-bit quantization (bitsandbytes, AutoGPTQ) for efficient execution on Google Colab T4, this system enhances diagnostic accuracy by proactively identifying missing details in medical discussions.

Performance Metrics

To evaluate HepaBot's accuracy, reliability, and efficiency, we will measure the following for each stage:

Speech Recognition & Speaker Diarization

Word Error Rate (WER) — Percentage of transcription errors versus a human-verified transcript; automated evaluation.

Diarization Error Rate (DER) — Percentage of audio time incorrectly attributed to the wrong speaker; automated comparison against annotated speaker labels.

Disease Prediction

Accuracy — Proportion of correctly classified liver disease outcomes; computed automatically against ground-truth diagnoses.

Precision, Recall & F1-Score — Balance false positives and false negatives to ensure clinical safety; generated via standard classification reports.

AUROC (Area Under ROC Curve) — Ability to distinguish between different disease classes; evaluated automatically.

Reporting & Summarization

LLM-Generated Summary Quality Score — A 0–1 score (target ≥0.85) using a prompt-based rubric (accuracy, completeness, coherence) judged by a fine-tuned LLaMA 2 model.

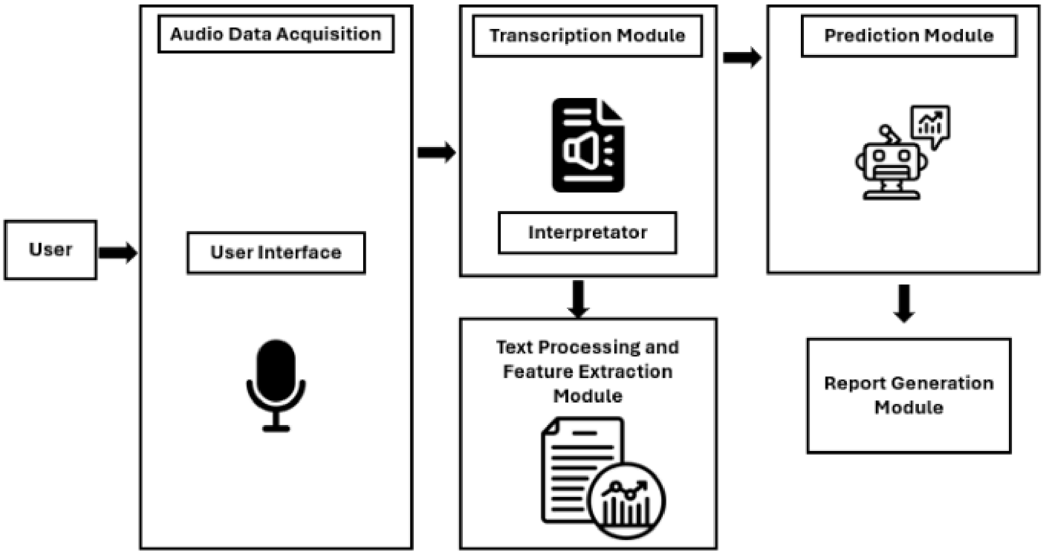
In-App Human Feedback Rating — Immediate 1–5 physician survey after report review measuring clarity, usefulness, and trustworthiness (target ≥4.5).

System Performance & Efficiency

Inference Time — Total latency from audio input to final report generation (target <10 seconds); measured via automated timing logs.

• **Memory & Compute Efficiency** — GPU/CPU usage and peak RAM consumption when running quantized models (4-bit LLaMA 2); profiled automatically.

a. Architecture Diagram



IV. Technical Features

- **Real-time Speech Recognition:** Utilization of state-of-the-art ASR models to transcribe doctor-patient interactions accurately.
- **Disease Prediction Engine:** Implementation of transformer-based models to extract symptoms and infer potential liver diseases.
- **Automated Report Generation:** AI-generated reports tailored for both patients and doctors.
- **User-Friendly Dashboard:** Web and mobile application for report access and review.
- **Data Security & Compliance:** Encryption protocols and HIPAA-compliant storage for sensitive medical data.

V. Timeline

The timeline for this project is given below:

Phase	Duration	Description
Phase 1	Week 1	Data Collection & Annotation
Phase 2	Week 2	Speech-to-Text Model Development
Phase 3	Week 3	Disease Prediction Model Implementation
Phase 4	Week 4	Report Generation & Summarization Module
Phase 5	Week 5	System Integration, Testing & Deployment

VI. Unique Value Proposition

- **Timesaving:** Reduces the burden of manual documentation, allowing doctors to focus more on patient care.
- **Enhanced Diagnosis:** AI-powered disease prediction aids in early detection of liver diseases.
- **Automated Reports:** Provides structured and actionable insights for both doctors and patients.
- **Improved Compliance:** Ensures proper documentation for medical and legal purposes.

VII. Future Aspects

- **Integration with EHR Systems:** Synchronization with electronic health records for efficient record-keeping.
- **Multilingual Support:** Expansion to support multiple languages for global adoption.
- **Personalized AI Recommendations:** Advanced AI-driven treatment recommendations based on patient history.
- **Voice Biometrics:** Incorporation of voice-based patient identification for enhanced security.

By implementing this AI-based voice assistant, we aim to revolutionize clinical workflows, enhance diagnostic accuracy in liver diseases, and improve patient care. This project will not only benefit doctors and hospitals but also contribute to the evolution of AI in healthcare.

References

1. IEEE Xplore, “Document 6135543,” [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6135543. [Accessed: Mar. 10, 2025].
2. M. Steinbach, G. Karypis, and V. Kumar, “A Comparison of Document Clustering Techniques,” 2000.
3. K. Lee, M. Chang, and J. Kim, “Fine-Tuning Language Models for Clinical Summarization,” arXiv preprint arXiv:1906.02239, 2019.

Submission By:
Anoosha Tahir 24280077
Rabia Aslam 24280019
Iqra Muzaffar 24280026
Hasnain Sohail 24280029
Ali Arsalan 24280048