International Journal on Science and Technology (IJSAT)
E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

Yashoda Technical Campus, Whadhe, Satara Healthcare Using AI-Voice & Vision

Mr. Dineshkumar Yadhav[1], Atharva Sonawane[2], Sanket Pawar[3], Pranav Palve[4], Piyush Jethwa [5], Karan M Patil[6]

[1]Mentor, [2,3,4,5,6]Student
[1,2,3,4,5,6]Yashoda Technical Campus, Satara

Abstract:

This project aims to develop an AI-powered healthcare assistant capable of diagnosing medical conditions based on **text**, **voice**, and **image inputs**, while also providing **doctor suggestions** and **medication recommendations** through both **text and speech**. It leverages machine learning, natural language processing (NLP), and computer vision to provide an accessible, interactive, and reliable healthcare support system. The system aims to reduce the burden on healthcare facilities and provide preliminary assistance to patients remotely.

Keywords: AI in healthcare, voice diagnosis, image-based diagnosis, NLP, doctor recommendation system, machine learning.

1. Introduction:

This project, **"Healthcare using AI,"** focuses on developing a smart diagnostic assistant that leverages multiple AI technologies to deliver healthcare recommendations through various user-friendly input modes—**text**, **voice**, and **image**. Users can interact with the system by typing symptoms, speaking naturally, or uploading relevant images (e.g., skin conditions), and receive probable diagnoses, doctor recommendations, and medication suggestions. The system also responds through text or speech, making it highly interactive and accessible even for those with limited literacy or visual impairments.

The core objective of this project is to build a **multi-modal healthcare support system** that mimics basic functions of a general physician by understanding and analyzing symptoms and suggesting relevant medical action. This system is particularly beneficial in **rural or remote areas**, where access to doctors may be limited, and can also serve as a **first-line triage tool** in

This initiative also aligns with global health goals, such as those set by the **World Health Organization (WHO)**, promoting the use of digital technology to achieve **universal health coverage**. While not intended to replace doctors, the system is designed to act as a **support tool** that can offer guidance, reduce diagnostic delays, and assist users in making informed decisions about their health.

2. Literature Survey:

The integration of Artificial Intelligence in healthcare has gained significant momentum in recent years. Multiple studies and real-world systems have explored the capabilities of AI to assist in medical diagnosis, patient communication, and healthcare decision-making. This section reviews existing technologies and research trends relevant to the components of this project, namely **text-based diagnosis**, **voice interaction**, **image-based diagnosis**, and **AI-driven recommendations**.

1. Text-Based Diagnosis Using NLP:

Natural Language Processing (NLP) has been widely used to interpret and analyze textual data in healthcare. Several AI systems, such as **Ada Health** and **Symptomate**, allow users to input symptoms in natural language and receive likely diagnoses.

- **Research by Esteva et al. (2019)** emphasized the use of NLP models trained on large medical datasets to map user-provided symptoms to disease categories.
- **Transformer-based models**, such as BERT and GPT, have revolutionized NLP in healthcare by understanding context better and improving the accuracy of symptom interpretation.
- Datasets like **MedQuAD** and **MediQA** provide annotated question-answer pairs in the healthcare domain to train models for medical query understanding.

2. Voice-Based Interaction:

Voice interaction enhances accessibility, particularly for users who are visually impaired or unfamiliar with typing.

- **Google's Medical Speech-to-Text API** and **Mozilla DeepSpeech** are two prominent tools enabling accurate voice-to-text conversion for healthcare applications.
- **WHO's digital health guidelines (2020)** stress the importance of voice-based technologies in low-resource settings.

Many digital health platforms like **Babylon Health** and **Suki AI** use conversational AI to facilitate doctor-patient interactions. These systems use **ASR (Automatic Speech Recognition)** combined with NLP to interpret spoken symptoms and convert them into structured data.

Our project integrates similar techniques to allow voice-based reporting of symptoms and vocal delivery of recommendations using Text-to-Speech (TTS).

3. Image-Based Diagnosis Using Computer Vision:

AI-based image analysis plays a critical role in diagnosing conditions like skin diseases, fractures, and respiratory issues through X-rays or photos.

- **Dermatologist-level classification** of skin cancer using CNNs (Esteva et al., Nature, 2017) marked a major breakthrough in image-based diagnostics.
- **MedMNIST**, **ISIC**, and **ChestX-ray14** are widely used datasets for training deep learning models for medical image classification.

Recent studies have shown that **transfer learning** with pretrained models like **ResNet**, **VGG16**, and **Inception** provides high accuracy even with limited medical image datasets.

Our system applies similar techniques to classify diseases from skin-related conditions using CNNs trained on filtered, labeled medical image datasets.

4. AI for Doctor and Medicine Recommendation:

Recommendation systems in healthcare involve complex decision-making and require an understanding of both symptoms and medical databases.

- **IBM Watson for Health** and **HealthTap** use AI-driven engines to suggest treatment options and direct users to specialists based on their input.
- Algorithms like **Decision Trees**, **Support Vector Machines (SVM)**, and **KNN classifiers** have been employed for medical decision support systems.

Our project applies **rule-based logic** integrated with ML models to recommend relevant doctors based on location, specialization, and symptom severity. Medicine recommendations are drawn from a curated database of common over-the-counter (OTC) treatments for initial relief.

5. Multi-Modal Healthcare Systems:

Few systems combine all three input modes—text, voice, and image—in a unified diagnostic tool. Some experimental platforms and research prototypes are beginning to explore this area.

- **MedWhat** and **Infermedica** offer multi-channel input capabilities but often rely on proprietary models and databases.
- **Academic studies (e.g., ACM Digital Library, 2020)** highlight the growing interest in combining voice and visual analysis for more accurate diagnoses.

Our project is unique in its approach to providing **a multi-modal interface** for healthcare advice, making diagnosis more flexible and user-centric.

3. Methodology:

1. **Data Collection & Preprocessing:**
    - Text data: Cleaned and tokenized symptom-disease mappings.
    - Voice data: Converted to text using speech recognition.
    - Image data: Resized, normalized, and augmented for training.

2. **Model Architecture:**
    - **Text Input:** Processed using NLP pipelines → Classification using logistic regression or transformer-based models.
    - **Voice Input:** Speech → Text → Same NLP pipeline.
    - **Image Input:** CNN model to classify disease types.
    - **Recommendation Engine:** Suggests nearby doctors (using location data) and common medicines from a database.
3. **Technologies Used:**
    - **Frontend:** HTML, CSS, JavaScript (for text and voice input interface)
    - **Backend:** Python (Flask/FastAPI), TensorFlow/PyTorch for ML models
    - **Database:** MongoDB/MySQL for storing symptoms, diseases, doctors, and medicine details.
    - **APIs:** Google Speech-to-Text, Text-to-Speech (TTS), and image recognition models.

Metrology:

To evaluate the system's effectiveness:

- **Diagnosis Accuracy (Text/Voice):** % of correct disease predictions based on symptoms
- **Image Model Accuracy:** CNN model classification accuracy on test set
- **Response Time:** Time taken for voice-to-text conversion and model inference
- **User Satisfaction:** Measured via mock user feedback
- **Speech-to-Text Accuracy:** Word Error Rate (WER)

Implementation:

1. **Symptom to Disease (Text):**
    - Input: "I have fever and body ache"
    - Output: Diagnosis: Dengue (90% confidence)
2. **Voice Diagnosis:**
    - Converts: "My throat hurts and I have fever" → Same pipeline as text.
3. **Image Diagnosis:**
    - Input: Rash image → Output: Possible Eczema (82% confidence)
4. **Doctor/Medicine Suggestion:**
    - Based on disease severity and type → Suggests specialist + OTC medication

4. Conclusion:

This AI-powered healthcare assistant demonstrates how multi-modal interfaces (text, voice, image) can improve accessibility and efficiency in medical diagnostics. It is especially useful in areas with limited healthcare access, offering preliminary support and direction. While not a replacement for medical professionals, this system can serve as a crucial first point of contact.

References:

1. Ghosh, R., et al. (2020). "AI in Healthcare: Diagnosis, Monitoring, and Therapy".
2. Vaswani et al. (2017). “Attention is All You Need” – Transformers in NLP
3. Simonyan, K., & Zisserman, A. (2014). "Very Deep Convolutional Networks for Large-Scale Image Recognition".
4. Google Speech Recognition API Documentation
5. https://www.tensorflow.org/
6. https://huggingface.co/transformers/
7. https://developer.mozilla.org/en-US/docs/Web/API/Web_Speech_API