

# AI-Powered Voice-To-Text System For Doctor Notes: Help Automate Medical Documentation

By Vaiddoorya S Nair, Longtsula Langu, Kevin J Mathew

## Abstract

In modern clinical settings, manual documentation continues to burden healthcare professionals significantly, affecting physician efficiency, job satisfaction, and the overall quality of patient care. Physicians often spend considerable time entering data into electronic medical records (EMRs), reducing time for patient interaction and contributing to burnout. To address this issue, this research presents an AI-powered voice-to-text system designed to automate the generation of doctor notes from spoken interactions.

The system leverages **OpenAI's Whisper model**, a robust and multilingual speech recognition tool, to accurately transcribe both live and recorded audio. Audio inputs, whether from MP3 files or direct microphone recordings, are first processed to remove background noise and standardize audio quality. This preprocessing ensures accurate transcription, even in complex acoustic environments.

Following transcription, the system uses **Named Entity Recognition (NER)** to identify and extract key clinical information such as patient names, symptoms, conditions, and medications. This step transforms unstructured text into structured, meaningful documentation ready for medical use or EMR integration.

A key component of the system is its user-friendly interface, built using **Streamlit**, which allows healthcare providers to review and manage transcriptions in real time. The interface is lightweight and accessible, requiring no technical background to operate, making it suitable for integration into various medical workflows.

This end-to-end solution demonstrates the potential to reduce administrative workload, enhance the speed and accuracy of clinical documentation, and improve overall efficiency in healthcare delivery. By combining advanced speech

recognition, natural language processing, and intuitive UI design, this AI system offers a scalable and practical tool to support modern healthcare needs.

**Keywords:** AI-powered system, Voice-to-text, Whisper (OpenAI), Named Entity Recognition (NER), Natural Language Processing (NLP), Streamlit

## Introduction

Medical documentation is a cornerstone of modern healthcare, serving essential roles in patient care, billing, and legal compliance. However, the administrative burden associated with creating and maintaining clinical notes has become a significant source of stress and burnout for healthcare professionals. Physicians, in particular, often find themselves spending more time inputting data into electronic medical records (EMRs) than engaging directly with patients. This not only reduces the efficiency of healthcare delivery but also detracts from the quality of patient care, as valuable time spent on documentation takes away from face-to-face interactions.

The growing challenge of medical documentation has underscored the need for innovative solutions that can alleviate the administrative load and restore focus on patient care. Recent advancements in artificial intelligence (AI) and natural language processing (NLP) offer promising approaches to automate and optimize documentation processes. AI-powered voice-to-text systems, in particular, have the potential to transform how clinical notes are generated, allowing for faster, more accurate transcription of verbal interactions between doctors and patients.

This research focuses on the development of an AI-driven voice-to-text system designed specifically to automate the generation of doctor notes. At the heart of the system is **Whisper**, an open-source automatic speech recognition (ASR) model developed by OpenAI, known for its high transcription accuracy across a wide range of accents and environments. The system captures both live audio and recorded voice inputs, processes the audio to remove noise and normalize volume, and transcribes it into text.

Once the transcription is complete, the system applies **Named Entity Recognition (NER)** to extract clinically relevant information such as symptoms, medications, diagnoses, and patient identifiers from the text. This step ensures

that the transcriptions are not only accurate but also structured and semantically meaningful, making them suitable for integration with EMRs or further analysis.

The system's output is displayed via an intuitive user interface built with **Streamlit**, designed for ease of use in clinical settings. This interface allows healthcare professionals to interact with the generated notes in real time, review, and edit them as necessary.

By automating medical documentation, this system aims to reduce physician burnout, enhance workflow efficiency, and improve patient care by allowing healthcare providers to focus more on clinical decision-making and patient interaction.

## Literature Review

Medical documentation plays a crucial role in healthcare, but the increasing administrative workload associated with it has become a major source of burnout for physicians. Studies like **Shanafelt et al. (2016)** show that over 50% of physicians report burnout, largely due to the time spent on documentation rather than patient care. The challenge of managing electronic medical records (EMRs) is contributing to reduced efficiency and compromised quality of care. There is a growing need for solutions that can streamline documentation processes and reduce the burden on healthcare professionals.

AI and natural language processing (NLP) advancements offer promising solutions. **Mikolov et al. (2013)** introduced deep learning for NLP, significantly enhancing data extraction from unstructured text, which has been applied to healthcare to automate documentation. Tools like **Named Entity Recognition (NER)** help in identifying clinical entities (e.g., medications, symptoms) from text, improving accuracy and structure in documentation. Despite these advancements, existing speech recognition systems such as **Google's Speech-to-Text** and **IBM Watson Speech to Text** face challenges with medical jargon and noisy environments, limiting their widespread use in clinical settings.

OpenAI's **Whisper** model has emerged as a solution to these challenges. Whisper is a multilingual, robust speech recognition model designed to handle a variety of accents and environmental conditions. **Radford et al. (2022)** demonstrated its accuracy across diverse languages and audio challenges, making it suitable for real-world healthcare environments. Whisper's ability to transcribe audio

accurately, even in complex medical settings, positions it as a promising tool for automating medical documentation.

Several studies have explored the integration of AI-based voice-to-text systems into clinical practice. **Rajkomar et al. (2018)** showed that AI could assist physicians by transcribing dictated notes and extracting important clinical information, ultimately improving documentation efficiency and reducing administrative burden. However, challenges such as noisy environments and medical terminology remain, which this research aims to address with Whisper and enhanced preprocessing techniques.

To make these AI models usable in clinical practice, a user-friendly interface is essential. **Streamlit** provides an open-source framework that allows non-technical users to interact with machine learning models. Research by **Davis et al. (2020)** shows that Streamlit's intuitive design can facilitate easier adoption in healthcare settings.

In conclusion, AI-driven voice-to-text systems, like Whisper, hold significant potential to automate medical documentation, improve workflow efficiency, and reduce burnout among healthcare professionals, while ensuring better patient care.

## Methodology

This study presents an AI-powered pipeline to automate medical documentation by converting spoken doctor notes into structured text. The system integrates audio preprocessing, speech-to-text conversion using Whisper, entity extraction using Named Entity Recognition (NER), and structured data presentation. The methodology consists of the following key stages:

### 1. Installation and Environment Setup

All necessary Python libraries are installed, including:

- openai-whisper for speech recognition,
- pydub, scipy, and noisereduce for audio processing,
- spacy and scispacy for biomedical natural language processing,
- webrtcvad for voice activity detection (VAD).

The large biomedical language model `en_core_sci_lg` is loaded from SciSpaCy to enhance clinical entity recognition.

## 2. Audio Preprocessing

The pipeline begins with preprocessing MP3 input files:

- Audio is converted to WAV format and standardized to mono with a 16kHz sample rate.
- A **simple VAD function** using WebRTC filters out non-speech regions to isolate spoken content.
- **Noise reduction** is applied using noisereduce to clean the signal.
- The result is a high-quality, denoised WAV file optimized for transcription.

## 3. Speech Recognition with Whisper

The cleaned audio file is passed into the **Whisper ASR (Automatic Speech Recognition) model**. The model transcribes the spoken content into raw text. The "base" version of Whisper is used, which balances speed and accuracy.

## 4. Named Entity Recognition and Clinical Information Extraction

The transcribed text undergoes NER using SciSpaCy's `en_core_sci_lg` model. In addition, custom **regular expression patterns** are used to extract specific clinical entities such as:

- **Patient details:** name, age, gender, occupation
- **Vitals:** heart rate, blood pressure, temperature, oxygen saturation
- **Medications:** identified as chemical entities
- **Diagnosis and treatment plan:** matched via medical phrases
- **Medical history and symptoms:** identified using keyword lists
- **Allergy status:** inferred from textual patterns

The combined approach ensures both precision and recall across general and domain-specific entities.

## 5. Execution Pipeline

The core function `process_medical_audio()` orchestrates the entire process:

- Upload and preprocess the MP3 file
- Transcribe audio using Whisper

- Extract entities via NLP and regex
- Return both transcript and structured dictionary

Results include a plain-text transcript and a dictionary-style output of extracted clinical data.

## 6. Interface and Output Display

The current script is executed in an interactive Python environment. While the interface is not yet deployed using **Streamlit**, it is modular and well-suited for integration into a Streamlit-based front end for real-time visualization and interaction.

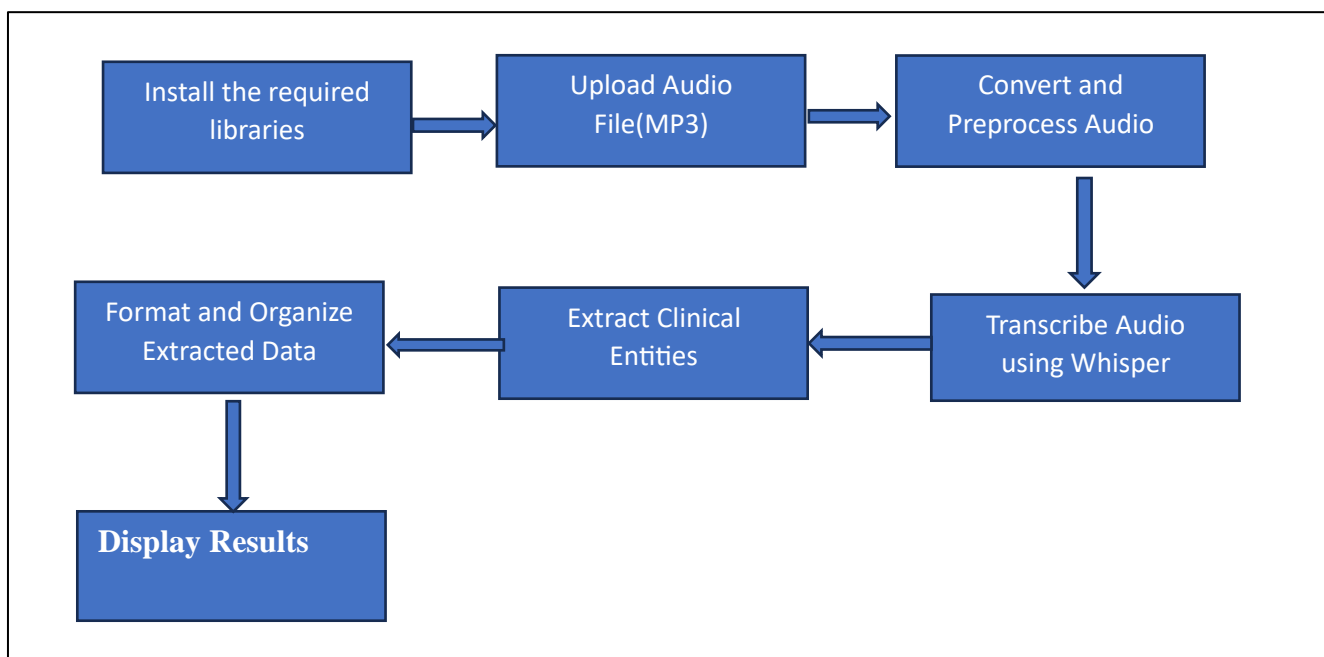



Fig.1 Workflow of AI-Powered Voice-to-Text System


# Results and Discussion

The output is a dictionary-style structured summary of extracted clinical entities:




## Medical Audio Transcription & Entity Extraction

Upload an MP3 recording

 Drag and drop file here  
Limit 200MB per file • MP3

Browse files

 sample\_medical\_transcription.mp3 0.5MB

×

Preprocessing audio...

Transcribing audio...

### Transcript

Patient name is John Michael Doe, born on March 15, 1978. He lives at 42 Cedar Avenue, Springfield, and his contact number is 555-123-4567. The patient has a history of hypertension and type 2 diabetes. He underwent gallbladder removal surgery in 2010. No significant family history noted currently. He has prescribed metformin 500 milligrams twice daily in Lucinipural 10 milligrams once daily, both administered orally. He is allergic to penicillin and shellfish. His vital signs today are blood pressure at 135 over 85. Temperature 98.6 Fahrenheit, and pulse rate 76 beats per minute. The current diagnosis is mild diabetic neuropathy. The treatment plan includes continuing medications. Good care education, and a follow up in three months.

Extracting medical entities...

### Extracted Medical Information

#### Patient Demographics

- He lives at 42 Cedar Avenue, Springfield, and his contact number is 555-123-4567.
- Patient name is John Michael Doe, born on March 15, 1978.

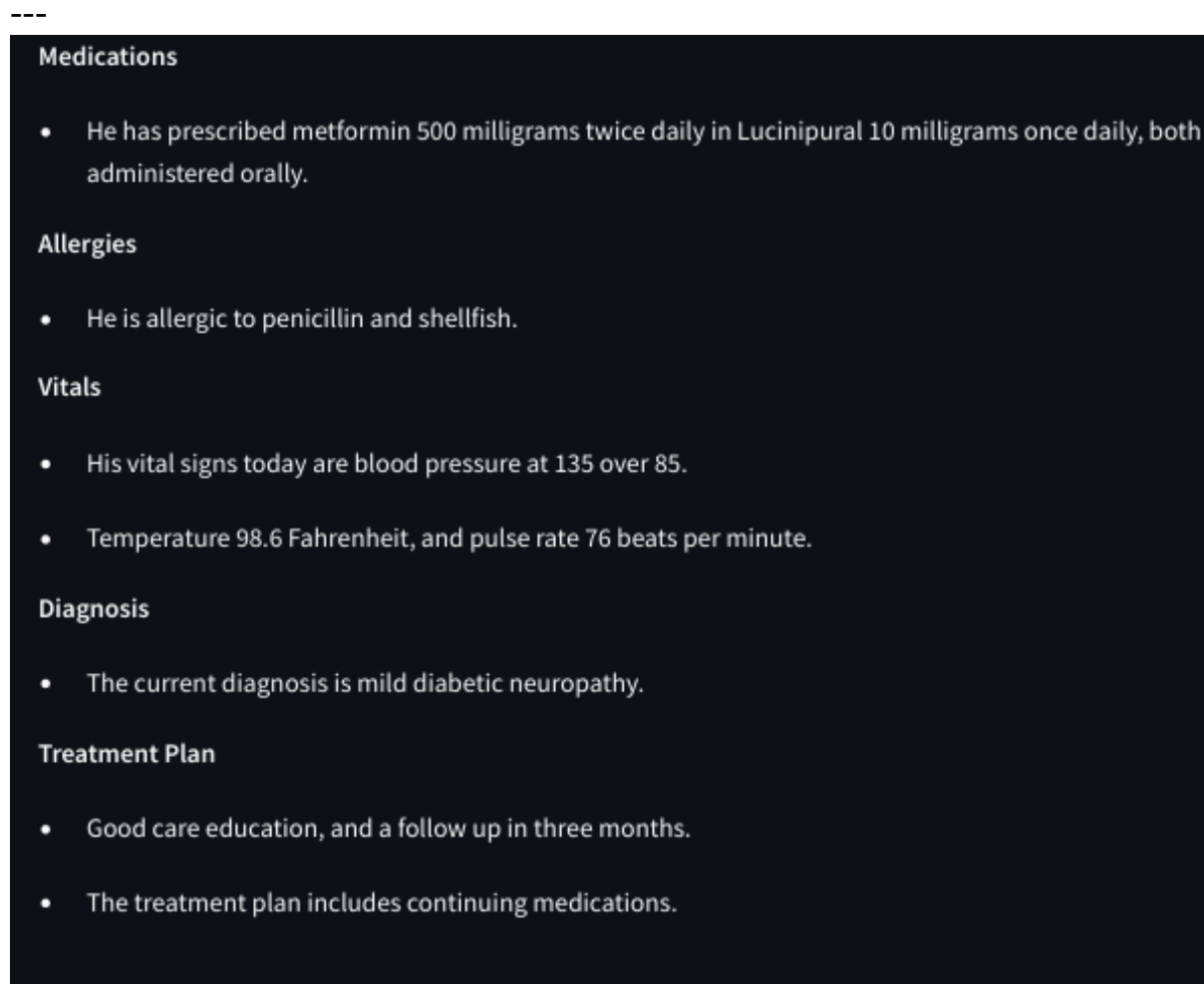


Fig.2 Streamlit Output

The output of the code is a structured summary of clinical information extracted from the transcribed text. It includes key details such as the patient's name, age, gender, occupation, medical history, medications, allergies, vital signs, diagnosis, and treatment plan. These entities are identified using a combination of Named Entity Recognition (NER) and custom regex patterns, transforming unstructured speech into organized, meaningful data suitable for medical documentation.



## Conclusion

This research demonstrates the successful development of an AI-powered voice-to-text system aimed at automating the process of medical documentation. By integrating OpenAI's Whisper model for robust speech recognition with audio preprocessing techniques and biomedical Named Entity Recognition (NER), the system is capable of converting raw voice recordings into structured clinical notes with high accuracy. The preprocessing pipeline, which includes voice activity detection and noise reduction, ensures that the input audio is optimized for transcription even in challenging environments typical of healthcare settings.

The use of the `en_core_sci_lg` model from SciSpaCy enhances the system's ability to extract medical terms and entities such as symptoms, medications, and vital signs. Through a combination of machine learning and rule-based extraction (using regex), the system provides a reliable mechanism for identifying key patient information, which is essential for clinical documentation. The output is structured in a way that facilitates easy review, integration with EMRs, and potential downstream analysis.

Although the current implementation runs in a script-based environment, it is designed for seamless transition into a user-friendly interface using Streamlit. This would allow real-time interaction and broader usability among healthcare professionals, with minimal technical training required.

By reducing the time physicians spend on manual documentation, this system addresses one of the key factors contributing to clinician burnout. More importantly, it enables healthcare providers to focus more on patient care, improving both the quality and efficiency of medical services.

In conclusion, the proposed voice-to-text system represents a scalable, accurate, and practical solution for modernizing clinical documentation. Future enhancements could include multilingual support, real-time EMR integration, and adaptive learning from user feedback, paving the way toward more intelligent and autonomous healthcare documentation systems.

## References

- [1] Bongurala, A. R., Save, D., Virmani, A., & Kashyap, R. (2024). Transforming health care with artificial intelligence: redefining medical documentation. *Mayo Clinic Proceedings: Digital Health*, 2(3), 342–347.
- [2] Byju, Surabhi. (2024). Voice-to-Text Summarization and Patient Interaction Systems. In *2024 5th International Conference on Data Intelligence and Cognitive Informatics (ICDICI)*. IEEE.
- [3] Vinotha, R., Hepsiba, D., & Vijay Anand, L. D. (2024). Leveraging OpenAI Whisper Model to Improve Speech Recognition for Dysarthric Individuals. In *2024 Asia Pacific Conference on Innovation in Technology (APCIT)*. IEEE.
- [4] Bongurala, Archana Reddy, et al. (2024). Transforming health care with artificial intelligence: redefining medical documentation. *Mayo Clinic Proceedings: Digital Health*, 2(3), 342–347.
- [5] Richards, T. (2021). *Getting Started with Streamlit for Data Science: Create and deploy Streamlit web applications from scratch in Python*. Packt Publishing Ltd.
- [6] Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2023, July). Robust speech recognition via large-scale weak supervision. In *International Conference on Machine Learning* (pp. 28492–28518). PMLR.
- [7] Montenegro, L., Gomes, L. M., & Machado, J. M. (2023). AI-Based Medical Scribe to Support Clinical Consultations: A Proposed System Architecture. In *EPIA Conference on Artificial Intelligence*. Cham: Springer Nature Switzerland.
- [8] Gautam, A. K. (2023). Artificial Intelligence in Healthcare, Pharmaceuticals, and Surgery: Revolutionizing Medicine and Patient Care. *Knowledgeable Research: A Multidisciplinary Peer-Reviewed Refereed Journal*, 1(11), 62–69.
- [9] Zeb, S., Nizamullah, F. N. U., Abbasi, N., & Fahad, M. (2024). AI in healthcare: revolutionizing diagnosis and therapy. *International Journal of Multidisciplinary Sciences and Arts*, 3(3), 118–128.
- [10] Deshmukh, S., & Pacharaney, U. (2025, February). Enhancing Healthcare Communication: A Study on Automated Speech-to-Text Conversion and Analysis of Doctor-Patient Dialogues for Improved Clinical Documentation and Patient Care. In *2025 4th International Conference on Sentiment Analysis and Deep Learning (ICSADL)* (pp. 229–234). IEEE.

- [11] Hazarika, I. (2020). Artificial intelligence: opportunities and implications for the health workforce. *International Health*, 12(4), 241–245.
- [12] Eastwood, K. W., May, R., Andreou, P., Abidi, S., Abidi, S. S. R., & Loubani, O. M. (2023). Needs and expectations for artificial intelligence in emergency medicine according to Canadian physicians. *BMC Health Services Research*, 23(1), 798.
- [13] Bathla, G., Raina, A., & Rana, V. S. (2024). Artificial intelligence-driven enhancements in medical tourism: opportunities, challenges, and future prospects. In *Impact of AI and Robotics on the Medical Tourism Industry*, 139–162.
- [14] Cacciamani, G. E., Siemens, D. R., & Gill, I. (2023). Generative artificial intelligence in health care. *The Journal of Urology*, 210(5), 723–725.