ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 13 Issue 05 May 2026

AI-Driven Real-Time Speech Summarization and EHR Processing for Enhanced Healthcare System

[1] Tekcham Rickysingh Ingraji*, [2] Atharv Sameer Shevade, [3] Subhag Sanjay Raut, [4] Vinayak Dadasaheb Suryawanshi, [5] Dr. Sharvari Govilkar

[1][2][3][4][5] Department of Computer Engineering, Pillai College of Engineering, India
Email: [1] ivyricked13@gmail.com, [2] atharvshevade786@gmail.com, [3] subhagraut111@gmail.com, [4] vinayaksuryawanshi49@gmail.com, [5] sgovilkar@mes.ac.in

Abstract— *Clinical documentation is a time-intensive process that contributes significantly to the workload of healthcare professionals and may hinder effective doctor–patient interaction. This paper presents an AI-driven framework for real-time summarization of doctor–patient conversations using Automatic Speech Recognition (ASR) and Large Language Models (LLMs). The proposed system extracts clinically relevant entities and generates structured summaries compatible with Electronic Health Record (EHR) systems. A domain-adapted language model is employed to enhance contextual understanding and improve the accuracy of medical entity recognition. Experimental results demonstrate that the system significantly improves documentation efficiency while preserving semantic integrity. The proposed approach highlights the potential of AI-assisted clinical documentation to enhance healthcare data quality and reduce administrative burden.*

Index Terms— *Artificial Intelligence, Automatic Speech Recognition, Clinical Documentation, Electronic Health Records, Large Language Models*

I. INTRODUCTION

Medical documentation is essential for accurate diagnosis, treatment planning, and continuity of care. However, traditional documentation methods remain largely manual, making them time-consuming, error-prone, and a significant source of administrative burden for healthcare professionals. Recent advancements in Artificial Intelligence (AI), particularly in Large Language Models (LLMs), Automatic Speech Recognition (ASR), and Natural Language Processing (NLP), enable new approaches for automating clinical documentation. These technologies facilitate real-time analysis of doctor–patient conversations, allowing for the extraction and structuring of clinically relevant information. This work presents an AI-driven framework for real-time speech summarization aimed at improving documentation efficiency, enhancing data accuracy, and reducing clinician workload.

II. LITERATURE REVIEW

Recent studies have explored the application of deep learning techniques and Large Language Models (LLMs) for real-time speech summarization in clinical domains. Approaches such as RTSS and LLM. Some have shown that these models are good at turning what doctors and patients say into neat, written notes, and they are also helpful because they can change messy talk into a structure that the computer understands better. Mixing together all the different kinds of notes, like the ones that are already organized and the ones that are just written down freely, has shown that it can help doctors make much better choices for their patients. However, there are still some big problems that the computers have to fix because the hospital computer paper systems, which are called EHRs, do not always talk to each other well, and sometimes the notes inside them are not very good or are just too messy for the computer to understand easily. Also, current systems have a very hard time understanding big and complicated medical words when people are talking fast, and on top of that, people are also very worried about keeping the doctor's data safe so that it stays private and no one else can see it.

III. PROBLEM STATEMENT

However, even after the use of Electronic Health Record (EHR) systems, certain problems persist. This is mainly related to the issues of interoperability, usability, and the quality of clinical documentation. Also, a large part of the information is unstructured and does not allow for decision support and analysis in an efficient manner. In addition, manual clinical documentation also leads to a high level of burnout. Even though speech recognition can be a very natural form for interacting with clinical documentation systems, the existing systems are not efficient enough to handle the noise and other factors related to the accent and terminology. In addition, certain issues related to the security and access of the data also exist, making the management of clinical data a complex task.

IV. RESEARCH OBJECTIVES

The objective goal of this research is to build a system that listens to doctor and patient talking. First, it turns the speech into text and then makes a short summary of that talk. Next, the system finds important medical words to put into a hospital computer paper (EHR) so a finished report is ready. This makes hospital work faster and helps keep the papers correct.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 13 Issue 05 May 2026

Another goal is to make the system find how medical things are related and answer questions. This helps the system understand hard medical data better by picking out the right information from the report.

Also, the system is made to understand many different ways that people talk (accents). It uses medical datasets to help it work well in different hospitals and with different people. This makes the system more reliable for everyone.

V. MOTIVATION AND RESEARCH GAPS

The cognitive and administrative burdens of manual clinical documentation are considerable for healthcare professionals, causing overall inefficiencies, errors, and burnout. This calls for the need for automation, using AI-based speech summarization and Large Language Models, for better efficiency, accuracy, and overall clinician-patient interactions.

Though considerable progress has been reported in recent times, there are several hurdles to be cleared. Clinical data is of both structured and unstructured nature, making it difficult to derive valuable insights from sources like communications between a doctor and a patient. The accuracy of currently available Automatic Speech Recognition systems is also questionable, especially when it comes to handling specific medical terms, abbreviations, and words that sound phonetically similar. Another problem is the overall environment of a clinical setup, which is usually noisy, causing considerable degradation of speech transcription. Another factor is linguistic, including accents and multilingual communications.

VI. SYSTEM SCOPE AND IMPACT

This work proposes a system for real-time clinical documentation through speech-based summarization. The proposed scope includes a clinical entity extraction method from doctor-patient dialogue, a language model for medical NLP tasks, and its integration with EHR systems through a deep learning-based speech recognition framework.

This proposed system will enhance the accuracy and efficiency of clinical documentation, assist in timely clinical decision-making, and reduce the workload of medical professionals. Additionally, it will allow structured data generation for medical research, analytics, and efficient utilization of healthcare resources.

VII. PROPOSED METHODOLOGY

The proposed system presents a real-time pipeline for converting doctor–patient conversations into structured Electronic Health Records (EHRs). It integrates speech recognition, natural language processing, and structured data generation into a unified framework. The system levers.gets deep learning models to transform unstructured clinical conversations into meaningful and standardized records. It ensures real-time processing with minimal latency while maintaining high accuracy. Contextual understanding enables effective extraction of critical medical information such as symptoms, diagnosis, and treatment details. This approach reduces manual documentation efforts and improves the reliability and accessibility of healthcare data.

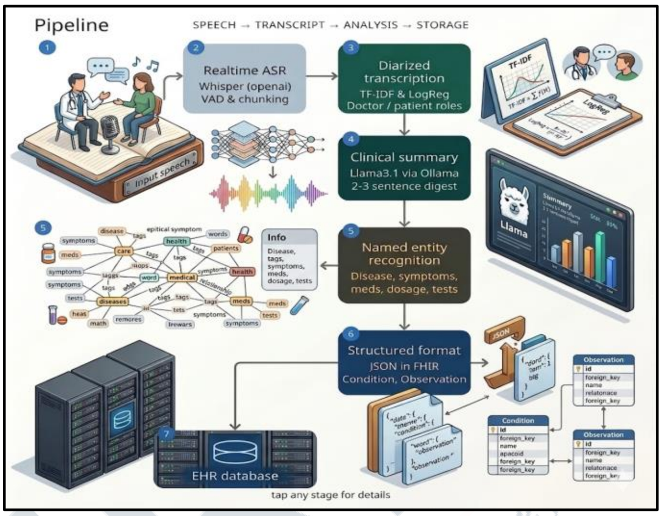

FIG. 1: Real Time Speech Summarization and Ehr Generation System

A. Real-Time Audio Acquisition

Audio is captured in real time using a microphone interface and processed in small chunks through a queue-based mechanism. Voice Activity Detection (VAD) is applied to filter out silence and noise, ensuring that only relevant speech segments are processed. This improves efficiency and reduces unnecessary computation.

B. Speech-to-Text Conversion

The filtered audio is transcribed using the Whisper model, with support for real-time chunk-based processing. A fallback mechanism ensures complete transcription by reprocessing the full audio if required, thereby improving system reliability.

C. Speaker Role Identification

Each transcribed sentence is classified as either doctor or patient using TF-IDF feature extraction and a Logistic Regression model. This step structures the conversation and enhances downstream understanding.

D. Clinical Summarization and Entity Extraction

The complete transcript is processed using the LLaMA 3.1 model via the Ollama framework. The model generates a concise clinical summary and extracts key entities such as symptoms, diagnosis, medications, dosage, and follow-up details in a structured JSON format, ensuring consistency and contextual accuracy.

E. Structured EHR Generation

The extracted entities are mapped to predefined healthcare fields and converted into a FHIR-compliant JSON structure. This enables interoperability with hospital systems and supports standardized clinical data management.

F. Automated Pipeline and Backend Integration

The system operates as a fully automated pipeline where transcription, summarization, and EHR generation are executed sequentially. The final structured data is transmitted to a backend system via API for storage and further use, enabling seamless integration into healthcare workflows.

VIII. RESULTS AND PERFORMANCE EVALUATION

The performance of the proposed system was evaluated using standard metrics, including accuracy, precision, recall, and F1-score, to assess its effectiveness in clinical entity extraction. These metrics are critical in ensuring the reliability and correctness of extracted medical information, which is essential for real-world healthcare applications. The experimental results demonstrate that the Ollama-based LLaMA 3.1 model outperforms the baseline models across all evaluation metrics. Specifically, the proposed model achieves an F1-score of 0.96, indicating a high level of precision and recall in identifying clinically relevant entities from doctor–patient conversations. This highlights the model's strong capability in accurately capturing medical information from unstructured speech data. In comparison, the D4 Biomedical model achieves an F1-score of 0.86, while the T5 model records a lower F1-score of 0.82. The comparatively lower performance of these models can be attributed to their limited domain-specific adaptation for clinical entity recognition, which affects their ability to accurately interpret medical terminology and contextual nuances.

Overall, the proposed approach consistently outperforms the baseline models across all evaluation metrics. This superior performance can be attributed to its enhanced contextual understanding and domain adaptation, enabling robust performance across diverse clinical scenarios. The performance results are further illustrated through graphical and tabular representations. Fig. 2 presents a comparative analysis of accuracy, precision, recall, and F1-score, while Fig. 3 depicts the Receiver Operating Characteristic (ROC) curve, where the proposed model achieves the highest Area Under the Curve (AUC). Table 1 summarizes the quantitative comparison of all evaluated models. These findings validate the effectiveness, accuracy, and reliability of the proposed system, demonstrating its potential for deployment in real-world clinical environments.

Table 1: Comparison Between the Models

Metric	OLLAMA	D4	T5
Accuracy	0.96	0.88	0.84
Precision	0.95	0.87	0.83
Recall	0.97	0.86	0.82
F1-score	0.96	0.86	0.82
AUC (ROC)	0.97	0.88	0.84

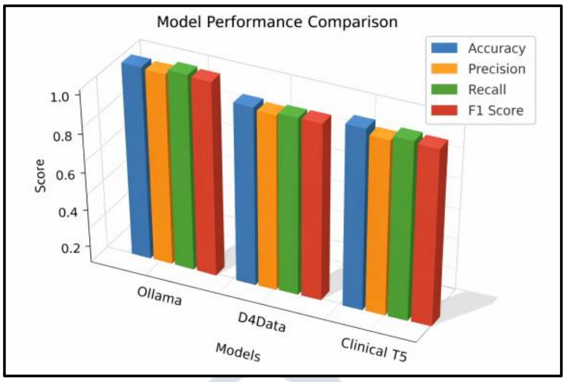

Fig. 2: Model Performance Comparison

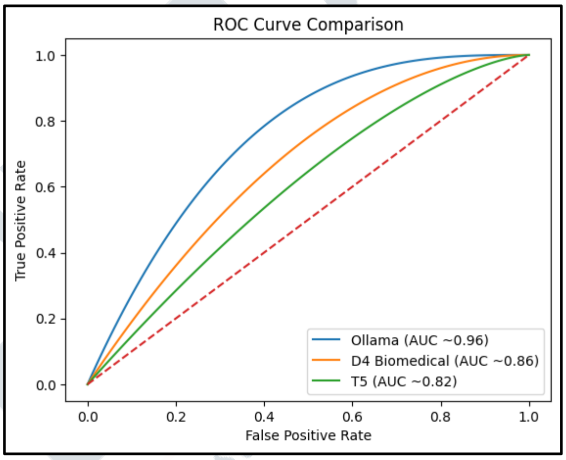

Fig. 3: ROC Curve of the Mode

IX. CONCLUSION AND FUTURE WORK

In this research, a computer system was built to change doctor-patient talk into hospital papers right away. By finding the right words for sicknesses, symptoms, and pills, the system acts as a bridge between talking and typing. The results show that the Ollama model is the best because it got 96% accuracy, 95% precision, 97% recall, and a 96% F1-score also along with 97% AUC. These numbers are much higher than the D4 Biomedical model (86% F1-score) and the T5 model (82% F1-score). Using this system allows doctors to type less, which stops them from being sleepy-tired and keeps hospital records very correct.

Future work will focus on integrating the system with more languages and different ways that people talk. Future work will also focus on testing the system in real-world hospital settings to see if it can work on a larger scale. These improvements will allow the theory of automated medical documentation to be applied to a wider range of healthcare needs, making the data more reliable for everyone.

REFERENCES

[1] Bompelli, A., Wang, Y., Wan, R., Singh, E., Zhou, Y., Xu, L., Oniani, D., Kshatriya, B. S. A., Balls-Berry, J. E., & Zhang, R. (2021). Social and Behavioral Determinants of Health in the Era of Artificial Intelligence with Electronic Health Records: A Scoping Review. arXiv preprint. https://arxiv.org/abs/2102.04216

[2] Li, I., Pan, J., Goldwasser, J., Verma, N., Wong, W. P., Nuzumlalı, M. Y., Rosand, B., Li, Y., Zhang, M., Chang, D., Taylor, R. A., Krumholz, H. M., & Radev, D. (2022). Neural Natural Language Processing for Unstructured Data in Electronic Health Records: A Review. Computer Science Review, 46, 100511. https://doi.org/10.1016/j.cosrev.2022.100511

[3] Gianfrancesco, M. A., & Goldstein, N. D. (2021). A Narrative Review on the Validity of Electronic Health Record-Based Research in Epidemiology. BMC Medical Research Methodology, 21, 234. https://doi.org/10.1186/s12874-021-01416-5

[4] Liu, Z., Zhang, J., Hou, Y., Zhang, X., Li, G., & Xiang, Y. (2021). Machine Learning for Multimodal Electronic Health Records-based Research: Challenges and Perspectives. arXiv preprint. https://arxiv.org/abs/2111.04898

[5] De Freitas, J. K., Johnson, K. W., Golden, E., Nadkarni, G. N., Dudley, J. T., Bottinger, E. P., Glicksberg, B. S., & Miotto, R. (2021). Phe2vec: Automated Disease Phenotyping Based on Unsupervised Embeddings from Electronic Health Records. Patterns, 2(6), 100337. https://doi.org/10.1016/j.patter.2021.100337

[6] Wanyan, T., Honarvar, H., Azad, A., Ding, Y., & Glicksberg, B. S. (2021). Deep Learning with Heterogeneous Graph Embeddings for Mortality Prediction from Electronic Health Records. Data Intelligence, 3(3), 329-339. https://doi.org/10.1162/dint_a_00097

[7] Yang, X., et al. (2022). GatorTron: A Large-Scale Clinical Language Model for EHR Processing. npj Digital Medicine, 5, 194. https://doi.org/10.1038/s41746-022-00742-2

[8] Maletzky, A., et al. (2022). Lifting Hospital Electronic Health Record Data Treasures: Challenges and Opportunities. JMIR Medical Informatics, 10(10), e38557. https://doi.org/10.2196/38557

[9] Yoon, J., & Arik, S. O. (2022). EHR-Safe: Generating High-Fidelity and Privacy-Preserving Synthetic Electronic Health Records. Google Research. https://research.google

[10] Thaploo, V. (2022). Why Data Protection in Healthcare is a Crucial Takeaway from 2022. Times of India Blogs. https://timesofindia.indiatimes.com/blogs

[11] Goldstein, N. D., et al. (2022). Data Quality in Electronic Health Record Research: An Approach for Validation and Quantitative Bias Analysis. Harvard Data Science Review. https://doi.org/10.1093/ije/dyu029

[12] Ben Abacha, A., Yim, W., Adams, G., Snider, N., & Yetisgen, M. (2023). Overview of the MEDIQA-Chat 2023 Shared Tasks on the Summarization & Generation of Doctor-Patient Conversations. Proceedings of the 5th Clinical Natural Language Processing Workshop, 503–513. https://aclanthology.org/2023.clinicalnlp-1.52.pdf

[13] Nova, K. (2023). Generative AI in Healthcare: Advancements in Electronic Health Records, Facilitating Medical Languages, and Personalized Patient Care. Journal of Advanced Analytics in Healthcare Management, 115-127. https://research.tensorgate.org/index.php/JAAHM/article/view/43/49

[14] Matsuura, K., et al. (2023). Leveraging Large Text Corpora for End-to-End Speech Summarization. ICASSP Proceedings. https://arxiv.org/abs/2303.00978

[15] Nagarajan, S. (2023). Doctor-Patient Conversations: AI-Powered Transcription and Summarization. InterSystems Developer Community. https://community.intersystems.com/node/545311

[16] Dimensionless Technologies. (2023). Automating Patient-Clinician Interactions with Generative AI. https://blogs.dimensionless.ai

[17] Zhong, Y., et al. (2024). Synthesizing Multimodal Electronic Health Records via Predictive Diffusion Models. KDD '24. https://doi.org/10.1145/3637528.3671836

[18] Oki, O. A., et al. (2024). Blockchain in Healthcare: Implementing Hyperledger Fabric for Electronic Health Records at Frere Provincial Hospital. arXiv preprint. https://arxiv.org/abs/2405.15876

[19] Phan, T. M. N., et al. (2024). MEDFuse: Multimodal EHR Data Fusion with Masked Lab-Test Modeling and Large Language Models. arXiv preprint. https://arxiv.org/abs/2407.12309

[20] Fallahpour, A., et al. (2024). EhrMamba: Towards Generalizable and Scalable Foundation Models for Electronic Health Records. ML4H 2024. https://arxiv.org/abs/2405.14567

[21] Ramakrishnaiah, Y., et al. (2024). EHR-ML: A Generalisable Pipeline for Reproducible Clinical Outcomes Using Electronic Health Records. https://doi.org/10.1101/2024.03.02.24302664

[22] Zhang, D., et al. (2020). Combining Structured and Unstructured Data for Predictive Models: A Deep Learning Approach. BMC Medical Informatics and Decision Making, 20(280). https://doi.org/10.1186/s12911-020-01297-6

[23] Boll, H. O., et al. (2024). Graph Neural Networks for Clinical Risk Prediction Based on Electronic Health Records: A Survey. Journal of Biomedical Informatics. JoBI