International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Web Scraping for Ovarian Cancer Detection: Utilizing Open-Source Whisper AI for Identifying Relevant Terminology and Improving Early Diagnosis

Vijayshri Khedkar[1], Pooja Bagane[1], Sonali Kothari[1], Anubha Gupta[1], Utkarsh Singh[1], Sahil Gupta[1], Tanya Agrawal[1], Dr. M. Karthikeyan[2]

Submitted: 11/05/2023 **Revised**: 16/07/2023 **Accepted**: 07/08/2023

Abstract: This research paper investigates the effectiveness of automatic speech recognition (ASR) using OpenAI Whisper module in detecting chemical word entities related to ovarian cancer from human speech. Ovarian cancer is a deadly disease that requires early detection for successful treatment. The proposed ASR system is based on deep learning models capable of recognizing complex speech patterns and distinguishing between different chemical terms related to ovarian cancer. Moreover, the detected chemical entities are used for web content search and retrieval, which can help in discovering useful information related to ovarian cancer. This study highlights the potential of using ASR technology for early detection and accurate identification of ovarian cancer-related chemical entities and utilizing them for retrieving relevant information from the web and opens new avenues for developing intelligent systems for disease diagnosis and treatment.

Keywords: *Automatic Speech Recognition; chemical named entity recognition; Ovarian Cancer; Natural language processing; Web Scraping*

1. Introduction

Ovarian cancer is a silent killer and one of the deadliest gynecological malignancies. Early detection and accurate diagnosis are crucial for improving patient outcomes. With advancements in natural language processing (NLP) and voice recognition technologies, voice assistants have emerged as promising tools for facilitating medical research and clinical decision-making. This study aims to develop a voice assistant that can accurately recognize ovarian cancer related terms from human speech input. The voice assistant will utilize ASR and NLP algorithms to analyze spoken words and identify keywords associated with ovarian cancer, such as "ovarian cancer," "ovarian tumor," "ovarian mass," "ovarian neoplasm," "CA-125," and "BRCA1/2 mutations," among others. The voice assistant will then search for relevant research papers from reputable scientific databases, such as PubMed, Scopus, and Google Scholar, using these keywords. The goal of this study is to help researchers, clinicians, and patients efficiently access and review the latest scientific literature related to ovarian cancer. By providing relevant research papers on ovarian cancer related terms, it can aid in evidence-based decision-making, facilitate scientific discovery, and promote knowledge dissemination in the field of ovarian cancer research.

[1] [1]Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune, India
ORCID ID: 0000-0001-6704-4823
ORCID ID : 0000-0001-9611-9601
[2] CSIR-NCL, Pune, India
** Corresponding Author Email: vijayshri.khedkar@sitpune.edu.in*

Automatic Speech Recognition (ASR) technology, which converts the spoken language into written text, plays a significant role in correctly recognizing chemical terms from human speech. The Whisper ASR system is used to accept the input from the user and provide transcriptions based on the same. ASR-generated transcriptions of chemical terms from human speech can be used for data analysis and retrieval purposes. Accurate recognition of chemical terms by ASR allows for efficient searching, indexing, and retrieval of relevant information from the transcriptions, enabling researchers to analyze and extract meaningful insights from their data.

2. Literature Review

Development of the Speech-to-Text Chatbot Interface Based on Google API discusses the development of a chatbot interface that uses Google's speech-to-text API to convert speech into text. The study presents a detailed methodology that involves integrating Google's speech-to-text API with a chatbot application, which allows users to interact with the system using natural language. The researchers evaluated the performance of the speech-to-text conversion system and the chatbot interface using various metrics, such as accuracy, response time, and user satisfaction [1]. The Whisper ROS Wrapper is a lightweight and efficient software module that enables automatic speech recognition (ASR) in embedded systems using the Robot Operating System (ROS) framework. It provides a simple interface for audio capture, speech recognition, and publishing the recognized text to the ROS network. The wrapper has been used in various applications, including mobile robotics, smart homes, robotic arm systems, and wearable devices [3].

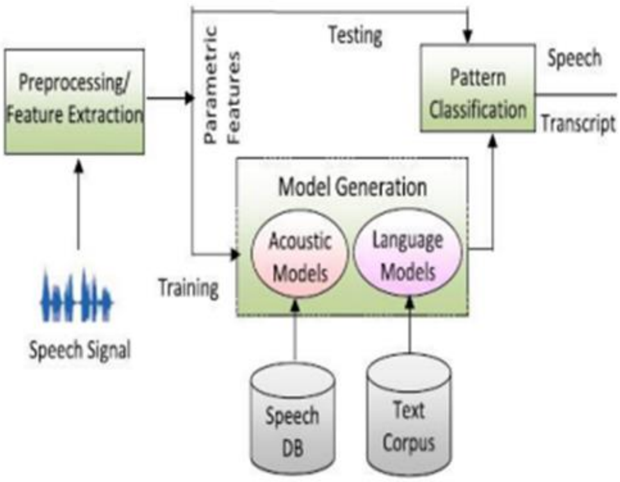

Fig. 1: System Architecture of the Automatic Speech Recognition System [4]

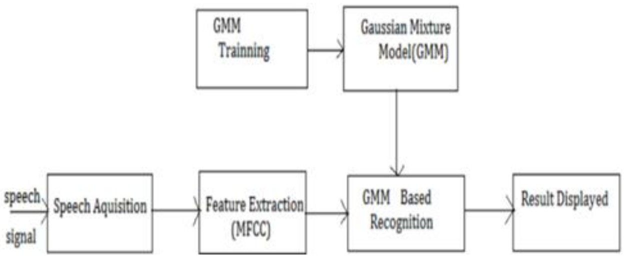

Fig. 2: General Block Diagram of Speech Signal Processing [7]

In [4], Suman K. Saksamudre discusses various methods for speech recognition. The paper provides a comprehensive overview of the traditional and modern techniques used for speech recognition, including acoustic phonetics, dynamic time warping, hidden Markov models, artificial neural networks, and deep learning approaches. Figure. 1 describes the functioning of a Speech Recognition System. The author highlights the advantages and limitations of each method and concludes that deep learning-based approaches have shown promising results in improving speech recognition accuracy. For those working in the subject of voice recognition, researchers and practitioners, the paper is an invaluable resource. The introduction of a weakly supervised method for automated speech recognition (ASR) has sparked advancements in speech recognition through the development of unsupervised pre-training techniques. The approach is based on a large-scale dataset of weakly labeled audio and text data, which is used to train a speech recognition system. The method is shown to achieve state-of-the-art performance on several ASR benchmarks, including clean and noisy speech [5]. Another effective technique utilizes automatic speech recognition (ASR) and natural language processing (NLP) to transcribe spoken language into written text and generate a summary. The proposed approach involves segmenting the audio into sentences, transcribing the speech into text using ASR, and then applying NLP techniques to summarize the text. The approach is shown to be effective in generating concise summaries of audio content and can be used for various applications, such as note-taking, accessibility, and knowledge management [6]. Gaussian Mixture models are used for representing Normally Distributed subpopulations within an overall population and the same can be used for converting speech into the written text that involves training a GMM on a dataset of speech signals and corresponding text transcriptions, segmenting the speech into phonemes, and estimating the likelihood of each phoneme given the speech signal. Finally, the phonemes are combined to form the final text transcription [7]. The general flow of processing and recognition of the speech signal is shown in Figure 2. Natural Language Management Systems under Variety of Accents uses Speech Recognition Algorithm that uses deep neural networks (DNN) to improve the performance of speech recognition systems for a variety of accents. The paper presents a detailed methodology that involves collecting a diverse dataset of speech signals and corresponding text transcriptions and training a DNN to recognize speech in various accents [8]. In [9], the author M. Benzeghiba reviews the impact of speech variability on automatic speech recognition (ASR) systems. He also discusses various techniques used to address the problem of speech variability, including feature normalization, speaker adaptation, and channel compensation. The review highlights the importance of addressing speech variability in the design and implementation of ASR systems to improve their accuracy and robustness in various real-world applications.

Table 1: Comparative study of existing work

	STRENGTH	**DRAWBACK**	**FINDING**
[11]	Pattern-based speech recognition is the most successful approach, reducing computation and redundancies in speech signals by extracting a limited number of parameters and using computer macro commands.	1. For speaker-independent speech recognition, parameters must be insensitive to the spoken language. 2. These systems require the ability to comprehend and store a vast range of	1. The usage of separate words, dependent systems, the total number of dictionary lexical items, language grammar, and controlled environmental conditions are five strategies that can be utilized to

		idiomatic expressions beyond standard language.	direct and facilitate voice recognition. 2. The speech recognition system for healthcare in German language achieved a word recognition accuracy of 92%-94% within a month of implementation, which improved to 97% for standardized texts.
[2]	1. The FreqDist function in the nltk library shows the user the frequency distribution of each word in the text. 2. Chemical data is extracted from the data using Chemdataextractor.	1. Python libraries were utilized for implementation and experimentation, which proved to be efficient, but lacked a user-friendly interface for non-coders. Therefore, developing an interface could be a crucial aspect of future development. 2. The program only stores the dictionary outside of its code, and if necessary, the word cloud can be saved externally.	A word is lemmatized when it is reduced to its lexical or root form, sometimes referred to as the lemma. Lemmatization is the process of converting words into a basic structure that can be used for text analysis, information retrieval, and machine learning activities. Chemdataextractor is an open-source software tool that is designed to

			automatically extract chemical information from scientific literature. The software uses natural language processing and machine learning algorithms to identify chemical entities, such as chemical names, formulas, and properties, and extract them from text.
[12]	Improved accuracy: ASR systems have made significant advancements in accuracy, allowing for more reliable and efficient speech recognition. Language flexibility: ASR systems can handle a wide range of languages, enabling multilingual and cross-lingual applications.	Background noise sensitivity: ASR systems can struggle to accurately recognize speech in noisy environments, affecting their performance in real-world scenarios. Speaker variability: ASR systems may encounter difficulties when dealing with accents, dialects, or variations in speech patterns, leading to	Research has shown that incorporating deep learning techniques, such as recurrent neural networks (RNNs) and convolutional neural networks (CNNs), has improved the accuracy and robustness of ASR systems. Various methods have been proposed to address background noise, including denoising algorithms and

		reduced accuracy.	beamforming techniques, to enhance ASR performance in noisy environments.
[13]	Comprehensive Comparison: The study provides a thorough evaluation and comparison of various automatic speech recognition (ASR) techniques. Broad Scope: The study encompasses a wide range of ASR techniques, including both traditional and state-of-the-art approaches.	Limited Timeframe: The study's findings are based on the research conducted within a specific timeframe, potentially excluding recent advancements in ASR techniques. Subjectivity in Evaluation: The evaluation of ASR techniques may involve subjective judgments, which can introduce bias into the study.	Comparative Performance: The study compares the accuracy, efficiency, and robustness of various ASR techniques, such as Hidden Markov Models (HMM), Deep Neural Networks (DNN), Recurrent Neural Networks (RNN), and Transformer-based models. It identifies the strengths and weaknesses of each technique in different scenarios. Language and Accent Variability: The research investigates the performance of ASR techniques across different languages and accents, highlighting variations in accuracy and the impact of data availability.
[14]	Comprehensive evaluation of a wide range of speech-to-text (STT) and text-to-speech (TTS) recognition systems, covering techniques, algorithms, and applications. In-depth analysis of system strengths and limitations, including performance, accuracy, usability, and technological advancements.	Scope limitations may exclude recent or specialized STT and TTS technologies, or focus on specific domains or languages. Subjective evaluations introduce potential bias based on researchers' preferences, expertise, or criteria.	Performance comparison of STT and TTS systems, highlighting accuracy, language support, noise robustness, and speaker variability. Considering real-time processing, multilingual support, the naturalness of speech synthesis, and adaptability to user requirements, application analysis is conducted across sectors such as transcription, voice assistants, accessibility, and customer service.

3. Whisper by Openai

The Whisper ASR system is based on a deep learning architecture that uses recurrent neural networks (RNNs), specifically a variant called long short-term memory (LSTM) network, which is a type of recurrent neural network with special gating mechanisms to capture long-term dependencies in sequential data. LSTMs are known for their ability to model sequential data, such as speech signals, and have been widely used in ASR tasks.

The Whisper ASR system is trained on a large amount of multilingual and multitask supervised data collected from the web, making it capable of converting spoken language into written text across different languages and domains. It has been trained on a massive amount of data to achieve high accuracy and performance in ASR tasks. The Whisper ASR API provides developers with an interface to utilize the Whisper ASR system for speech recognition capabilities in their applications or services, without having to train or fine-tune the model themselves.

The proposed model leverages the Whisper model's capabilities to process unstructured speech data and extract meaningful information related to ovarian cancer and different pre-processing techniques, feature representations, and model configurations are explored to optimize the performance of the Whisper model for this specific task.

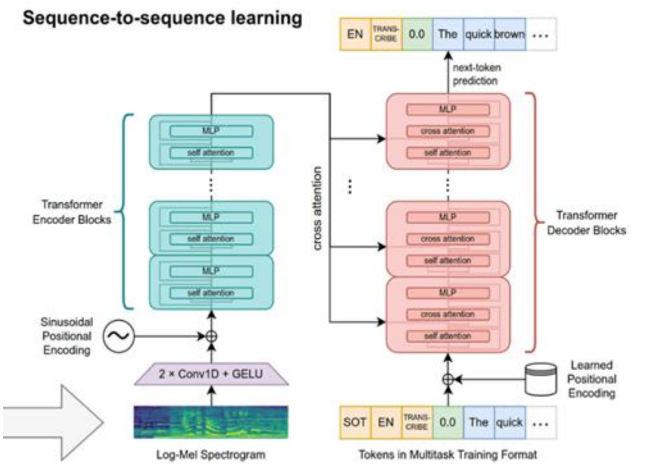

Fig. 3: A sequence-to-sequence Transformer model of Whisper [9]

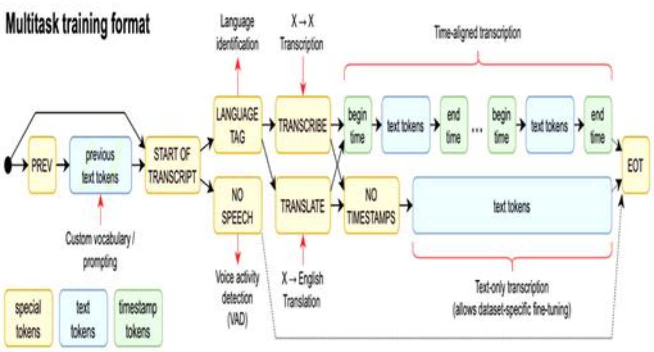

Fig. 4: Multitask Training Format of Whisper [9]

4. Proposed Methodology

Speech Acquisition:

To ensure high-quality speech acquisition, a high-quality microphone is selected and its placement and positioning is optimized. A directional microphone is used, which selectively captures sound from the participant's mouth while minimizing ambient noise. The microphone is placed at an appropriate distance and angle from the participant's mouth to capture the sound waves with high fidelity. The recorded speech is then stored in memory.

Speech Preprocessing:

The encoder in Whisper is a deep neural network that takes the raw audio waveform as input and processes it using multiple layers of convolutional or recurrent neural networks (RNNs). The first step in preprocessing an audio waveform is to transform it into a spectrogram, which is a graphic representation of the frequency content of the audio signal across time. The spectrogram is then fed into the encoder network, which learns to extract meaningful features from the audio signal.

The encoder network in Whisper is typically a convolutional neural network (CNN) or a bidirectional RNN (BiRNN) that is designed to capture both the temporal and spectral features of the audio signal. The output of the encoder network is a fixed-length bottleneck vector that summarizes the most important features of the audio signal.

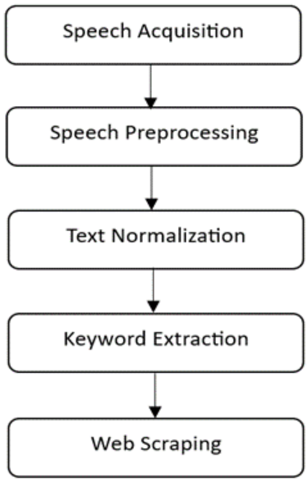

Fig. 5: Speech recognition process flow

The bottleneck vector is then fed into the decoder network, which generates the output speech signal. The decoder network in Whisper is typically an RNN that takes the bottleneck vector as input and generates the corresponding spectrogram. The spectrogram is then converted into the final speech signal using a vocoder, which is a signal processing algorithm that can convert the spectrogram back into an audio waveform.

Text normalization:

The output from the speech to text conversion process is normalized to remove stop words, punctuation marks, and other non-textual elements. This involves converting the text to lowercase, removing numbers and special characters, and tokenizing the text into individual words.

Keyword Extraction:

This is accomplished either manually by compiling a vocabulary of chemical terminology from published works or automatically using a program like PubChem. The preprocessed text is then compared against the dictionary of chemical terms to identify relevant entities. Any words or phrases that match the terms in the dictionary are extracted as chemical entities.

PubMed Web Scraping:

Once the keywords are extracted, they are used to search the PubMed database for relevant articles using web scraping techniques. This involves writing a script that sends a request to the PubMed database, retrieves the search results, and extracts the relevant information such as article title, author name, abstract, and publication date.

5. Results

The proposed model takes a voice input and extracts the chemical terms from the given input. The proposed model was given the following line-“What is the role of cisplatin in ovarian cancer”. All the words were then separately compared with the words from the dictionary which includes chemical entities related to ovarian cancer. From this, the words “cisplatin”, “ovarian” and “cancer” were found to be present in the dictionary. These three words were then converted back into a string with ‘+’ symbol between each of the words. The string obtained is “ovarian+cisplatin+cancer”. This string was then used to obtain the information (Title of the paper, Abstract of the paper, Year of publishing, Authors, Journal, Digital Object Identifier) regarding n number of research papers from PubMed as well as the top n relevant links from Google. Here, the number ‘n’ is defined by the user. The information related to the research papers was downloaded as a .csv file.

This technique can save time and effort in manually searching for and selecting important cancer-related language, allowing researchers to focus on the analysis and interpretation of their findings. The pre-trained model can also deliver reliable and consistent results, decreasing the possibility of human error.

6. Conclusion

It is vital to stress, however, that the usage of such technologies should not be used to substitute researchers' critical thinking and analysis skills. It should be utilized as an additional tool to help in the research process. Additionally, researchers should confirm that the identified cancer-related locution is contextually acceptable and appropriately reflects the intended meaning.

Overall, web-based assistance for recognizing cancer-related locution utilizing speech synthesis with pre-trained models has the potential to improve the efficiency and accuracy of oncology research, but it should be utilized with caution and with a critical eye on the data it produces.

7. Future Scope

With the tremendous advancements in the field of artificial intelligence (AI) and natural language processing (NLP), it is now possible to construct sophisticated models that can effectively identify cancer-related terms and concepts from audio data. Such models may be trained on enormous databases of cancer-related lectures and research articles, allowing them to learn to recognize the most popular terminology, concepts, and phrases used in the subject. Once trained, these models might be used to automatically transcribe spoken conversations, extract essential cancer-related concepts and phrases, and even generate written summaries or research papers.

One potential application for such technology would be to provide real-time help to researchers and medical professionals during cancer-related talks or presentations. The model might listen to the speech, identify significant concepts and words, and offer suggestions or comments to the speaker to improve communication and understanding.

Another possible application would be to assist researchers in drafting research papers on cancer-related issues. The model may take notes during discussions or presentations, identify the most essential concepts and phrases, and then generate a draft of the article for the researcher to revise and edit.

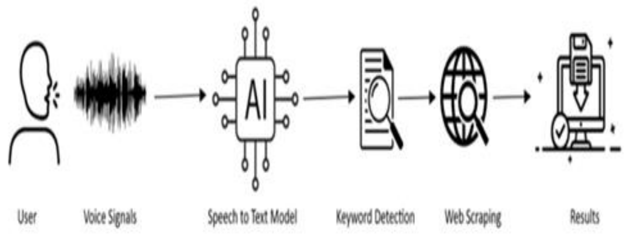

Fig. 6: Working of the proposed mode

Overall, the future of web-assisted cancer locution detection utilizing speech synthesis with a pre-trained model appears bright. Seeing more sophisticated and accurate models should be anticipated that can support researchers and medical practitioners in their work on cancer-related themes as AI and NLP technologies progress.

References

[1] Nataliya Shakhovska, Oleh Basystiuk, Khrystyna Shakhovska: Development of the Speech-to-Text Chatbot Interface Based on Google API(2022), Researchgate

[2] Dr. Sonali Kothari Tidke, Adhiraj Dev Goswami, Prof. Vijayshri Khedkar, Muskaan Agrawal, Anvita Gupta, Kajal Jaggi: Identification of chemical entities from prescribed drugs for ovarian cancer by text mining of medical records (2022), IEEE

[3] Andrés A. Ramírez-Duque, Mary Ellen Foster: A Whisper ROS Wrapper to Enable Automatic Speech Recognition in Embedded Systems(2023), HRCI 2023

[4] Suman K. Saksamudre, P.P. Shrishrima, R.R. Deshmukh: A Review on Different Approaches for Speech Recognition System(2015), International Journal of Computer Applications

[5] Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, Ilya Sutskever: Robust Speech Recognition via Large-Scale Weak Supervision (2019), IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) Pages: 5961-5965, DOI: 10.1109/ICASSP.2019.8682574

[6] Vinnarasu A., Deepa V. Jose: Speech to text conversion and summarization for effective understanding and documentation(2019), International Journal of Electrical and Computer Engineering (IJECE)

[7] Virendra Chauhan, Shobhana Dwivedi, Pooja Karale, Prof. S.M. Potdar: Speech to Text Converter Using Gaussian Mixture Model(2016), International Research Journal of Engineering and Technology(IRJET)

[8] Irina Gurtueva , Olga Nagoeva , and Inna Pshenokova: Speech recognition algorithm for natural language management systems under variety of accents(2020), E3S Web of Conferences

[9] M. Benzeghiba, R. De Mori, F. Dufour, and P. J. Godfrey, “Automatic speech recognition and speech variability: A review,” Speech Communication, vol. 57, pp. 109-129, 2014.

[10] Santosh K. Gaikwad, Bharti W. Gawali, Pravin Yannawar: A review on Speech Recognition Technique

[11] S. Ajami, "Use of speech-to-text technology for documentation by healthcare providers," Int. J. Healthc. Technol. Manag., vol. 15, no. 1, pp. 23-32, 2016.

[12] Wiqas Ghai, Navdeep Singh: “Literature Review on Automatic Speech Recognition”, International Journal of Computer Applications (0975 – 8887), Vol. 41, No. 8, March 2012.

[13] Michelle Cutajar, Edward Gatt, Ivan Grech, Owen Casha, Joseph Micallef: “Comparative Study of Automatic Speech Recognition techniques”, IET Signal Processing, January 2013.

[14] Ayushi Trivedi, Navya Pant, Pinal Shah, Simran Sonik, Supriya Agrawal: “Speech to text and text to speech recognition system – A Review”, IOSR Journal of Computer Engineering, Vol. 20, Issue 2, Ver. 1 (March – April 2018), PP 36-43.

[15] Ms. Sweta Minj. (2012). Design and Analysis of Class-E Power Amplifier for Wired & Wireless Systems. International Journal of New Practices in Management and Engineering, 1(04), 07 - 13. Retrieved from http://ijnpme.org/index.php/IJNPME/article/view/9

[16] Faris, W. F. . (2020). Cataract Eye Detection Using Deep Learning Based Feature Extraction with Classification. Research Journal of Computer Systems and Engineering, 1(2), 20:25. Retrieved from https://technicaljournals.org/RJCSE/index.php/journal/article/view/7

[17] Sherje, N.P., Agrawal, S.A., Umbarkar, A.M., Kharche, P.P., Dhabliya, D. Machinability study and optimization of CNC drilling process parameters for HSLA steel with coated and uncoated drill bit (2021) Materials Today: Proceedings,