2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS)
Karnataka, India. Jun 28-29, 2024

2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS) | 979-8-3503-8269-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICITEICS61368.2024.10624940

Chrome Extension for Speech-to-Text Conversion and Text Summarization using NLP

Gayatri Patil
Computer Science and Technology
Usha Mittal Institute of Technology, SNDTWU
Mumbai, India
gkp512003@gmail.com

Krithika Saravanan
Computer Science and Technology
Usha Mittal Institute of Technology, SNDTWU
Mumbai, India
krithisarva7@gmail.com

Bhakti Sapariya
Computer Science and Technology
Usha Mittal Institute of Technology, SNDTWU
Mumbai, India
bhaktisapariya16@gmail.com

Ms. Prajakta Gotarne
Assistant Professor of Computer Science and Technology
Usha Mittal Institute of Technology, SNDTWU
Mumbai, India
prajakta2490@gmail.com

***Abstract*—In the evolving landscape of communication, the increasing reliance on web conferencing and online meetings has become the norm for various purposes, ranging from educational lectures to business discussions. However, the virtual nature of these interactions poses challenges in capturing the essence of conversations, leading to the need for efficient tools. This research introduces a Chrome extension designed to address this challenge by incorporating automatic Speech-to-Text conversion and Text Summarization using Natural Language Processing (NLP). The extension aims to enhance communication on platforms such as Google Meet, Zoom, Teams, and more online meetings. By seamlessly transcribing spoken language into text and providing summarization capabilities, this tool aims to alleviate the time-consuming and distracting task of minute-taking, ensuring that crucial details are never missed during discussions. The dual-functionality of Speech-to-Text conversion and Text Summarization not only addresses the challenges posed by the virtual nature of modern communication but also establishes a practical solution for optimizing the communication experience in diverse settings.**

***Index Terms*—Natural Language Processing (NLP), Text Summarization, Minute-taking, Speech to text conversion**

I. Introduction

In the ever-evolving landscape of technological innovation, Natural Language Processing (NLP) emerges as a pivotal force, empowering systems to comprehend and manipulate textual or spoken language. This transformative technique not only facilitates the understanding of instructions but also opens avenues for extracting valuable insights from large volumes of textual data. Our research seeks to harness the potential of NLP within the context of Google Chrome extensions, introducing a novel approach that seamlessly transforms speech into written text while generating concise summaries of content directly within the browsing experience.

Our proposed system, known for enhancing the functionality of the popular web browser, serves as a versatile tool that can modify the interaction capabilities of web pages. Our research strategically integrates the power of speech-to-text technology and NLP algorithms into a Chrome extension, revolutionizing the way users engage in online meetings. This dynamic combination not only addresses the challenges posed by the virtual nature of modern communication but also introduces a user-friendly and efficient tool for diverse purposes.

Recognizing the broad spectrum of applications[6] for text summarization and speech-to-text conversion, our research delves into the core purpose of these NLP tasks. Automatic text summarization, a key component of our investigation, aims to present original material in a semantically concise format. This condensed representation not only expedites the reading process but also ensures a clear understanding of the document's key ideas, expressed in plain and accessible language. In response, our Chrome extension not only aims to revolutionize information retention during discussions but also seeks to redefine the efficiency of these interactions. By minimizing the traditionally time-consuming and distracting practice of minute-taking, the extension ensures that participants can engage in a dialogue without fear of missing critical details, fostering a more productive and inclusive communication environment.

This proposal has been put forward for a variety of reasons, endeavors to bridge the realms of NLP, technology, and communication dynamics, presenting a comprehensive solution that extends beyond the technical landscape to impact the way individuals interact and collaborate in virtual and offline settings.

II. Literature Review

The paper[1] proposes that Understanding the intricate nuances of human language has been a longstanding challenge in the field of information systems research. By delving into the depths of NLP, we aim to elucidate the progress made in computational text analysis and the challenges that persist in achieving a comprehensive comprehension of human language. The research highlights the transformative impact of NLP on converting human language into text. The advancements in computational text analysis and the successful application of NLP techniques across various linguistic tasks demonstrate the potential of these technologies. However, the persistent challenges related to different accents and pronunciation variations underscore the need for continued research and development to enhance the robustness and adaptability of NLP systems. As we delve into the subsequent sections of this research paper, we aim to contribute to the ongoing discourse by exploring innovative solutions to address these challenges and further advance the field of NLP.

The paper[2] explored the integration of business intelligence and statistical approaches through automatic Speech-to-Text conversion and Text Summarization. The model employed a Multi-Layered Text Summarization technique to optimize the outcome of transcribed text, achieving a compression rate of 60 percent. The automated categorization module exhibited notable performance, attaining a commendable accuracy of 70 percent and an overall F-score of 0.81349.

Incorporating emotive capabilities such as tone, the system aimed to enhance user perception. However, the paper acknowledged a limitation, highlighting the imperative need to perfect and fine-tune the algorithm utilized in the categorization modules.

The paper [3] presents a Speech Recognition model designed to transform user-provided speech data into text format in their preferred language. The primary objective of this research is to construct a speech recognition model that empowers even individuals with low literacy levels to seamlessly interact with computer systems using their regional language. To achieve this, the model incorporates Multilingual features, building upon the existing Google Speech Recognition model by integrating principles derived from natural language processing.

However, the system acknowledges a limitation, recognizing potential challenges in speech recognition systems when dealing with specific languages or accents. The accuracy of transcription may be affected if users possess a strong accent or communicate in less common languages. This limitation underscores the importance of ongoing efforts to enhance the system's adaptability and broaden its language recognition capabilities.

The paper[4] proposes a comprehensive review paper that explores diverse approaches for generating summaries from extensive text documents. The paper outlines methods capable of producing either Abstractive (ABS) or Extractive (EXT) summaries, emphasizing their role in efficiently condensing large volumes of data into concise representations. The overarching goal is to save users' time and resources in navigating and extracting valuable information from extensive textual content.

To assess the quality and relevance of the generated summaries, the paper suggests the utilization of common evaluation metrics such as ROUGE score and TF-IDF scores. These metrics serve as quantitative measures to evaluate the effectiveness of different summarization methods.

However, the research also highlights a notable limitation in the field – the ongoing challenge of ensuring accuracy and relevance in the generated summaries. Despite the application of diverse algorithms and evaluation metrics, there is no specific model identified that consistently produces the best summaries. This acknowledgment underscores the complexity of the task and the need for continued research and innovation to enhance the reliability and effectiveness of text summarization methods.

The paper[5] proposes to analyze the text data to measure similarity, facilitating the correction process based on essay responses.

The key technology employed in this study is Automated Essay Scoring (AES), a computational approach that automatically determines scores from text documents. The primary goal is to streamline and automate the correction and scoring of essays by leveraging computer-based applications. Through the utilization of AES, the study seeks to enhance efficiency in grading processes, providing a more automated and standardized method for evaluating essay responses.

The paper[6] proposes a research study focused on developing an efficient method for speech recognition and text summarization. It highlights the challenges associated with variations in speech, including differences in pace, dialect, and pronunciation, which can hinder effective communication. The interdisciplinary field of computational linguistics is identified as instrumental in overcoming these challenges by leveraging speech recognition technologies.

The primary objective of the research is to present a method that seamlessly converts speech into text while providing a succinct summary of the content. This approach is deemed valuable for various applications, such as creating lecture notes and summarizing lengthy documents or catalogs. The proposed method is positioned as an easy and effective solution to enhance the understanding and accessibility of spoken content.

The research work emphasizes the integration of speech recognition and text summarization, showcasing its potential to facilitate tasks that require the extraction of essential information from spoken language. The efficiency of the pro-

posed method is validated through extensive experimentation, suggesting its practical applicability across diverse scenarios.

III. Proposed System

The proposed system aims to bridge the gap in digital communication by introducing a Chrome extension that leverages advanced NLP techniques to facilitate real-time speech-to-text conversion and automatic text summarization within web conferencing platforms like Google Meet, Zoom, and Teams. This innovation seeks to address the prevalent challenge of efficiently capturing and distilling the essence of spoken discourse in online meetings a need that has become increasingly critical in contemporary remote work and learning environments. By automating the transcription and summarization processes, the extension not only enhances accessibility and comprehension for participants but also significantly reduces the time and effort traditionally required to manually document and summarize key points from digital meetings. This system is particularly beneficial for users who aim to focus on the ongoing conversation without worrying about note-taking, ensuring that no critical information is missed.

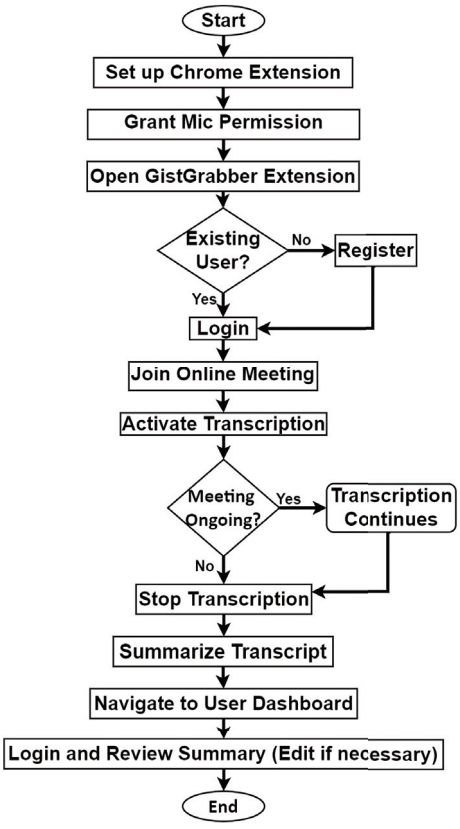

Fig. 1. Workflow Diagram of the Proposed System

A. Proposed System Procedure

Based on Figure 1, procedure for the proposed model is as follows:

1) Download the Chrome extension zip file. Then, in the Chrome browser, navigate to chrome://extensions/, enable "Developer mode", and use the "Load unpacked" option to select and upload the extension zip file.
2) After installation, give the extension permission to use your computer's microphone to capture audio.
3) Access the extension's interface to either log in to your existing account or register for a new one with your credentials.
4) Enter your online meeting through whichever online web conferencing platform you're using (Zoom, Teams, etc.).
5) With the meeting in progress, activate the transcription feature in the Chrome extension.
6) Conclude the transcription by stopping the feature once your meeting ends.
7) After ensuring all relevant data is captured, you can close the extension.
8) Go through the transcribed data to make sure it accurately reflects the meeting. Make any needed edits or remove irrelevant sections.
9) Using the summarization feature, create concise and informative meeting notes.
10) Navigate to your user dashboard on the platform associated with the Chrome extension.
11) Log in and review the summarized notes, ensuring they meet your requirements for clarity and conciseness.

IV. Methodology

The methodology of our research paper entails a comprehensive approach divided into modules. Through systematic steps, we aim to integrate advanced natural language processing techniques into the Chrome extension framework. Our methodology encompasses real-time audio extraction, transcription, and backend processing for text summarization using machine learning algorithms. By leveraging these modules, our objective is to create a user-friendly dashboard interface for seamless interaction and control over the generated content, ensuring efficient utilization in various online communication scenarios.

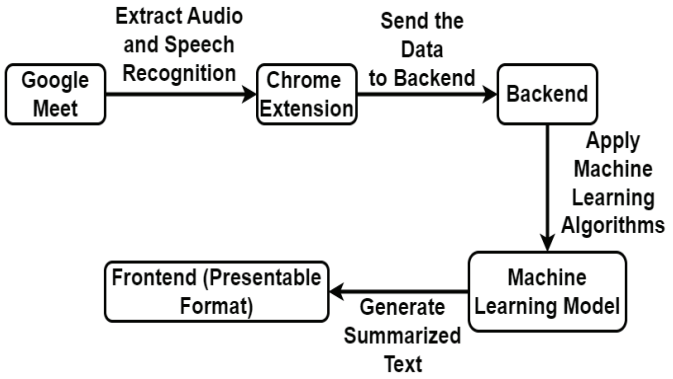

Fig. 2. Basic Working of the System

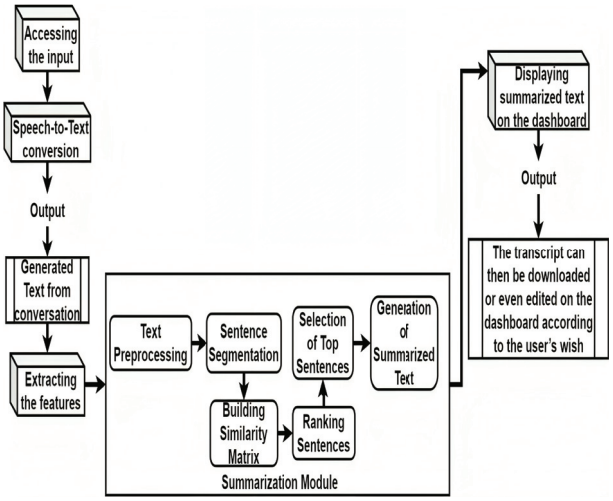

Fig. 3. System Architecture of the Project

A. Extension Initialization

A Chrome extension is a compact software tool designed to enrich a user's online experience by augmenting their browser's capabilities. It operates through the browser's functionalities to modify or amplify its performance. Featuring an intuitive popup interface, it streamlines user interaction. Once the extension is uploaded, users are prompted to grant microphone permissions, crucial for voice recognition. This extension streamlines user authentication, ensuring smooth login or registration directly within the extension interface. Once authenticated, users unlock a comprehensive array of functionalities, including seamless participation in virtual meetings across diverse platforms. Notably, the extension simplifies meeting transcription, enabling users to initiate and halt the process effortlessly.

B. Real-time Audio Extraction and Transcription

This module focuses on real-time audio extraction and transcription within the context of Google Meet sessions. As participants engage in conversation, our Chrome extension actively extracts audio data, monitoring each speaker's input and capturing their speech. Leveraging the Microsoft Azure Speech-to-Text SDK, this data is swiftly processed, converting spoken words into written text in real-time. Microsoft Azure Speech to Text SDK is a cutting-edge tool designed for accurate and efficient transcription tasks. It stands out for its robustness, offering high accuracy even in noisy environments or with varying accents. Its real-time capabilities align perfectly with our objective of providing users with immediate access to meeting transcripts as discussions progress.

Microsoft Azure Speech to Text SDK employs advanced machine learning algorithms, continuously improving its transcription accuracy through ongoing training with diverse datasets. Its adaptability makes it an ideal choice for our Chrome extension, ensuring reliable performance across different scenarios. Moreover, it seamlessly integrates with our platform, facilitating smooth implementation and operation. By harnessing the power of Microsoft Azure Speech to Text SDK, our Chrome extension delivers a seamless and invaluable experience, enabling users to effortlessly follow and engage with meeting discussions through real-time transcription.

C. Text Summarization

Upon extracting the transcript in text format, the system engages in Machine Learning techniques for text summarization.

1) Exploration of Summarization Approaches: In the domain of text summarization for our Chrome extension, we initiated a thorough exploration, examining diverse summarization methodologies with the aim of efficiently and precisely capturing the core of spoken content. Considering the risk of potential misinterpretation of information inherent in abstractive summarization, the team opted for the more nuanced extractive summarization method.

2) Utilization of NLP Algorithms: To implement extractive summarization, the system leverages TF-IDF (Term Frequency-Inverse Document Frequency) and cosine similarity algorithms. These algorithms play a pivotal role in identifying significant phrases and key information within the transcript.

3) Implementation of TF-IDF and Cosine Similarity: TF-IDF serves as a cornerstone for identifying the significance of words within the transcript. By quantifying the importance of each term based on its frequency in the document and across a corpus of documents, TF-IDF enables us to discern key points and phrases essential for summarization.

We employ cosine similarity to measure the resemblance between sentences. This metric evaluates the angle between two sentence vectors, allowing us to gauge their cosine similarity irrespective of their length.[5]

4) Execution of Summarization Process: The summarization process is executed as follows:

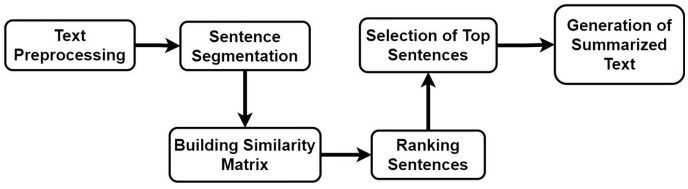

Fig. 4. Summarization Process of the Backend Module

1) Text Preprocessing and Sentence Segmentation:
   The initial phase involves parsing the extracted transcript, breaking it down into individual sentences, and ensuring its cleanliness for subsequent analysis.
2) Building Similarity Matrix:
   Through a meticulous process, we construct a similarity

matrix among sentences, gauging their semantic proximity using cosine similarity metrics. This step lays the foundation for identifying the most salient sentences.

3) Ranking Sentences:
   Employing the network's PageRank algorithm, we assign scores to each sentence based on its centrality within the semantic network constructed earlier. This step facilitates the identification of pivotal sentences crucial for summarization.
4) Selection of Top Sentences:
   Armed with ranked sentences, we curate the summary by selecting the most significant ones, adhering to a predefined threshold for conciseness.
5) Generation of Summarized Text:
   Finally, utilizing the selected sentences, we generate a coherent and condensed summary, encapsulating the essence of the original discourse succinctly.

This approach to text summarization underscores the commitment to precision, ensuring that the generated summaries are impactful and formal, without compromising on accuracy or clarity.

D. User Dashboard Interface

In this section, we detail the methodology employed for the development and implementation of the User Dashboard module, which serves as an integral component of our Chrome extension for speech-to-text conversion and text summarization. The summarized text is made available to the user through a user-friendly dashboard interface as meeting minutes. The system primarily comprises two interconnected components: the backend, responsible for user authentication, processing speech-to-text conversion and text summarization, and the frontend, which encompasses the User Dashboard.

1) Frontend Development with ReactJS: The User Dashboard is meticulously crafted using ReactJS, a JavaScript library renowned for its component-based architecture and efficient rendering capabilities. This section outlines the key steps involved in frontend development.

a) User Authentication Functionality: Users are provided with a streamlined login/register experience, allowing them to access the dashboard securely using their credentials. Leveraging ReactJS's state management capabilities, we implement an authentication mechanism to authenticate users and authorize access to the dashboard.

b) Moment of Meeting Management: The User Dashboard empowers users to effortlessly view and edit moments of meeting generated from speech-to-text conversion and text summarization processes. Utilizing ReactJS components, we design an intuitive interface that enables users to interact with and manipulate meeting transcripts seamlessly.

c) Asynchronous User Activities: To ensure concurrent usage by multiple users without compromising data integrity, we implement asynchronous handling of user activities within the dashboard. By leveraging ReactJS's asynchronous capabilities and efficient state management, users can perform actions simultaneously, such as editing transcripts or sharing meeting summaries.

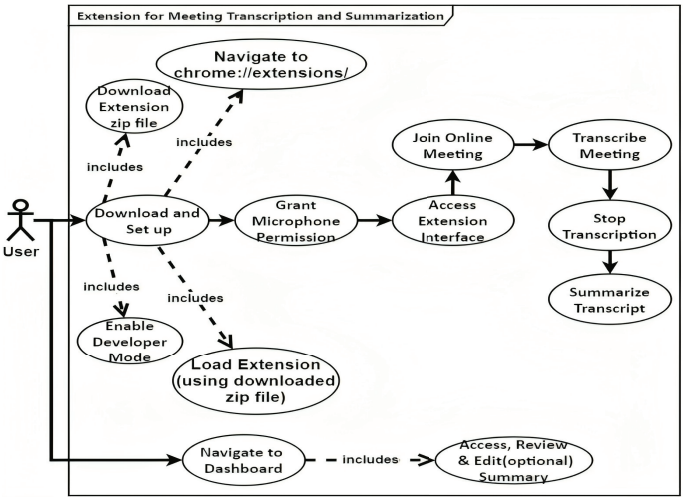

Fig. 5. System Use Case Diagram

d) Integration with REST APIs: Seamless integration with Django REST APIs is paramount for facilitating data exchange between the frontend and backend components. Through the use of fetch requests, the User Dashboard interacts with the server-side APIs to retrieve, update, and manipulate meeting data in real time.

2) Backend Development with Django: The backend is skillfully developed using Django, a well-regarded Python web framework known for its straightforwardness, adaptability, and reliability. The User Dashboard interacts with the backend system to retrieve and display meeting transcripts generated by the speech-to-text conversion and summarization processes. Data flow between the frontend and backend components is facilitated through Django REST APIs, ensuring efficient communication and synchronization of user activities.

a) Notes and Moment of Meeting Management: The backend consists of app modules responsible for listing, editing, and deleting notes and moments of meetings (MOM). Leveraging Django's Model-View-Controller (MVC) architecture, we design models to represent transcripted text and MOM, along with corresponding views and controllers for CRUD (Create, Read, Update, Delete) operations. The SQLite database, provided by Django, efficiently stores and retrieves data related to notes and MOM.

b) User Authentication: User authentication is seamlessly integrated into the backend using Django's built-in authentication system. This ensures secure access to the application, allowing users to register, log in, and manage their accounts securely. User account information, including usernames, passwords (hashed for security), and authentication tokens, are stored in the SQLite database. This ensures that user credentials are securely managed and validated during the authentication process.

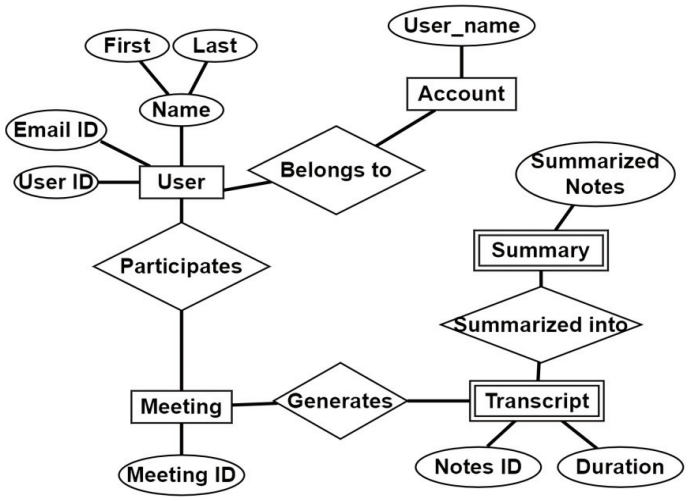

Fig. 6. Entity Relationship Diagram

c) Text Summarization: To facilitate text summarization, a dedicated module within the backend interacts with the SQLite database to retrieve textual data for processing. Leveraging natural language processing (NLP) techniques and external libraries, such as NLTK (Natural Language Toolkit), we implemented algorithms to summarize textual content efficiently. Upon summarization, users are notified via email, enabling them to access and review the summarized transcripts conveniently.

V. Result and Analysis

As a result, the proposed system effortlessly merges speech-to-text conversion and text summarization technologies to elevate the effectiveness of online meetings. By automating transcription and summarization tasks, the system endeavors to reduce the need for manual note-taking, allowing participants to engage fully in discussions without overlooking vital information.

1) Upon logging into Google Meet and enabling the Chrome extension, the system initiates by recording speech or audio input via a microphone.
2) Utilizing speech-to-text technology, the Chrome extension converts the recorded audio into a textual transcription of the meeting.
3) The transcribed content is sent to the backend, where Machine Learning algorithms are employed to produce comprehensive meeting summaries.
4) Subsequently, the meeting minutes are formatted into presentable notes and displayed on the user's dashboard in the frontend interface.

VI. Conclusion

In our research, we aimed to make the task of documenting lengthy speeches during online meetings easier and less time-consuming. To achieve this, we developed a Chrome extension that serves as an online meeting assistant. This tool automates the process of generating meeting minutes by utilizing speech-to-text conversion and text summarization techniques. By doing so, it allows users to concentrate more on participating in the conversation during video conferencing platforms like Google Meet. Our extension is a step towards increasing productivity and efficiency in virtual communication environments.

VII. Future Work

In the future, we aim to significantly enhance the functionality of our Chrome extension by incorporating support for multiple languages in both speech recognition and text summarization. This expansion will make the extension more accessible and user-friendly for a global audience. Furthermore, we plan to integrate a feedback mechanism within the extension. This will enable users to provide their feedback and suggestions, allowing us to refine and improve the tool based on their needs and preferences. Additionally, we are looking to develop the extension into a cross-browser platform to broaden its accessibility.

Moreover, an exciting development in our roadmap includes the addition of speaker recognition functionality. This feature will enable the extension to not only recognize but also trace the identity of the speaker, enhancing the utility of the tool for applications involving multiple speakers.

Acknowledgment

We are extremely grateful to the Department of Computer Science and Technology, Shreemati Nathibai Damodar Thackersey Women's University for providing us with the platform and all necessary materials required to carry out this study.

References

[1] R. Kaur and A. Kaur, "Text Generator using Natural Language Processing Methods," 2023 International Conference on Computational Intelligence, Communication Technology and Networking (CICTN), Ghaziabad, India, 2023, pp. 421-425, doi: 10.1109/CICTN57981.2023.10140271.

[2] Hegdepatil and K. Davuluri, "Business Intelligence based novel Marketing Strategy Approach using Automatic Speech Recognition and Text Summarization," 2021 2nd International Conference on Computing and Data Science (CDS), Stanford, CA, USA, 2021, pp. 595-602,doi: 10.1109/CDS52072.2021.00108.

[3] S. Bano, P. Jithendra, G. L. Niharika and Y. Sikhi, "Speech to Text Translation enabling Multilingualism," 2020 IEEE International Conference for Innovation in Technology (INOCON),Bangluru, India, 2020, pp. 1-4, doi: 10.1109/INOCON50539.2020.9298280.

[4] Rahul, S. Adhikari and Monika, "NLP based Machine Learning Approaches for Text Summarization," 2020 Fourth International Conference on Computing Method- ologies and Communication (ICCMC), Erode, India, 2020, pp. 535-538, doi: 10.1109/ICCMC48092.2020.ICCMC-00099.

[5] R. Lahitani, A. E. Permanasari and N. A. Setiawan, "Cosine similarity to determine similarity measure: Study case in online essay assessment," 2016 4th International Conference on Cyber and IT Service Management, Bandung, Indonesia, 2016, pp. 1-6, doi:10.1109/CITSM.2016.7577578.

[6] International Journal of Electrical and Computer Engineering (IJECE) Vol. 9, No. 5, October 2019, pp. 3642 3648 ISSN: 2088-8708, DOI: 10.11591/ijece.v9i5.pp3642-3648.

[7] T. Tyagi, L. Dhari, Y. Nigam and R. Nagpal, "Video Summarization using Speech Recognition and Text Summarization," 2023 4th International Conference for Emerging Technology (INCET), Belgaum, India, 2023, pp. 1-7, doi: 10.1109/INCET57972.2023.10169901.