Meeting Summarizer Using NLP

LEKHANA R[1], NISHATH ANJUM[2], SYED SAQEEB[3], THANUSHREE B S[4], IRFAN KHAN[5]
[1, 2, 3, 4]*Information Science Engineering, Ghousia College of Engineering*
[5]*Asst Proffessor, CSE, Ghousia College of Engineering*

Abstract- This project proposes a hybrid Natural Language Processing (NLP) system for automating meeting summarization and action item extraction. The system converts meeting audio into text using speech-to-text technology, preprocesses the data, and applies a combination of extractive and abstractive summarization techniques utilizing transformer models such as T5 and BART. Additionally, it extracts critical action items including tasks, deadlines, and responsible individuals through named entity recognition and dependency parsing. The designed web-based platform enhances productivity by delivering clear, concise meeting summaries and structured follow-up actions. Evaluation on real and synthetic datasets demonstrates improved accuracy and effectiveness over traditional methods, making it a valuable tool for efficient organizational communication.

Keywords— Meeting summarization, Natural language processing, Extractive summarization, Abstractive summarization, Action item extraction, Speech-to-text, Transformer models, Named entity recognition, Hybrid NLP system, Organizational productivity.

I. INTRODUCTION

In the evolving landscape of digital collaboration, conferences constitute a fundamental mechanism through which organizations align objectives, share knowledge, and coordinate strategies. From corporate meeting room to academic seminars and team huddles, the volume and frequency of meetings have hurled exponentially. However, the valuable insights, decisions, and action points emerging from these encounters often remain inadequately captured or dispersed, discouraging effective communication and follow-up.

Automated meeting summarization aims to transform raw meeting content—whether in radio frequency audio, video recordings, or textual transcripts—into concise, intelligible, and informative abstracts. These summaries provide a high-level overview of discussions, key decisions, and tasks assigned, significantly reducing cognitive load for participants and extending accessibility to non-attendees. Moreover, automatic extraction of action items such as to-dos, responsible shareholders, and deadlines enhances responsibility and project management efficiency.

Traditional note-taking methods suffer from restrictions such as human error, irregularity, and delayed availability. This has inspired considerable interest in automated meeting summarization systems that control advances in Natural Language Processing (NLP) and artificial intelligence.

Automated meeting summarization goals to transform raw meeting content—whether in radio frequency audio, video recordings, or textual transcripts—into crisp, coherent, and informative summaries. These summaries provide a high-level outline of discussions, key decisions, and tasks assigned, significantly reducing intellectual load for participants and extending accessibility to non-attendees. Moreover, automatic extraction of action items such as to- dos, responsible sponsors, and time limit enhances responsibility and project management competence.

II. LITERARY REVIEW

Automated meeting summarization and action item extraction have rapidly evolved as crucial research areas within Natural Language Processing (NLP), owed to the increasing importance of efficient communication in organizational and isolated work settings. The difficulty of meetings, categorized by multi-speaker interactions, spontaneous speech, and diverse domains, makes this an especially challenging task discrete from predictable manuscript summarization.

Early developments in speech-to-text transcription relied heavily on statistical models such as Hidden Markov Models (HMMs) and Gaussian Mixture Models (GMMs). However, these approaches often

© OCT 2025 | IRE Journals | Volume 9 Issue 4 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I4-1712104

fell short in handling diverse accents, noisy environments, and spontaneous speech typical in meetings. Recent advances leverage deep learning methods including end-to-end architectures like DeepSpeech and OpenAI's Whisper model. Whisper, in particular, validates superior multilingual transcription capabilities and robustness against background noise, making it highly suitable for real-world meeting applications.

In the province of text summarization, initial research focused principally on extractive techniques. Methods such as TF-IDF- based sentence selection, graph-based algorithms like TextRank, and centroid-based approaches acknowledged the most relevant condemnations based on lexical similarity and position. Regardless of their computational efficiency, these methods classically produced summaries with limited coherence, frequent redundancy, and missing nuanced conversational context.

The advent of transformer-based deep learning models modernized abstractive summarization. Architectures like BERTSUM, Pointer-Generator Networks, and the T5 and PEGASUS transformers introduced the capability to generate fluent, reworded summaries preserving semantic content beyond direct extraction. By engaging attention mechanisms and encoder-decoder designs, these models capture global context and syntactic dependencies..

III. PROPOSED METHODS

The proposed system is an NLP-based AI framework designed to automate meeting summarization and extract action items from both audio recordings and textual transcripts. The system integrates state-of-the-art techniques spanning speech-to-text conversion, text preprocessing, hybrid summarization, and information extraction to deliver accurate, concise, and actionable meeting outputs.

The proposed NLP-based AI System integrates state-of-the- art natural language processing and machine learning techniques to automatically summarize meeting content and extract actionable tasks. The system's architecture includes the following key components:

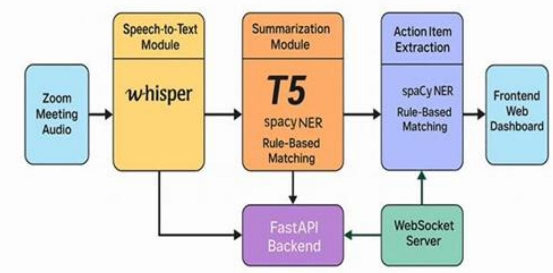


Fig. 1. Architectural diagram

Furthermore, as far as we are aware, no other programs have succeeded in providing a real-time meeting report. In order to create a web extension for lecture summaries, we want to create a web application. With the help of this study, we can distill an hour-long meeting into a single document that is simple to use and offers a concise summary. Both business and education could make use of this application. The proposed system's architecture is shown in Figure 1. The investigation of numerous libraries, such as the hugging face transformer Spacy NER (rule-based matching), whisper , T5 provided us with a systematic understanding of the libraries available on the market.

1. An extractive summarizer identifies key sentences by measuring semantic relevance using TF-IDF and word embedding's to filter essential content.

2. The system accepts meeting inputs as audio files (in formats such as .wav, .mp3) or textual transcripts (.txt, .docx). For audio files, a speech-to-text engine is employed to transcribe spoken content into text.

3. An extractive summarizer identifies key sentences by measuring semantic relevance using TF-IDF and word embeddings to filter essential content

Handwritten Text Recognition.

Artificial Intelligence: Handwritten recognition is one of the most challenging pattern recognition tasks. It is used in a variety of settings, including courtrooms and postal services for postal code recognition. First, a test photograph with white backdrop and black text is taken.

Speech-to-Text: The Whisper model, chosen for its robust multilingual support and noise tolerance, transcribes audio inputs. Depending on volume and latency requirements, smaller or base variants are used to balance accuracy and inference speed.

Procedure is used to analyze the features of the extracted handwritten text, and the extraction method is utilized to extract the compactness of each text's pixels. Around 90% of the time, the system is thought to be accurate.

Deep Neural Networks((DNN): The accuracy and speed of handwriting detection have increased because to the development of deep neural networks. People can now rely on features and algorithms that were obtained from data instead of ones that were manually created[6]. All previous methods rely on a different system where the image is initially divided, the extracted image's angle, curve, and height are fixed, and there are constraints on the text's size and orientation. However, people's handwriting varies widely and is frequently extremely challenging to read, therefore these methods are either incorrect or ineffectual. A neural network, which is trained from data and is easily adaptable to new datasets, has all of these implicit processes and is unrestricted by any constraints. The structure is seen in Figure 2. With a 7.6% error rate compared to other cloud- based API is 14.6%, this algorithm [6] is trained at 145 DPI and outperforms all other models.

Figure 2 shows the Model Schematics. On the left, there is a CNN Encoder and Transformer Decoder. On the right, there is the Transformer Layer with an optionally Localised Self Attention.

B. Text Summarization

1) Extractive Text Summarization using Sentence Ranking: The most often occurring terms are those that are most pertinent, according to the primary idea underpinning extractive text summary. The technique first creates a frequency matrix, which is a list outlining the frequency with which each word occurs, and then it filters out the English words that are used the most. To determine a phrase's prominent context, the ratings of each word in the phrase are taken into account. The summarizer will extract the sentence fragments with the highest weighted frequency and present a summary.

2) Abstractive Text Summarization using BERT Model: BERT and K-Means are also used by extractive text summarization. The paragraphs are tokenized into sentences using the BERT model using a pipeline. Clustering is embedded using K-means. The last step is to choose the sentences closest to the centroid. Uses the PyTorch-pre-trained-BERT package for the BERT for Text Embedding algorithm. Clustering is accomplished with the help of K-

IV. RESULTS

The proposed application focuses on three different modules:

A. Key-Frame Extraction

The user launches the summarizing program, which extracts the keyframes (Figure 3) from the movie. A comparison algorithm is run through all the frames sequentially to check the similarity of the frames. If two frames are within the threshold, then the latest frame will be made the keyframe and the previous frame will be discarded. This helps reduce the redundancy of data and provides a more accurate measure of the words that have been utilized in the meeting. This method is selected assuming the host erases the canvas once complete and can give a more definitive measure of the different sections in the meeting. Figure 4 shows the key frames extracted from the chosen video using Convolutional Neural Networks by comparing the extracted images to select only the frames that are distinct from the preceding frames that have been extracted.

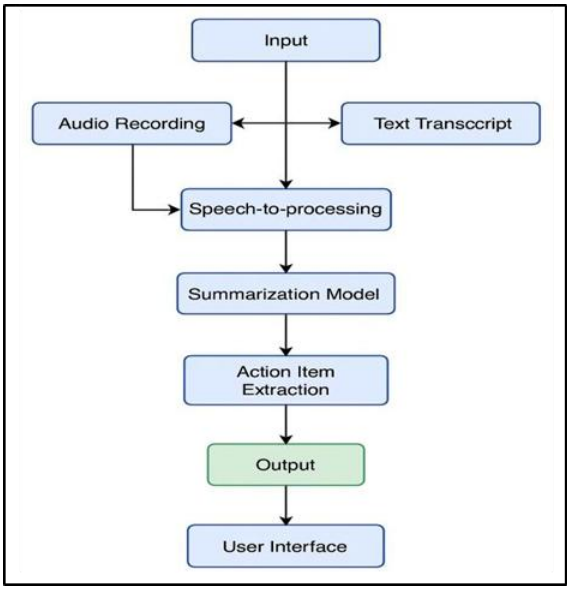


Fig. 2. DNN Model Schematics

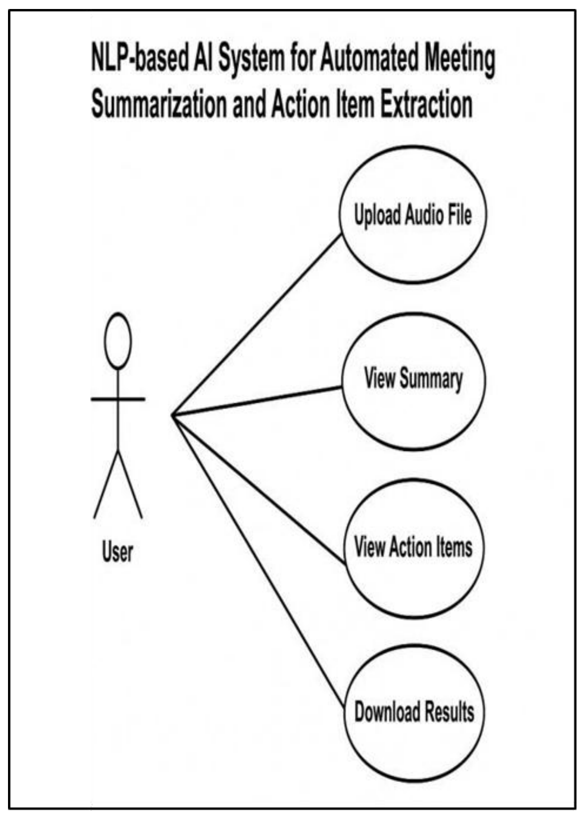
Means and Gaussian Mixture Models. A very modest number of phrases are required so that the model can successfully capture the context of the entire lecture. The model [20] included additional sentences which were beyond what was requested.

3) K Nearest Neighbours (KNN): The degree of feature values and feature similarity are taken into account in a modified version of the KNN algorithm. After selecting the K distinct words that are similar as the K nearest neighbours, the label for the beginner entity is decided by voting on the labels of the K closest neighbours. A binary category has been assigned to the text summation. A string of text that has been broken up into paragraphs using carriage returns makes up the input. [9].

4) TFRSP: This algorithm combines extractive and abstractive summarization techniques. In the first step, the Term Frequency-Inverse Document Frequency algorithm is coupled with the Text Ranking algorithm. The output of the is given as an input to the abstractive algorithm. The summary is generated using abstractive text summarization. The sequence-to-sequence model, a supervised learning technique with training and testing datasets, is used for abstractive text summarization. The summary increases the ROUGE score of the existing methods by 38.42 per cent.

Currently, there exists a way to summarize the video by using the MSR-VTT dataset which consists of over 10,000 YouTube clips along with their summaries but to use this model we need to process the entire video lecture at once which is inefficient and time-consuming. This method is a more improved implementation by using screenshots and audio from the host instead of the entire video. This allows us to the pipeline by using threads which reduces the time consumed while processing the video. Since this is used mainly for live meetings this reduces the time for the generation of the report.

Our application can record audio, video, and text from it. It can also translate speech to text and provide a summary. There is no market-available software that can offer all these features and effectively summarize the meeting. Therefore, there is room for potential future expansion and improvement of our research.



The use case diagram for the NLP-based AI System for Automated Meeting Summarization and Action Item Extraction visually represents the interaction between the user and the system. The user initiates the process by uploading an audio file of the meeting, which is then processed by the backend pipeline. Once the file is uploaded, the system transcribes the audio using speech-to-text conversion. The transcribed text is passed through NLP models that generate a concise summary of the discussion and extract action items. The user can then view the generated summary and the list of action items through the interface. Additionally, the system offers the functionality to download the final results for future reference or reporting purposes. This use case diagram captures the essential functionalities provided to the user, emphasizing the system's focus on accessibility,

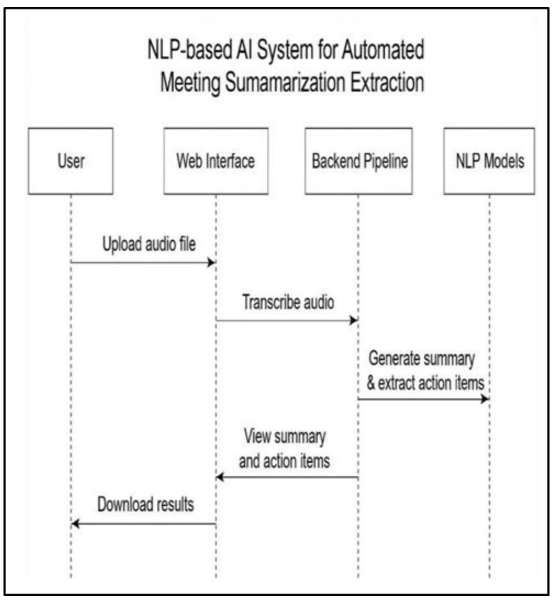


Fig. 3. Diagrammatic representation of a key frame extraction from video.

Output:
Video Description: The input video consists of a college lecture with slides about the topic.
Video Duration: 4:21
Frames Extracted: 24

Fig. 4. Frames extracted from the video

B. Speech Recognition(SR)

The application breaks up the audio into distinct segments after identifying the host's voice in it. These segments are produced using the time frames that were recorded during the frame extractor's production of unique frames. After being recorded, the audio is saved in the.wav format. The speech from the wav file is recognized using the Google Speech API and grouped with the relevant text created from the keyframe. The workflow for converting speech to text is shown in Figure 5, and the code .

The activity diagram for the NLP-based AI System for Automated Meeting Summarization and Action Item Extraction illustrates a systematic flow of how the system be in either audio or textual format. If the input is in audio form, the system employs a speech-to-text module to convert the audio into written text for further processing. If the input is already in text format, the system bypasses transcription and moves directly into natural language preprocessing, where the text is cleaned, segmented, and prepared for analysis processes meeting input and generates meaningful output for the user. The process begins when a user initiates the system by uploading or recording meeting data, which can

C. Text Summarization

The text generated from the above modules is passed to the text summarizer which summarizes the context and returns a summarized text document. Figure 7 shows the workflow of text summarization and Table I presents the original text and summarized words with % of reduction achieved.

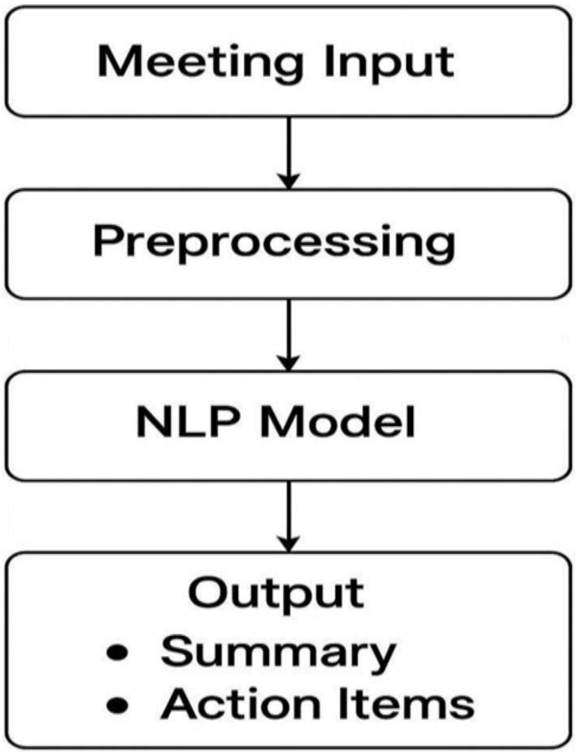


Fig. 7. Workflow of Text Summarization

TABLE I. SAMPLE INPUT TEXT AND ITS SUMMARY WITH REDUCTION PERCENTAGE

Sample Number	No. of words	No. of words after summarization	Reduction Percentage
1	190	86	53.01 %
2	150	80	46.45 %
3	94	48	43.47 %
4	123	63	57.72 %
5	210	89	57.62 %

After completely integrating all the modules together the application was run on a few small lecture videos to determine the time taken to generate the summary as depicted in Table II.

TABLE II. VIDEO SUMMARIZER RESULTS

Sample Number	Video Length	Total Time Taken by the Mod- ule
1	15 mins	325 secs
2	30 mins	574 secs

To get these values, the modules are run independently of the video but when taking into consideration that a lot of these operations are happening simultaneously as the lecture is being played, the time that a user must wait to get the summary is almost negligible.

V. CONCLUSION

This project successfully developed a hybrid NLP-based meeting summarization system that automates the transcription, summarization, and action item extraction from meeting audio and transcript data. By leveraging advanced transformer models such as T5 and BART for abstractive and extractive summarization, combined with named entity recognition and dependency parsing for precise task identification, the system effectively converts unstructured meeting content into clear, concise summaries and relevant actionable outputs

REFERENCES

[1] Kongthon, Alisa; Sangkeettrakarn, Chatchawal; Kongyoung, Sarawoot; Haruechaiyasak, Choochart (October 27–30, 2009). Implementing an online help desk system based on conversational agents. MEDES '09: The International Conference on Management of Emergent Digital EcoSystems. France: ACM. doi:10.1145/1643823.1643908.

[2] Mitchell, Tom (1997). Machine Learning. New York: McGraw Hill. ISBN 0-07-042807-7. OCLC 36417892.

[3] A. P. Singh, R. Nath and S. Kumar, "A Survey: Speech Recognition Approaches and Techniques," 2018 5th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), 2018, pp. 1-4, doi:10.1109/UPCON.2018.8596954.

[4] N. Chumuang and M. Ketcham, "Model for Handwritten Recognition Based on Artificial Intelligence," 2018 International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI- NLP), 2018, pp. 1-5, doi: 10.1109/iSAI-NLP.2018.8692958.

[5] J. N. Madhuri and R. Ganesh Kumar, "Extractive Text Summarization Using Sentence Ranking," 2019 International Conference on Data Science and Communication (IconDSC), 2019, pp. 1-3, doi: 10.1109/IconDSC.2019.8817040.

[6] S. S. Desai, D. Rajput and K. Patil, "An approach for Text Recognition from Document Images," 2020 IEEE Bangalore Humanitarian Technology Conference (B-HTC), 2020, pp. 1-5, doi: 10.1109/B-HTC50970.2020.9297939.

[7] Du, Y. et al. (2020) “PP-OCR: A practical ultra-lightweight OCR

[8] system,” arXiv [cs.CV]. doi: 10.48550/ARXIV.2009.09941.

[9] Ghadage, Y. H. and Shelke, S. D. (2016) “Speech to text conversion for multilingual languages,” in 2016 International Conference on Communication and Signal Processing (ICCSP). IEEE, pp. 0236– 0240.

[10] Jo, T. (2017) “K nearest neighbor for text summarization using feature similarity,” in 2017 International Conference on Communication, Control, Computing and Electronics Engineering (ICCCCEE). IEEE, pp. 1–5.

[11] Jolad, B. and Khanai, R. (2019) “An Art of Speech Recognition: A Review,” in 2019 2nd International Conference on Signal Processing and Communication (ICSPC). IEEE, pp. 31–35.

[12] Lakkhanawannakun, P. and Noyunsan, C. (2019) “Speech Recognition using Deep Learning,” in 2019 34th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC). IEEE, pp. 1–4.

[13] Ozsoy, M. G., Alpaslan, F. N. and Cicekli, I. (2011) “Text summarization using Latent Semantic Analysis,” Journal of information science, 37(4), pp. 405–417. doi: 10.1177/0165551511408848.

[14] Raundale, P. and Shekhar, H. (2021) “Analytical study of Text Summarization Techniques,” in 2021 Asian Conference on Innovation in Technology (ASIANCON). IEEE, pp. 1–4.

[15] Sharma, N. and Sardana, S. (2016) “A real time speech to text conversion system using bidirectional Kalman filter in Matlab,” in 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI). IEEE, pp. 2353–2357.

[16] Singh, S. S. and Karayev, S. (2021) “Full page handwriting recognition via image to sequence extraction,” in Document Analysis and Recognition – ICDAR 2021. Cham: Springer International