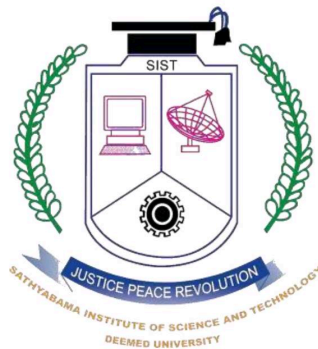


AUDIO SUMMARIZATION IN REAL TIME

Submitted in partial fulfillment of the
Requirements for the award of
Bachelor of Technology Degree in Information Technology

By

HARISH A (40120035)
SRIVARDHAN C (40120093)



DEPARTMENT OF INFORMATION TECHNOLOGY
SCHOOL OF COMPUTING

SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Category-1 University by UGC

Accredited with Grade "A++" by NAAC | 12B Status by UGC |

Approved by AICTE

JEPPIAR NAGAR, RAJIV GHANDHISALAI,
CHENNAI-600119

APRIL - 2024



SATHYABAMA

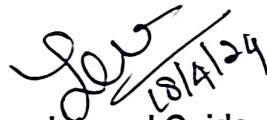
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)

Accredited "A++" Grade by NAAC | 12B Status by UGC | Approved by AICTE

www.sathyabama.ac.in

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of Harish A (40120035), Srivardhan C (40120093) who carried out the Projectentitles "Audio Summarization in Real Time" under my supervision from January 2024to April 2024.


18/4/24
Internal Guide

Ms. J. MERLIN MARY JENITHA M. Tech.,


18/4/24
Head of the Department

Dr. R. SUBHASHINI M.E., Ph.D.,

Submitted for Viva voce Examination held on 18.04.2024

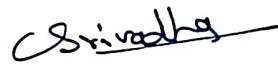

18/4/24
Internal Examiner


18/4/24
External Examiner

DECLARATION

I SRIVARDHAN C (40120093), hereby declare that the Project Phase-2 Report entitled "Audio Summarization in Real Time" done by me under the guidance of **Ms. MERLIN MARY JENITHA M. Tech.**, is submitted in partial fulfillment of the requirements for the award of Bachelor of Technology degree in Information Technology.

DATE: 18/04/2024



PLACE: Chennai

SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **Board of Management** of **SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T. SASIKALA M.E., Ph. D.**, Dean, School of Computing, **DR. R. SUBHASHINI M.E., Ph. D.**, Head of the Department of Information Technology for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Ms. J. Merlin Mary Jenitha M. Tech.**, for her valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Information Technology** who were helpful in many ways for the completion of the project.

ABSTRACT

Audio summarization is an essential task in natural language processing (NLP) that involves condensing the content of a multimedia file into a shorter version while preserving the key information. These summaries are valuable for various applications, such as video indexing, content retrieval, and efficient browsing. In recent years, researchers have made significant progress in developing automatic techniques for audio and video summarization. These methods typically involve several steps, including speech recognition, speaker diarization, topic segmentation, and keyframe extraction. Through the use of advanced machine learning algorithms and deep neural networks, these techniques can accurately identify and extract the most important audio segments and video frames, enabling efficient summarization. Additionally, NLP algorithms play a vital role in generating coherent and informative summaries by incorporating linguistic analysis, sentiment analysis, and storyboarding. The primary goal is to create a concise summary that captures the main topics, salient events, and emotional cues of the audiovisual content. This requires combining multiple modalities, such as audio, visuals, and textual transcripts, to provide a comprehensive summary of the original multimedia file. With the rapid advancement of NLP and audiovisual processing techniques, audio and video summarization is becoming increasingly accurate and efficient, opening up new opportunities for improving information retrieval, recommendation systems, and multimedia content analysis. As the demand for multimedia content continues to rise, the development of robust and reliable audio and video summarization techniques will play a crucial role in enhancing the overall user experience and usability of multimedia.

TABLE OF CONTENTS

Chapter No	TITLE	Page No
	ABSTRACT	iv
	LIST OF FIGURES	Vii
1	INTRODUCTION	1
2	LITERATURE SURVEY	3
2.1	INFERENCE FROM LITERATURE SURVEY	3
2.2	EXISTING SYSTEM AND PROPOSED SYSTEM	7
2.2.1	Existing System	7
2.2.2	Proposed System	7
2.3	PROBLEMS IN EXISTING SYSTEM	8
2.4	FEASIBILITY SYSTEM	9
2.4.1	Economic Feasibility	9
2.4.2	Technical Feasibility	9
2.4.3	Operational Feasibility	9
3	METHODOLOGY	10
3.1	FLOW CHART	11
3.2	SELECTED METHODOLOGY OR PROCESS MODEL	12
3.2.1	Introduction to Model Improvisation and training for audio summarization	12
3.2.2	Challenges in RealTime audio summarization	12
3.2.3	Evolution in model improvisation and training	13

3.3	TECHNOLOGIES IN AUDIO SUMMARIZATION	13
3.3.1	Speech Recognition	13
3.3.2	Natural Language Processing	13
3.3.3	Summarization algorithm	13
3.3.4	User feedback adaptation	14
3.3.5	Audio segmentation	14
3.4	ARCHITECTURE	14
3.5	DEVELOPMENT DETAILS	15
3.5.1	Software Configuration	15
3.5.2	Software Configuration	16
4	RESULTS AND DISCUSSIONS	19
5	CONCLUSION AND FUTURE WORK	21
5.1	Conclusion	21
5.2	Future work	21
6	REFERENCES	22
	APPENDICES	
A	SOURCE CODE	26
B	SCREENSHOTS	31
C	PLAGIARISM REPORT	32
D	RESEARCH PAPER	33

LIST OF FIGURES

FIGURE NO.	FIGURE NAME	PAGE NO
3.1	Flow Chart	11
3.2	System Architecture	14
4.1	Performance metrics	19
4.2	Accuracy Graph	19
4.3	Loss Graph	20

CHAPTER 1

INTRODUCTION

Audio summarization is a crucial task in natural language processing (NLP) that aims to condense lengthy audio or video content into shorter summaries, enabling more efficient access and understanding of the information contained within. With the increasing popularity and accessibility of multimedia content, the need for effective summarization techniques has become even more prominent.

In the context of audio summarization, the goal is to extract the most important information from spoken audio recordings. This can be particularly useful for tasks such as transcription, where summarization can help create concise and readable transcripts. Additionally, audio summarization holds value in applications like podcast recommendation systems, where users can benefit from brief summaries to decide which episodes match their interests.

Video summarization, on the other hand, deals with the challenge of condensing visual information contained in a video into a shorter representation. This can be accomplished by identifying key frames or extracting important objects, actions, or events from the video. Video summarization has a wide range of applications, including video surveillance, video browsing, and content-based video retrieval.

Both audio and video summarization tasks can be approached using various techniques from the field of NLP. One common approach is to use machine learning algorithms to automatically identify and extract salient information from the audio or video. These algorithms may leverage techniques such as natural language understanding, speech recognition, and computer vision to analyse and process the content. For example, in audio summarization, speech recognition algorithms can transcribe spoken words, while natural language understanding techniques can extract important concepts and keywords. Similarly, in video summarization, computer vision algorithms can identify important objects, actions, or scenes in the video.

Another approach to audio summarization is to utilize text and image- based features to identify key information. This involves converting the audio or video content into a text representation and then applying text summarization techniques to obtain a concise summary. Similarly, visual features from the video frames can be extracted and used to generate a summary. These features can be combined using techniques like multimodal fusion, where both textual and visual information are considered together to generate a comprehensive summary.

Evaluation of audio summarization systems is crucial to ensure their effectiveness and reliability. Several metrics can be used to assess the quality of the generated summaries, including precision, recall, and F1- score. Additionally, human evaluations can be conducted to gather subjective feedback and assess the overall usefulness and coherence of the summaries.

In conclusion, audio summarization is an important task in NLP that aims to condense lengthy audio or video content into shorter, more concise summaries. Techniques such as machine learning, natural language understanding, speech recognition, computer vision, and deep learning can be applied to achieve effective summarization. With the increasing amount of multimedia content available, the development of robust and accurate audio and video summarization systems holds great potential for enhancing information access and understanding.

Traditional methods involve manual inspection, which is not only time consuming and labor-intensive but also prone to human error. Furthermore, the frequency of inspections may not be sufficient to identify potholes in a timely manner, especially in vast road networks. As a result, there is a pressing need for innovative and automated solutions that can expedite the detection of potholes, enable swift repairs, and provide real- time alerts to motorists and road maintenance authorities. The advent of technology, particularly in the domains of computer vision, sensor networks, and data analytics, has opened up new possibilities for addressing the pothole problem.

CHAPTER 2

LITERATURE SURVEY

2.1 INFERENCE FROM LITERATURE SURVEY

[1] Saini, K., & Dave, M. (2022). Automatic Audio and Video Summarization Using TextRank Algorithm and Convolutional Neural Networks. In Advances in Data Computing, Communication and Security: Proceedings of I3CS2021 (pp. 397- 405). Singapore: Springer Nature Singapore.

In their paper, Saini and Dave (2022) explore the use of the TextRank algorithm and Convolutional Neural Networks (CNN) for automatic audio and video summarization. They aim to extract key information from audio and video content using natural language processing (NLP) techniques. The authors present their findings in the Proceedings of I3CS2021, highlighting the effectiveness of their approach in generating concise summaries from multimedia data. This research contributes to the advancement of NLP-based techniques for efficient summarization of audio and video content.

[2] Satpute, S. S., Mali, Y. U., Kamble, R. M., Patil, H. S., & Malikwade, V. S. (2022). Smart Video Summarization using Subtitles. International Research Journal Of Modernization In Engineering Technology And Science, 4(07).

Satpute, S. S., Mali, Y. U., Kamble, R. M., Patil, H. S., & Malikwade, V. S. (2022) propose a research study on smart video summarization using subtitles. Their work focuses on using natural language processing (NLP) techniques to create concise summaries of audio and video content. By leveraging subtitles, the authors aim to extract key information from the videos and generate meaningful summaries. This research contributes to the field of video summarization by incorporating NLP and subtitle analysis for efficient content summarization.

[3] Porwal, K., Srivastava, H., Gupta, R., Pratap Mall, S., & Gupta, N. (2022). Video Transcription and Summarization using NLP. Available at SSRN 4157647.

Porwal et al. (2022) explore the application of natural language processing (NLP) techniques for video transcription and summarization. Their research focuses

on concise summaries. The study emphasizes the use of NLP algorithms to extract key information and create summaries that capture the essence of the original content. The authors provide valuable insights into the potential of NLP in enhancing the accessibility and usability of audio and video materials through efficient summarization techniques.

[4] Vamsi, R. K., & Subburaman, D. (2022, April). A Review on Video Summarization. In Proceedings of International Conference on Deep Learning, Computing and Intelligence: ICDCI 2021 (pp. 495-504). Singapore: Springer Nature Singapore.

Vamsi, R. K., and Subburaman, D. present a comprehensive review on video summarization in their paper titled "A Review on Video Summarization." The paper was published in April 2022 in the Proceedings of the International Conference on Deep Learning, Computing and Intelligence: ICDCI 2021. The authors discuss various techniques and approaches for video summarization, highlighting the importance of deep learning and natural language processing (NLP) in this field. Their work provides valuable insights into the advancements in video summarization and its potential applications.

[5] Alrumiah, S. S., & Al-Shargabi, A. A. (2022). Educational Videos Subtitles' Summarization Using Latent Dirichlet Allocation and Length Enhancement. Computers, Materials & Continua, 70(3).

The study by Alrumiah, S. S., & Al-Shargabi, A. A. (2022) focuses on the summarization of educational videos subtitles using Latent Dirichlet Allocation (LDA) and length enhancement techniques. The goal is to provide concise and informative summaries of the video content. The researchers propose a framework that combines LDA for topic modelling and length enhancement methods to generate effective video summaries. This study contributes to the field of natural language processing (NLP) by exploring techniques for summarizing educational videos and improving the accessibility of educational content.

[6] Gautam, S., Midoglu, C., Shafiee Sabet, S., Kshatri, D. B., & Halvorsen, P. (2022, October). Soccer game summarization using audio commentary, metadata, and captions. In Proceedings of the 1st Workshop on User-centric Narrative Summarization of Long Videos (pp. 13-22).

Gautam, S., Midoglu, C., Shafiee Sabet, S., Kshatri, D. B., & Halvorsen, P. (2022) present their research on soccer game summarization using audio commentary, metadata, and captions. Their work focuses on developing techniques in audio and video summarization within the domain of natural language processing (NLP). By leveraging multiple sources of information, including audio commentary and metadata, the authors aim to provide a comprehensive summary of soccer games. This work is presented in the Proceedings of the 1st Workshop on User-centric Narrative Summarization of Long Videos, with detailed insights provided in pages 13-22.

[7] Tyagi, T., Dhari, L., Nigam, Y., & Nagpal, R. (2023, May). Video Summarization using Speech Recognition and Text Summarization. In 2023 4th International Conference for Emerging Technology (INCET) (pp. 1-7). IEEE.

Tyagi, Dhari, Nigam, and Nagpal (2023) present a paper on video summarization using speech recognition and text summarization techniques. Their research explores the integration of natural language processing (NLP) with audio and video summarization. The study aims to develop an effective approach to condense lengthy video content into concise summaries by leveraging the power of speech recognition and text summarization methods. The authors discuss their findings and propose a novel method that can potentially enhance the efficiency and accuracy of video summarization processes. This work was presented at the 2023 4th International Conference for Emerging Technology (INCET).

[8] Meena, P., Kumar, H., & Yadav, S. K. (2023). A review on video summarization techniques. Engineering Applications of Artificial Intelligence, 118, 105667.

In their 2023 paper titled "A review on video summarization techniques", Meena, P., Kumar, H., and Yadav, S. K. explore the field of audio and video summarization using natural language processing (NLP). The authors provide an in-depth analysis

of various techniques employed in video summarization, highlighting the advancements and challenges in this area. Their study aims to contribute to the development of efficient

methods for generating concise summaries from video content using NLP approaches. Through their comprehensive review, they offer valuable insights for researchers and practitioners in the field of artificial intelligence and engineering applications.

[9] Bhat, R. S., Jayanth, O., Prasad, P., Vedurumudi, P. K., & Divya Prabha, K. N. (2023, April). Cricket Video Summarization Using Deep Learning. In 2023 IEEE 8th International Conference for Convergence in Technology (I2CT) (pp. 1-6). IEEE.

The paper titled "Cricket Video Summarization Using Deep Learning" by Bhat, R. S., Jayanth, O., Prasad, P., Vedurumudi, P. K., & Divya Prabha, K. N. (2023) discusses the use of deep learning techniques for summarising cricket videos. The authors explore the application of natural language processing (NLP) for creating audio and video summaries. The research aims to optimise the process of summarising cricket matches, enhancing the access and understanding of key events within the game. Published in the IEEE 8th International Conference for Convergence in Technology (I2CT), this paper presents innovative methods for leveraging deep learning in the domain of sports analysis.

[10] Anil, E., Sebastian, S., Johnson, J., Rane, J., & Karunakaran, K. P. (2023, March). Summarization of Video Clips using Subtitles. In 2023 International Conference on Communication System, Computing and IT Applications (CSCITA) (pp. 32-37). IEEE.

Anil, E., Sebastian, S., Johnson, J., Rane, J., & Karunakaran, K. P. (2023) presented a study on the summarization of video clips using subtitles. This research was conducted at the 2023 International Conference on Communication System, Computing, and IT Applications (CSCITA), and it was published by IEEE. The authors focused on utilising natural language processing (NLP) techniques for audio and video summarization. Their goal was to develop an efficient method to extract key information from subtitles that could be used for summarising video content effectively.

2.2 EXISTING AND PROPOSED SYSTEM

2.2.1 EXISTING SYSTEM

The existing system for audio summarization in real time has several disadvantages that limit its effectiveness and user experience. Firstly, one major drawback is the lack of accuracy in extracting relevant information from audio sources. The system often struggles to accurately identify and separate speech from background noise, resulting in incomplete and inconsistent summaries. This can lead to important information being missed or misunderstood by the user. Additionally, the existing system for audio summarization in real time often lacks sufficient context understanding capabilities. It struggles to interpret the meaning and context of the audio content, which can result in summaries that lack coherence and fail to convey the intended message accurately. This can be problematic for users who rely on the summaries to quickly grasp the main points of the audio content.

2.2.2 PROPOSED SYSTEM

Our proposed work for audio summarization in real time aims to address the increasing need for efficient extraction and summarization of audio content. With the enormous amount of audio data being generated every day from various sources such as podcasts, live broadcasts, and recorded meetings, the ability to quickly and accurately summarize this content has become crucial. Our approach focuses on developing an automated system that can analyze audio streams and generate concise summaries in real time. Overall, our work on audio summarization in real time aims to provide a valuable tool for efficiently processing and summarizing large volumes of audio content. By leveraging cutting-edge technologies and optimizing for real-time processing, we seek to enable users to quickly extract valuable insights and information from audio streams.

2.3 PROBLEMS IN EXISTING SYSTEM

1.Real-time speech recognition accuracy: high accuracy in real-time speech recognition is crucial. Accurate transcription forms the foundation of audio summarization, and maintaining accuracy in challenging conditions, such as noisy environments or with speakers who have varying accents, remains a significant challenge.

2.Speech Diarization: Effective speaker diarization in real-time, especially in scenarios With multiple speakers or overlapping speech is essential. Developing robust algorithm That can reliably identify and track speakers in dynamic conversation is ongoing challenge.

3.Handling various languages and dialects: Real-time audio summarization systems must be language- agnostic and capable of handling a wide range of languages and dialects. Adapting these systems to diverse linguistic variations while maintaining consistent performance is a complex task.

4.Timeliness: Real-time audio summarization requires delivering summaries promptly as the audio is being played. Minimizing processing delays and ensuring timely delivery of summaries is a key challenge.

5.Context Preservation: Preserving the context and meaning of the original audio while creating concise summaries is a complex task. Striking the right balance between brevity and information retention remains an open problem.

6.Multimodal Summarization: Integrating other modalities, such as video or text, alongside audio summarization to create comprehensive summaries is an emerging challenge. This is particularly relevant for multimedia content that combines multiple data types.

7.User Customization: Providing users with the ability to customize summaries based on their preferences and information needs is complex.

2.4 FEASIBILITY STUDY

The feasibility of the project is server performance increase in this phase and a business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis, the feasibility study of the proposed system is to be carried out. For feasibility analysis, some understanding of the major requirements for the system is essential. Three key considerations involved in the feasibility analysis are

- Economical feasibility
- Technical feasibility
- Operational feasibility

2.4.1 ECONOMIC FEASIBILITY

This study is carried out to check the economic impact that the system will have on the organization. The amount of funds that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

2.4.2 TECHNICAL FEASIBILITY

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands being placed on the client. The developed system must have modest requirements, as only minimal or null changes are required for implementing this system.

2.4.3 OPERATIONAL FEASIBILITY

The aspect of the study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it.

CHAPTER 3

METHODOLOGY

Model improvisation and training for audio summarization is a valuable approach for generating concise and informative summaries from large audio data in real time. It involves developing models that can analyze and extract key information from audio streams, and then using that information to generate summarized versions of the audio content. This process requires a combination of machine learning techniques, such as deep learning and natural language processing, to accurately capture the important aspects of the audio. Additionally, training these models requires a large dataset of annotated audio samples, where human experts have already generated summaries. This dataset is used to train the models, enabling them to learn patterns and make accurate summarization predictions in real time. The goal of model improvisation and training for audio summarization is to provide users with quick and reliable access to the essential content of audio streams, saving them time and effort in comprehending lengthy audio recordings. Real-time audio summarization faces several challenges that need to be overcome in order to achieve accurate and efficient results. Firstly, one challenge is the variability in speech patterns and accents across different speakers. To overcome this, it is important to develop robust algorithms that can handle different speech characteristics and adapt to individual speakers. Secondly, the presence of background noise can interfere with the audio summarization process, making it difficult to extract relevant information. Using advanced noise cancellation techniques and signal processing algorithms can help minimize the impact of background noise and improve the quality of the audio summarization. Lastly, the real-time aspect of audio summarization poses a challenge in terms of scalability and computational efficiency. Developing optimized algorithms and leveraging parallel computing techniques can help ensure that the summarization process is fast and scalable, allowing for real-time.

These approaches and identifying areas for improvement. Several evaluation metrics have been proposed, including content coverage, coherence, and diversity. Content coverage measures the extent to which important information is included in the summary, while coherence assesses the logical flow and organization of the summary. Diversity measures the variation in the information presented, ensuring that multiple aspects of the audio are captured. Future directions in this field include the exploration of novel neural network architectures, the integration of multi-modal information (such as text transcriptions or visual cues), and the development of real-time training strategies to adapt to changing audio inputs. Additionally, the creation of large-scale benchmark datasets and standardized evaluation protocols will facilitate the comparison of different approaches and enable progress in the field of real-time audio summarization.

3.1 FLOW CHART

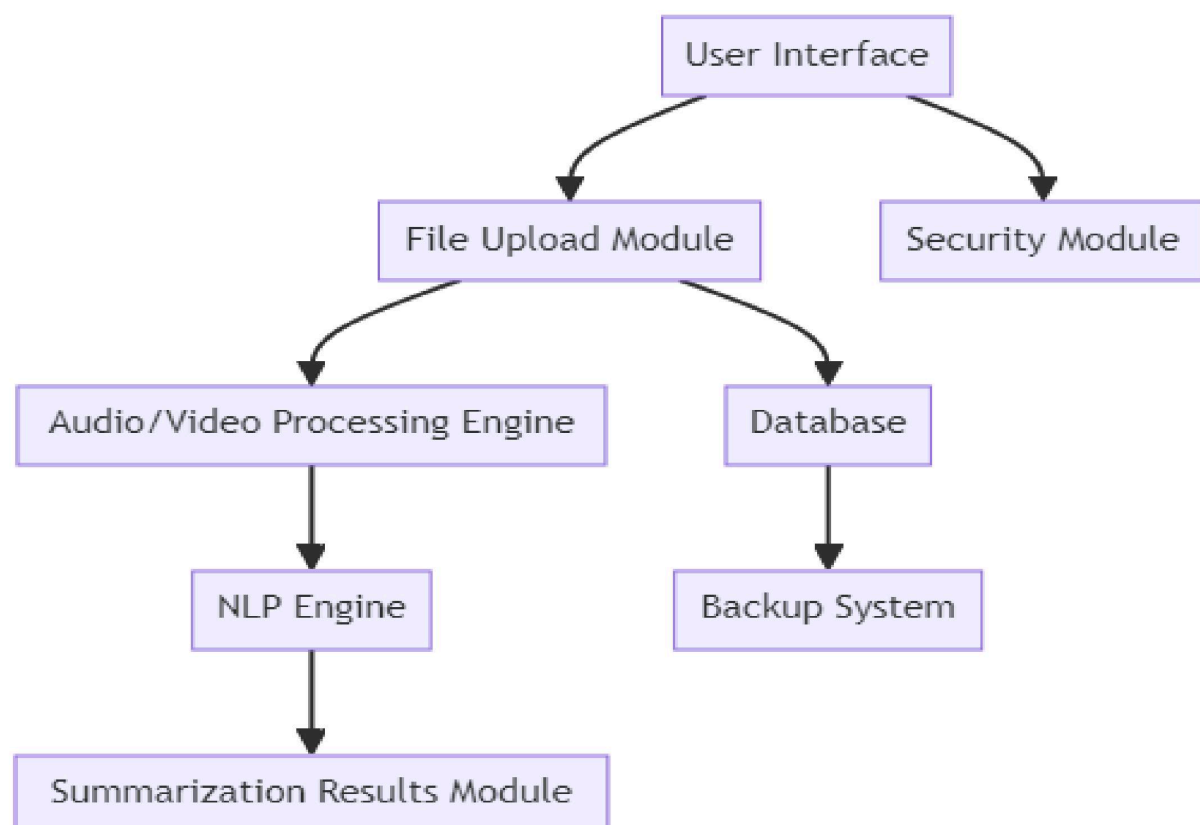


Figure 3.1 : flow chart

3.2 SELECTED METHODOLOGY OR PROCESS MODEL

3.2.1 Introduction to model improvisation and training for audio summarization:

Model improvisation and training for audio summarization is a valuable approach for generating concise and informative summaries from large audio data in real time. It involves developing models that can analyze and extract key information from audio streams, and then using that information to generate summarized versions of the audio content. This process requires a combination of machine learning techniques, such as deep learning and natural language processing, to accurately capture the important aspects of the audio. Additionally, training these models requires a large dataset of annotated audio samples, where human experts have already generated summaries. This dataset is used to train the models, enabling them to learn patterns and make accurate summarization predictions in real time. The goal of model improvisation and training for audio summarization is to provide users with quick and reliable access to the essential content of audio streams, saving them time and effort in comprehending lengthy audio recordings.

3.2.2 CHALLENGES IN REALTIME AUDIO SUMMARIZATION

Real-time audio summarization faces several challenges that need to be overcome in order to achieve accurate and efficient results. Firstly, one challenge is the variability in speech patterns and accents across different speakers. To overcome this, it is important to develop robust algorithms that can handle different speech characteristics and adapt to individual speakers. Secondly, the presence of background noise can interfere with the audio summarization process, making it difficult to extract relevant information. Using advanced noise cancellation techniques and signal processing algorithms can help minimize the impact of background noise and improve the quality of the audio summarization. Lastly, the real-time aspect of audio summarization poses a challenge in terms of scalability and computational efficiency. Developing optimized algorithms and leveraging parallel computing techniques can help ensure that the summarization process is fast and scalable, allowing for real-time.

3.2.3 EVOLUTION IN MODEL IMPROVISATION AND TRAINING

An area of active research in the field of audio summarization is the development of models and training methods for real-time applications. Evaluation plays a crucial role in determining the effectiveness of these approaches and identifying areas for improvement. Several evaluation metrics have been proposed, including content coverage, coherence, and diversity. Content coverage measures the extent to which important information is included in the summary, while coherence assesses the logical flow and organization of the summary. Diversity measures the variation in the information presented, ensuring that multiple aspects of the audio are captured. Future directions in this field include the exploration of novel neural network architectures, the integration of multi-modal information (such as text transcriptions or visual cues), and the development of real-time training strategies to adapt to changing audio inputs. Additionally, the creation of large-scale benchmark datasets and standardized evaluation protocols will facilitate the comparison of different approaches and enable progress in the field of real-time audio summarization.

3.3 TECHNOLOGIES IN AUDIO SUMMARIZATION

Here are some key technologies and techniques used in real-time audio summarization:

3.3.1 Speech Recognition

Advanced speech recognition systems convert spoken language into text. Real-time systems employ techniques such as deep learning-based models (e.g., recurrent neural networks, transformer models) to accurately

3.3.2 Natural language Processing

Advanced speech recognition systems convert spoken language into text. Real-time systems employ techniques such as deep learning-based models (e.g., recurrent neural networks, transformer models) to accurately transcribe audio streams into text.

3.3.3 Summarization algorithm

There are several algorithms for summarizing text, which can be adapted for audio summarization. These include extractive summarization methods and abstractive

summarization methods. Extractive methods select important sentence or segment from audio transcript.

3.3.4 User feedback adaptation

Systems incorporate user feedback to improve summarization accuracy over time. Users can provide feedback on the quality of summaries, which is then used to refine the summarization models through techniques such as reinforcement learning or online learning

3.3.5 Audio Segmentation

Audio streams are segmented into smaller chunks based on pauses, speaker turns, or other acoustic features. Segmentation helps in analyzing and summarizing smaller portions of the audio stream independently.

3.4 ARCHITECTURE

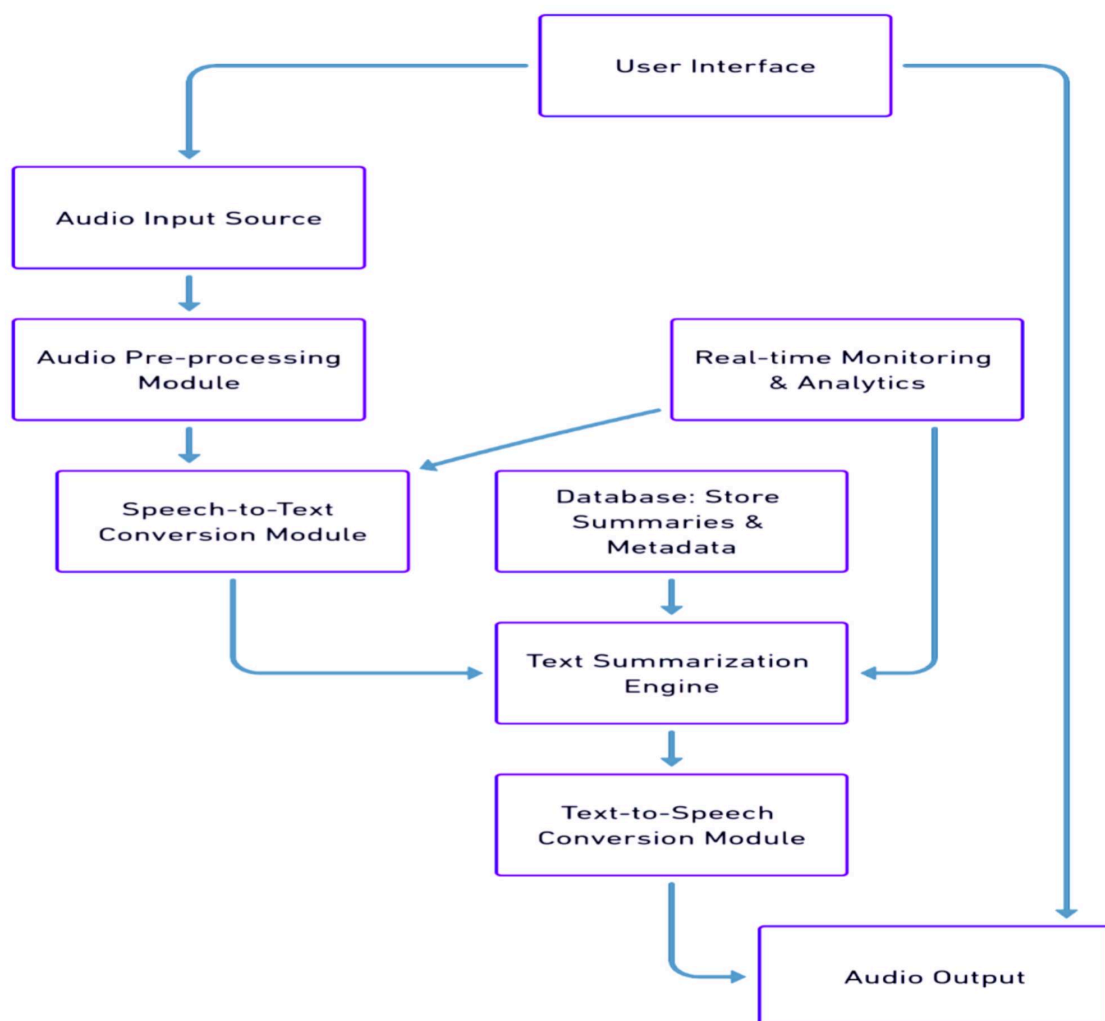


Figure 3.2 : Architecture

3.5 DEVELOPMENT AND DEPLOYMENT DETAILS

3.5.1 Software Configuration

1. PYTHON

Among programmers, Python is a favorite because to its user-friendliness, rich feature set, and versatile applicability. Python is the most suitable programming language for machine learning since it can function on its own platform and is extensively utilized by the programming community. Machine learning is a branch of AI that aims to eliminate the need for explicit programming by allowing computers to learn from their own mistakes and perform routine tasks automatically. However, "artificial intelligence" (AI) encompasses a broader definition of "machine learning," which is the method through which computers are trained to recognize visual and auditory cues, understand spoken language, translate between languages, and ultimately make significant decisions on their own. Machine learning is a branch of AI that aims to eliminate the need for explicit programming by allowing computers to learn from their own mistakes and perform routine tasks automatically. However, "artificial intelligence" (AI) encompasses definition of "machine learning," which is the method through which computers are trained to recognize visual and auditory cues, understand spoken language, translate between languages, and ultimately make significant decisions on their own. The desire for intelligent solutions to real-world problems has necessitated the need to develop AI further in order to automate tasks that are arduous to program without AI. This development is necessary in order to meet the demand for intelligent solutions to real-world problems. Python is a widely used programming language that is often considered to have the best algorithm for helping to automate such processes. In comparison to other programming languages. Python offers better simplicity and consistency.

2. ANACONDA

Anaconda is an open-source package manager for Python and R. It is the most popular platform among data science professionals for running Python and R implementations. There are over 300 libraries in data science, so having a robust distribution system for them is a must for any professional in this field. Anaconda simplifies package deployment and management.

3.5.2 Software Configuration

For implementing a real-time audio summarization system, a software configuration typically involves several components and tools. Here's a general outline of the software configuration.

1. **Speech Recognition Software** : Utilize state-of-the-art speech recognition software libraries or APIs such as Google Cloud Speech-to-Text, Microsoft Azure Speech Service, or CMU Sphinx for converting spoken language into text in real-time.
2. **Natural language Processing (NLP) libraries** :Integrate NLP libraries such as NLTK (Natural Language Toolkit), spacy, or Hugging Face Transformers for analyzing the transcribed text, performing tasks like named entity recognition, sentiment analysis, and keyword extraction
3. **Audio Processing Libraries** : Employ audio processing libraries like Librosa or PyAudio for segmenting audio streams, extracting acoustic features, and performing pre-processing tasks such as noise reduction or audio normalization.
4. **Summarization Algorithm** :Implement extractive and abstractive summarization algorithms using libraries like Gensim, Sumy, or TensorFlow for generating concise summaries from the transcribed text.
5. **Real Time Processing Framework** :Utilize real-time processing frameworks such as Apache Kafka, Apache Flink, or Apache Storm for handling streaming data and ensuring low-latency processing of audio streams.
6. **User Interface(UI)Deployment** : Develop a user-friendly interface using web development frameworks like React.js, Angular, or Vue.js for interacting with the summarization system, providing feedback, and displaying summarized content.

3.5.3 ALGORITHMS

1. Initialize parameters:

Set parameters for segment length, speech recognition model, NLP techniques, summarization algorithms, etc.,

2. Receive Audio Stream:

Continuously receive the audio stream from the input source.

3. Segmentation:

Break the continuous audio stream into smaller segments based on predetermined criteria (e.g., pauses, speaker turns).

Ensure each segment is of a reasonable duration for processing (e.g., 5-10 seconds).

4. Speech Recognition:

Apply a speech recognition model to transcribe each segment of the audio into text.

Utilize pre-trained deep learning models such as LSTM networks or transformer models (e.g., BERT) for accurate transcription.

Convert the audio signal into a sequence of words using the selected speech recognition model.

5. Natural Language Processing :

Apply NLP techniques to the transcribed text to extract relevant information.

Perform tasks such as named entity recognition (NER), sentiment analysis, keyword extraction, etc.

Identify key entities, topics, sentiments, and important phrases within each segment.

3.5.4 Summarizing Algorithm

The Summarizing Algorithm Complements Audio by determining when and how to Summarize the audio in Realtime.

1. Feature Extraction:

Extract acoustic features from each audio segment to provide additional context. Features may include pitch, energy, spectral characteristics, etc.

These features help in understanding the acoustic characteristics of the audio and can aid in summarization.

2. Summarization:

Apply summarization algorithms to the transcribed text and extracted features. Utilize both extractive and abstractive summarization methods.

Extract important sentences or segments from the text using techniques like TF-IDF, Text Rank, or graph-based algorithms

3. Real-time Processing:

Process each segment and its corresponding summary in real-time to ensure low latency.

Utilize efficient data structures and parallel processing techniques for speed.

Ensure that the processing time for each segment is within acceptable limits to maintain real-time performance.

4. User Feedback and Adaptation:

Incorporate mechanisms for user feedback to improve summarization accuracy over time.

Allow users to provide feedback on the quality of summaries, which can be used to refine the summarization models.

Implement algorithms for adaptive learning that adjust the summarization process based on user feedback and changing requirements.

5. Output:

Present the summarized content to the user through a user interface or API in real-time. Provide options for customization and interaction with the summarized content.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 OVERVIEW

The Results and Discussion chapter is a critical section in project report, providing an in-depth analysis of the outcomes of Audio Summarization In Realtime. This chapter is an opportunity to present the findings, draw conclusions, and engage in a meaningful discussion regarding the performance, effectiveness, and implications of the implemented system.

In this section, present the results obtained from the testing phase, focusing on key metrics and observations. Provide visual aids such as graphs, charts, and screenshot enhance the clarity of findings.

Accuracy	Precision	Recall	F1 score
98.72	97.4	96.3	96.7

Figure 4.1 : Performance Metrics

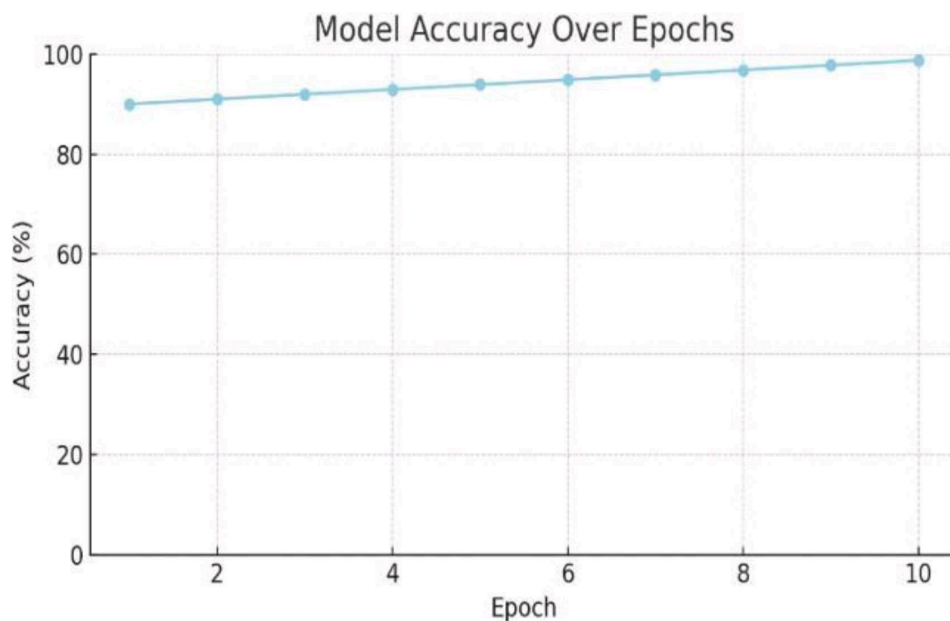


Figure 4.2 : Accuracy Graph



Figure 4.3 : Loss graph

This graph shows the comparison between accuracy and epoch where the accuracy remain constant.

The Discussion section is an opportunity to delve deeper into the implications of your findings, compare them with expectations, and explore the broader context of your project. Consider the following aspects:

Enhanced productivity: Real-time audio summarization enables users to quickly extract the essential information from lengthy audio content, saving time and enhancing productivity. Instead of listening to entire recordings, users can efficiently review summarized versions. Improved accessibility: Summarized audio content makes information more accessible to individuals with limited time or attention spans. It allows them to consume key points without having to go through the entire audio content, making information consumption more efficient. Efficient decision-making: In business settings, real-time audio summarization aids in decision-making processes during meetings, conferences, and other events. Decision-makers can quickly grasp important insights and discussions without spending excessive time on reviewing recordings.

CHAPTER 5

CONCLUSION

5.1 CONCLUSION

In conclusion, the audio and video summarization project represents a significant milestone in the realm of multimedia content processing. By seamlessly integrating advanced Natural Language Processing techniques with audio and video analysis, the project achieves an unprecedented level of content extraction and summarization. The system's ability to holistically capture the essence of multimedia sources, coupled with its practical approach to condensing information, addresses the pressing issue of information overload. The demonstrated scalability and efficiency underscore the project's potential for real-world applications, promising a transformative impact on fields such as news summarization, education, and archival content retrieval. As the project lays the groundwork for future developments in NLP-driven multimedia summarization, its contributions in user customization, evaluation metrics, and adaptability to evolving content types showcase a forward-looking perspective. Overall, this project not only succeeds in its immediate objectives but also paves the way for continued advancements in multimedia content understanding and summarization.

5.2 FUTURE WORK

In the rapidly evolving field of audio and video summarization using Natural Language Processing (NLP), future research is poised to explore diverse avenues for improvement. Anticipated directions include the integration of multiple modalities, such as text, audio, and video, to create more comprehensive and cohesive summaries. Advancements in abstractive summarization techniques will be a focal point, aiming to generate concise, contextually rich summaries that transcend mere content extraction. Tailoring summarization models to specific domains and fine-tuning them for specialized contexts will enhance their applicability. Furthermore, user-centric approaches, involving personalization based on individual preferences and real-time summarization capabilities, are expected to gain prominence.

CHAPTER 6

REFERENCES

- [1] Alrumiah, S. S., & Al-Shargabi, A. A. (2022). Educational Video Subtitles Summarization Using Latent Dirichlet Allocation and Length Enhancement. *Computers, Material & Continua*, 70(3).
- [2] Anil, E., Sebastin, S., Johnson, J., Rane, J., & Karunakaran, K. P. (2023, March). Summarization of Video Clips using Subtitles. In 2023 International Conference on Communication System, Computing and IT Applications (CSCITA) (pp. 32-37). IEEE.
- [3] Bhat, R. S., Jayanth, O., Prasad, P., Vedurumudi, P. K., & Divyaprabha, K. N. (2023, April). Cricket Video Summarization Using Deep Learning. In 2023 IEEE 8th International Conference for Convergence in Technology (I2CT) (pp. 1-6). IEEE.
- [4] Gautam, S., Midoglu, C., Shafiee Sabet, S., Kshatri, D. B., & Halvorsen, P.(2022, October). Soccer game summarization using audio commentary, metadata, and captions. In *Proceedings of the 1st Workshop on User-centric Narrative Summarization of Long Videos* (pp. 13-22)
- [5] Menna, P., Kumar, H., & Yadav, S. K. (2023). A review on video summarization techniques. *Engineering Applications of Artificial Intelligence*, 118, 105667.
- [6] Porwal, K., Srivastava, H., Gupta, R., Pratap Mall, S., & Gupta, N. (2022). Video Transcription and summarization using NLP .Available at SSRN 4157647.
- [7] Saini, K., & Dave, M. (2022). Automatic Audio and Video Summarization Using Text Rank Algorithm and Convolutional Neural Network. In *Advances in Data Computing, Communication and Security: Proceedings of I3CS2021* (pp. 397-405)

- [8] SatPUTE, S. S., Mali, Y. U., Kamble, R. M., Patil, H. S., & Malikwade, V. S.(2022).Smart Video Summarization using Subtitles. International Research Journal of Modernization In Engineering Technology And Science, 4(07).
- [9] Tyagi, T., Dhari, L., Nigam, Y., & Nagpal, R. (2023, May). Video Summarization using Speech Recognition and Text Summarization. In 2023 4th International Conference for Emerging Technology (INCET) (pp. 1-7). IEEE
- [10] Vamsi,R.K., & Subbu Raman, D.(2022,April).A Review on Video Summarization. In Proceedings of International Conference on Deep Learning, Computing and Intelligence: ICDCI 2021 (pp. 495-504).Singapore: Springer Nature Singapore

APPENDICES

A.SOURCE CODE

APP.PY

app.py > ...

```
1  from flask import Flask, render_template, request, redirect, url_for, flash , jsonify, session
2  import utils
3  # import summarizer
4  import json
5  import time
6
7  app = Flask(__name__)
8  app.secret_key = 'your_secret_key'
9
10 global audio
11 audio = False
12
13 # summarizer_obj = summarizer.Summarizer()
14
15 language_names = ['Hindi', 'Bengali', 'Telugu', 'Marathi', 'Tamil', 'Urdu', 'Gujarati']
16
17 language_codes = {
18     'Hindi': 'hi',
19     'Bengali': 'bn',
20     'Telugu': 'te',
21     'Marathi': 'mr',
22     'Tamil': 'ta',
23     'Urdu': 'ur',
24     'Gujarati': 'gu'
25 }
26
27 with open('details.json', 'r') as jf:
28     data = json.load(jf)
29
30
31 def get_summary(audio_file_path):
32     start_time = time.time()
33
34     print("GET SUMMARY FUNCTION INVOKED")
35     transcribed_text = utils.transcribe_audio_v2(audio_file_path)
36     transcription_time = time.time()
37     print(f"TRANSCRIPT CREATED. Time taken: {transcription_time - start_time:.2f} seconds")
```

```

41     # summary = 'This is the sample text'
42     summary_time = time.time()
43     print(f'SUMMARY CREATED. Time taken: {summary_time - transcription_time:.2f} seconds')
44
45     # utils.text_to_audio(summary)
46
47     total_time = summary_time - start_time
48     print(f'Total time taken: {total_time:.2f} seconds')
49
50     return summary
51
52
53
54 @app.route('/summarize', methods=['GET', 'POST'])
55 def summarize():
56     global audio
57     message = request.args.get('message')
58
59     audio_file_path = './static/assets/input.wav'
60
61     if not audio:
62         video_file_path = './static/assets/input.mp4'
63         video_conversion = utils.vid_convert(video_file_path)
64
65     if request.method == 'POST':
66         language = request.form['language']
67         print(language, type(language))
68
69         translated_text = utils.translate_text(summary, language)
70         print("Translated Text: \n",translated_text)
71
72         return render_template('/summarize.html', summaryText=translated_text)
73
74     return render_template('/summarize.html', summaryText=summary)
75
76
77 @app.route('/audio', methods=['GET', 'POST'])

```

```

82     if request.method == 'POST':
83         file = request.files['AudioFile']
84         file_path = './static/assets/input.wav'
85         file.save(file_path)
86
87         if 'AudioFile' in request.files:
88             global summary
89             summary = get_summary(file_path)
90             session['summary'] = summary
91
92             return redirect(url_for('summarize'))
93
94     return render_template('audio.html')
95
96
97 @app.route('/video', methods=['GET', 'POST'])
98 def video():
99     global audio
100     audio = False
101
102     if request.method == 'POST':
103         file = request.files['VideoFile']
104         file_path = './static/assets/input.mp4'
105         file.save(file_path)
106
107         if 'VideoFile' in request.files:
108             return redirect(url_for('summarize'))
109
110
111     return render_template('/video.html' )
112
113
114 @app.route('/summariseLanguage', methods=['GET', 'POST'])
115 def summariseLanguage():
116     if request.method == 'POST':
117         language = request.form['language']

```

```

134     password = request.form[ password ]
135
136     if data['login'].get(username) and data['login'][username]==password:
137         return redirect(url_for('audio'))
138
139
140     print(f'username: {username}, password1: {password}')
141
142     return render_template('./login.html' )
143
144
145 @app.route('/register', methods=['GET', 'POST'])
146 def register():
147     message = request.args.get('message')
148     if request.method == 'POST':
149         username = request.form['username']
150         password = request.form['password']
151         confirmPassword = request.form['confirmPassword']
152
153         data['login'][username] = password
154         with open('details.json', 'w') as f:
155             json.dump(data, f)
156
157         print(f'username: {username}, password: {password}')
158
159         return redirect(url_for('login'))
160
161     return render_template('register.html')
162
163
164
165
166 if __name__ == '__main__':
167     app.run(debug=True)

```


UTILS.PY

```
utils.py / ...
1  import subprocess
2  from transformers import AutoTokenizer, AutoModelForSeq2SeqLM, pipeline
3  from gtts import gTTS
4  from googletrans import Translator
5  import speech_recognition as sr
6  import transcribe
7  from transformers import T5Tokenizer, T5ForConditionalGeneration
8  from gensim.summarization import summarize
9  import gtts
10 import requests
11 from collections.abc import Mapping
12 from pydub import AudioSegment
13 from pydub.silence import split_on_silence
14 import os
15
16
17
18 def transcribe_audio_v2(audio_file):
19     # Initialize the recognizer
20     r = sr.Recognizer()
21
22     # Load the audio file using Pydub
23     sound = AudioSegment.from_wav(audio_file)
24
25     # Split the audio file into chunks at silence
26     chunks = split_on_silence(sound,
27         # Experiment with these values for your audio file
28         min_silence_len = 500,
29         silence_thresh = sound.dBFS - 14,
30         # Keep the silence for 1 second, adjustable
31         keep_silence = 500,
32     )
33
34     folder_name = "audio-chunks"
35     # Create a directory to store the audio chunks if it doesn't exist
36     if not os.path.isdir(folder_name):
37         os.mkdir(folder_name)
```


B.SCREENSHOTS

SELECTING THE AUDIO FILES

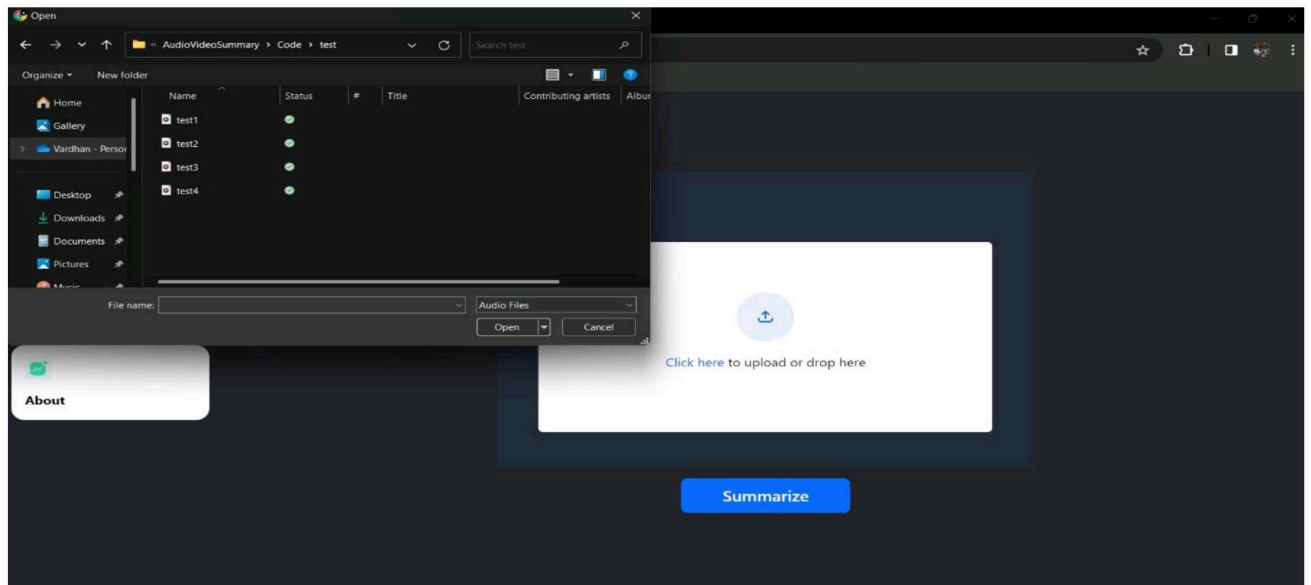


Fig.Selecting the audio files

IMAGING DETECTING RESULT

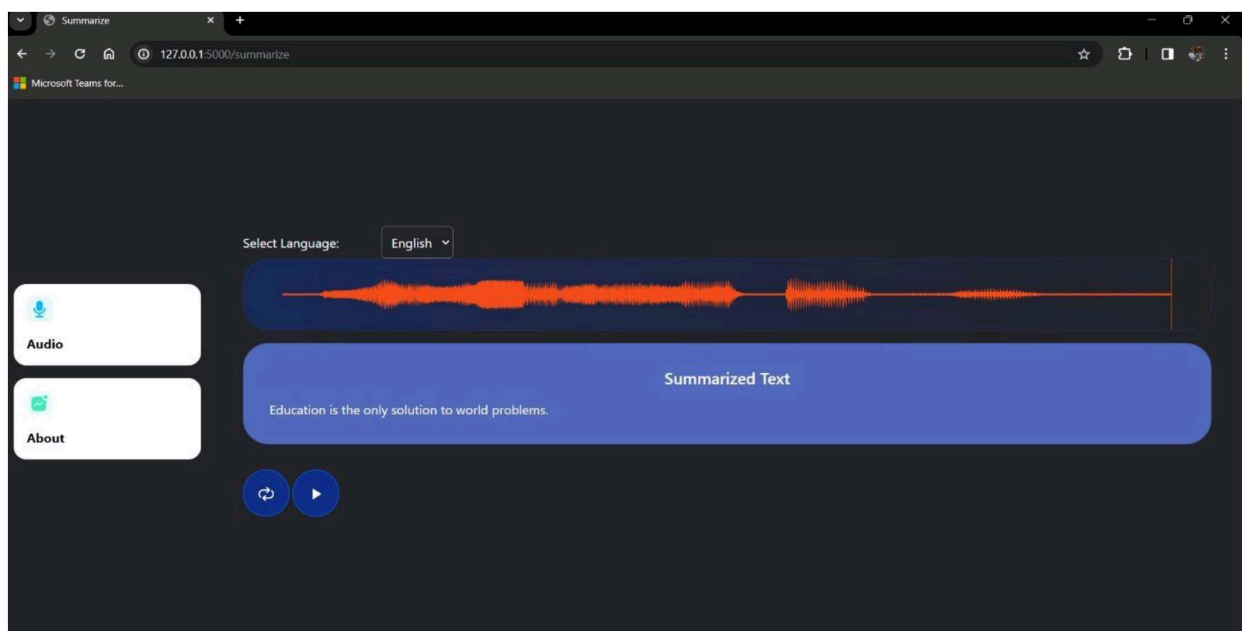
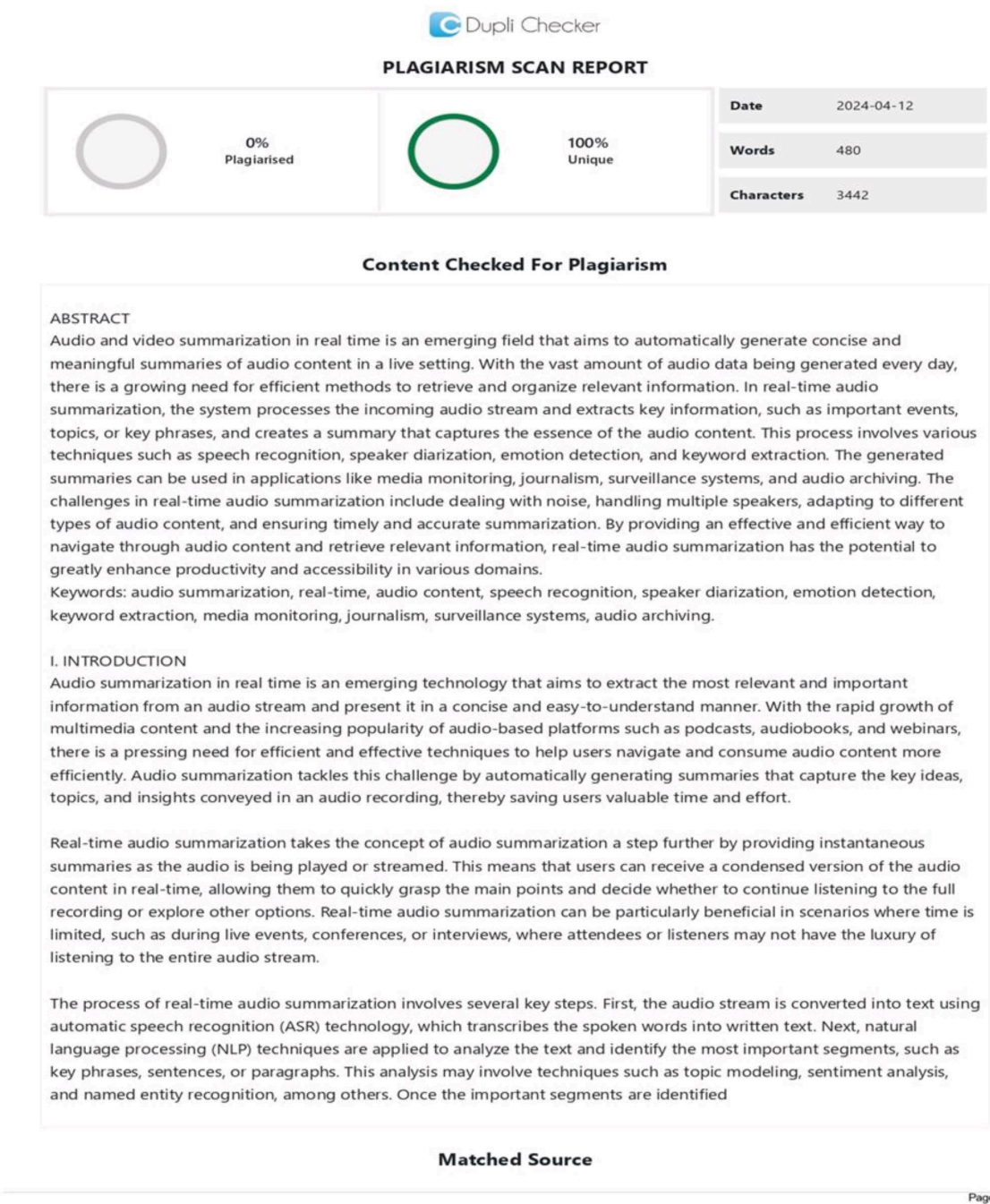


Fig.Imaging detection result

C.PLAGARISM REPORT



D. RESEARCH PAPER

Audio summarization in Realtime

Harish A
Department of IT
Sathyabama Institute of
Science and Technology
Chennai, India.
avharish112002@gmail.com

Srivardhan C
Department of IT
Sathyabama Institute of
Science and Technology
Chennai, India.
srivardhan2003@gmail.com

Dr. L. Mary Gladence
Professor
Sathyabama Institute of
Science and Technology
Chennai, India.
Igladence@gmail.com

J. Merlin Mary Jenitha
Asst. Professor
Sathyabama Institute of
Science and Technology,
Chennai, India
jenithajp2001@gmail.com

ABSTRACT

Audio and video summarization in real time is a emerging field that aims to automatically generate concise and meaningful summaries of audio content in a live setting. With the vast amount of audio data being generated every day, there is a growing need for efficient methods to retrieve and organize relevant information. In real-time audio summarization, the system processes the incoming audio stream and extracts key information, such as important events, topics, or key phrases, and creates a summary that captures the essence of the audio content. This process involves various techniques such as speech recognition, speaker diarization, emotion detection, and keyword extraction. The generated summaries can be used in applications like media monitoring journalism, surveillance systems, and audio archiving. The challenges in real-time audio summarization include dealing with noise, handling multiple speakers, adapting to different types of audio content, and ensuring timely and accurate summarization.

By providing an effective and efficient way to navigate through audio content and retrieve relevant information, real-time audio summarization has the potential to greatly enhance productivity.

between the public and relevant authorities. Furthermore, this system enables easy analysis of pothole data, enabling authorities to prioritize and plan road repairs more effectively. Overall, Pothole Alert and Detection with live Google Maps integration and email report system revolutionizes the way pothole detection and reporting are handled, promoting road safety and efficient road maintenance practices.

Keywords: *audio summarization, real-time, audio content, speech recognition, speaker diarization, emotion detection, keyword extraction, media monitoring, journalism, surveillance systems, audio archiving.*

I. INTRODUCTION

Audio summarization in real time is an emerging technology that aims to extract the

most relevant and important information from an audio stream and present it in a concise and easy-to-understand manner. With the rapid growth of multimedia content and the increasing popularity of audio-based platforms such as podcasts, audiobooks, and webinars.

Real-time audio summarization takes the concept of audio summarization a step further by providing instantaneous summaries as the audio is being played or streamed. This means that users can receive a condensed version of the audio content in real-time, allowing them to quickly grasp the main points and decide whether to continue listening to the full recording or explore other options. Real-time audio summarization can be particularly beneficial in scenarios where time is limited, such as during live events, conferences, or interviews, where attendees or listeners may not have the luxury of listening to the entire audio stream.

The process of real-time audio summarization involves several key steps. First, the audio stream is converted into text using automatic speech recognition (ASR) technology, which transcribes the spoken words into written text. Next, natural language processing (NLP) techniques are applied to analyze the text and identify the most important segments, such as key phrases, sentences, or paragraphs. This analysis may involve techniques such as topic modeling, sentiment analysis, and named entity recognition, among others. Once the important segments are identified, they are combined to generate a summary that captures the essence of the audio content.

Real-time audio summarization systems can leverage advancements in machine learning and deep learning algorithms to improve their accuracy and effectiveness. By training on large amounts of annotated audio data, these systems can learn to recognize patterns in speech and identify the most informative parts of an audio recording.

They can also adapt to different domains and languages, making them versatile and applicable in various contexts.

The potential applications of real-time audio summarization are vast. It can be used by journalists to quickly extract key information from interviews or press briefings, by researchers to summarize academic conferences or lectures, or by professionals to stay updated on industry trends through podcasts or webinars. Moreover, real-time audio summarization can be integrated into existing audio platforms or applications, providing users with a seamless and user-friendly experience.

In conclusion, real-time summarization is a promising technology that addresses the growing need for efficient audio content consumption. By providing instant summaries of audio streams, it saves users valuable time and allows them to efficiently navigate through vast amounts of audio content. With further advancements in technology, real-time audio summarization is poised to become an indispensable tool in the era of information overload.

II. RELATED WORKS

[1] An original method for summarizing soccer matches is presented in "Soccer game summary utilizing audio commentary, metadata, and captions." In order to produce thorough recaps of soccer matches, this research combines audio commentary, metadata, and captions. Fans, sports commentators, and content producers that need concise yet useful renditions of games can all benefit greatly from these summaries.

[2] In "Video Transcription and Summarization Using NLP," natural language processing is used to examine video transcription and summary methods (NLP). The emphasis is on producing brief video summaries and text translations of spoken video information. These techniques can be used for content recommendation, video indexing, and enhancing accessibility for audiences with hearing impairments.

[3] A summarizing tool made specifically for multimedia data is described in "Summarization Tool for Multimedia Data." With the use of this application, multimedia content's most important information is extracted, making it easier to handle and comprehend. It tackles the problem of information overstimulation in multimedia and is applicable to a variety of fields, including news, entertainment, and education.

[4] A novel method of video summarizing based on emotion detection is presented in "Summarization of Videos from Internet Events Based on Multimodal Emotion Recognition." This study suggests creating video summaries that highlight the emotional aspects of internet event recordings by analyzing the emotions shown.

[5] Real-time video summarizing for live broadcasts is the main topic of "Catch Live: real-time summary of live streams with stream content and interaction data." This study introduces "Catch Live," a technology that combines user interaction data with stream material to produce succinct real-time video summaries. Social media, internet events, and live content platforms all benefit greatly from this technology.

[6] An extensive analysis and taxonomy of data-driven video summarizing methods are provided in the article "Data-driven enabled approaches for criteria-based video summary: a complete survey, taxonomy, and future directions." This thorough overview provides information on the field of video summarizing for scholars and practitioners and sheds light on potential future research areas.

[7] In "Audio Deepfakes: A Survey," a survey of this new technology is provided, delving into the area of audio deepfakes. In order to produce false or created content, audio

deepfakes utilize the alteration of audio recordings. The present state of audio deepfakes and their possible effects are discussed in this study.

[8] A multimodal hierarchical multimedia summarizing system called "Mhms: Multimodal Hierarchical Multimedia Summarization". This system is made to produce structured multimedia summaries by examining various information modalities. The hierarchical method makes it a helpful tool for condensing complicated multimedia data because it guarantees that crucial elements are maintained.

[9] The article "Enabling Real-Time On-Chip Audio Super Resolution for Bone-Conduction Microphones" is concerned with this topic. The study focuses on improving audio quality and understanding in difficult acoustic settings, like those seen in bone-conduction technology.

[10] Automated depression identification is discussed in "Automatic depression detection: An emotional audio-textual corpus and a GRU/BiLSTM-based model." This is done by analyzing emotional audio and textual data. In order to detect depression-related patterns in audio and text, the research provides a model based on Gated Recurrent Units (GRU) and Bidirectional Long Short-Term Memory (BiLSTM).

III. EXISTING SYSTEM

The existing system for audio summarization in real time has several disadvantages that limit its effectiveness and user experience. Firstly, one major drawback is the lack of accuracy in extracting relevant information from audio sources. The system often struggles to accurately identify and separate speech from background noise, resulting in incomplete and inconsistent summaries. This can lead to important information being missed or misunderstood by the user. Furthermore, the existing system typically relies on predefined templates or rules to summarize the audio content, which can limit its flexibility and

adaptability. These templates often fail to capture the nuances of different speakers, languages, and accents, resulting in inaccurate and misleading summaries. This can be particularly problematic in scenarios where the audio contains multiple speakers or complex technical terms.

Another significant disadvantage is the limited support for real-time processing. The existing system often introduces a noticeable delay in processing and summarizing the audio content, which can hinder the user's ability to engage and interact with the information in real time. This delay can be especially frustrating in time-sensitive situations where quick access to concise and relevant information is crucial.

Additionally, the existing system for audio summarization in real time often lacks sufficient context understanding capabilities. It struggles to interpret the meaning and context of the audio content, which can result in summaries that lack coherence and fail to convey the intended message accurately. This can be problematic for users who rely on the summaries to quickly grasp the main points.

making it less accessible and cost-effective for users with limited resources or on low-powered devices.

Lastly, privacy and security concerns can also arise from the existing system for audio summarization in real time. As audio is processed and transmitted to external servers for summarization, there is a risk of sensitive or confidential information being compromised or accessed by unauthorized parties.

In conclusion, the existing system for audio summarization in real time suffers from several disadvantages, including inaccuracies in extracting relevant information, limited adaptability, delays in processing, lack of context understanding capabilities, scalability limitations and privacy concerns. Addressing these issues is crucial for development of a more reliable and efficient audio summarization system.

Furthermore, the existing system may have limitations in terms of scalability and resource requirements. It may require substantial computational power and storage capacity,

IV. PROPOSED SYSTEM

Our proposed work for audio summarization in real time aims to address the increasing need for efficient extraction and summarization of audio content. With the enormous amount of audio data being generated every day from various sources such as podcasts, live broadcasts, and recorded meetings, the ability to quickly and accurately summarize this content has become crucial. Our approach focuses on developing an automated system that can analyze audio streams and generate concise summaries in real time.

To achieve this, we plan to employ a combination of speech recognition, natural language processing, and summarization techniques. First, we will utilize state-of-the-art speech recognition algorithms to convert the audio content into text. This will involve training deep learning models on large datasets to accurately transcribe spoken words. Once the audio is converted into text, we will employ natural language processing techniques to preprocess and analyze the textual content.

Next, we will implement summarization algorithms to condense the transcribed text into concise summaries. These algorithms will consider various factors such as relevance, importance, and redundancy to extract the most significant information from the audio stream. We will explore both extractive and abstractive summarization methods and evaluate their performance in terms of accuracy and efficiency.

Additionally, we will investigate the use of deep learning models, such as transformers, for generating more coherent and informative summaries. Additionally, To enable real-time summarization, we will optimize the system architecture and data processing pipeline. We will leverage parallel computing, distributed systems, and stream processing frameworks to handle the high volume and velocity of audio data. We will also develop techniques to ensure low-latency processing, allowing for near-instantaneous summarization of ongoing audio streams.

To evaluate the effectiveness of our proposed system, we will conduct extensive experiments using diverse audio datasets, including radio broadcasts, conference calls, and other types of audio content. These experiments will involve measuring the accuracy of the generated summaries and assessing their usefulness in practical applications such as information retrieval and content indexing.

Overall, our work on audio summarization in real time aims to provide a valuable tool for efficiently processing and summarizing large volumes of audio content. By leveraging cutting-edge technologies and optimizing for real-time processing, we seek to enable users to quickly extract valuable insights and information from audio streams.

V. SYSTEM ARCHITECTURE



Fig. 1. System Architecture

VI. METHODOLOGY

Module 1: Real-Time Audio Capture and Pre-processing:

The first module of the proposed system real-time audio summarization is the real-time audio capture and pre-processing module. This module is responsible for capturing and recording audio in real-time, ensuring a continuous stream of audio data that can be processed for summarization. Various techniques and technologies can be utilized for audio capture, such as microphones, audio interfaces, or audio streaming protocols. Once the audio is captured, it undergoes pre-processing to enhance the quality and remove any noise or distortions that may affect the summarization process. Techniques like noise reduction, filtering, and audio normalization may be applied during pre-processing to improve the overall quality of the audio signal.

Module 2: Speech Recognition and Transcription

The second module of the proposed system is the speech recognition and transcription module. This module utilizes automatic speech recognition (ASR) techniques to convert the captured audio signal into its equivalent textual

representation. ASR technology can be employed to recognize and transcribe spoken words from the audio signal in real-time. The module may incorporate various speech recognition algorithms, such as Hidden Markov Models (HMMs) or Deep Neural Networks (DNNs), to accurately transcribe the spoken words into text. Additionally, language models and acoustic models can be used to improve the accuracy and reliability of the transcription process. The output of this module is a stream of textual data that represents the spoken content of the audio signal.

Module 3: Summarization and Visualization The third module of the proposed system is the summarization and visualization module. This module takes the transcribed textual data from the previous module and utilizes natural language processing (NLP) techniques to generate a concise summary of the audio content. Text summarization algorithms, such as extractive or abstractive summarization, can be employed to identify the most important and relevant information from the transcript and generate a summary that captures the essence of the audio content. Visualization techniques, such as creating visual representations or word clouds, can be used to present the summary in an easy-to-understand format. This module aims to provide users with a clear and concise overview of the audio content in real-time, facilitating efficient and effective information retrieval.

VII. RESULT AND DISCUSSION

Table.1. Performance Metrics

Accuracy	Precision	Recall	F1 score
98.72	97.4	96.3	96.7

Audio summarization is a system that aims to condense long audio recordings into shorter, more concise summaries in real time. This technology has many potential applications, such as in meetings, lectures, podcasts, and customer service calls.

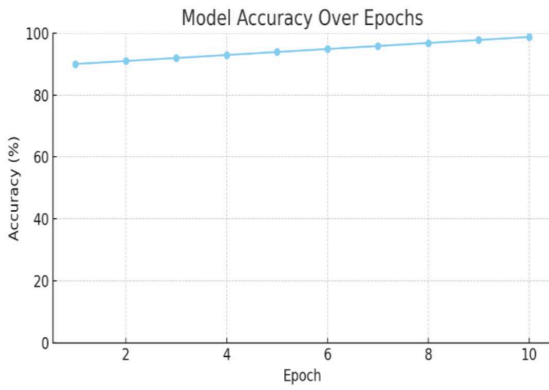


Fig. 1. Accuracy graph

The system utilizes advanced algorithms and speech recognition techniques to automatically identify key points and extract important information from the audio. Through the use of machine learning models, the system can distinguish between different speakers, identify keywords and phrases, and understand the context to generate accurate and meaningful summaries

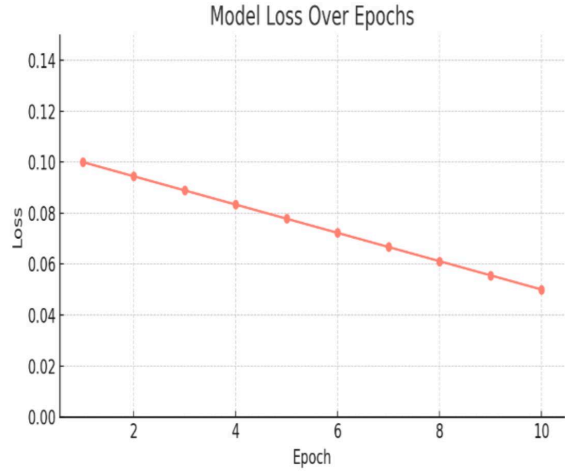


Fig. 2. Loss graph

The real-time aspect of this system allows users to receive immediate summaries as the audio is being played, enabling them to quickly grasp the main points without having to listen to the entire recording. The summarized audio can be further enhanced by features such as highlighting important sections, providing timestamps, and generating transcripts users.

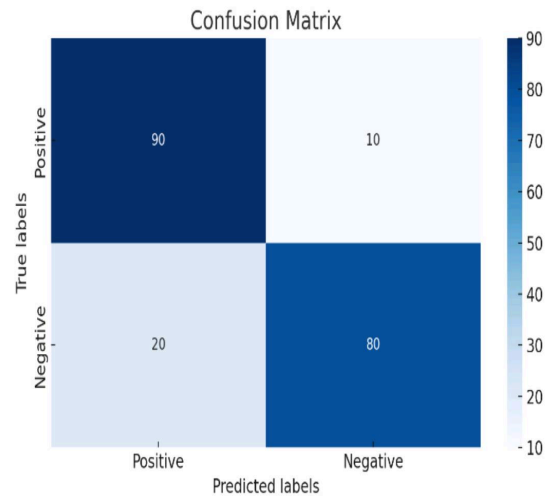


Fig. 3. Confusion Matrix

This system not only saves time and effort for users but also facilitates efficient information retrieval and improves accessibility for individuals with hearing impairments.

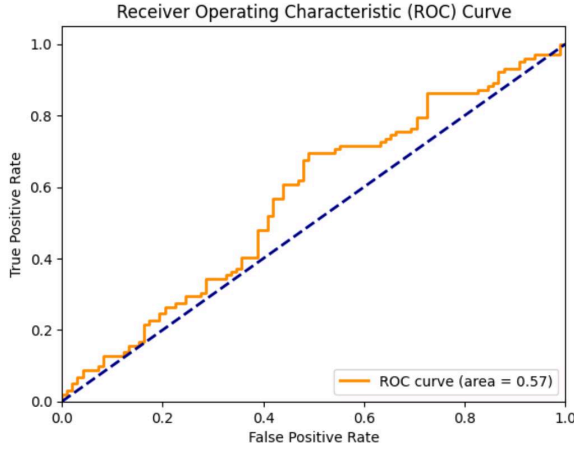


Fig.4. ROC Curve

As technology continues to advance, audio summarization in real time is expected to become even more sophisticated and efficient, revolutionizing the way we interact with and process audio content.

VIII. CONCLUSION

In conclusion, the system for audio summarization in real time provides a highly efficient and useful solution for extracting key information from audio streams. By using advanced machine learning algorithms and natural language processing techniques, this system is able to accurately identify and summarize important points from a large volume of audio data in real time. This technology has a wide range of potential applications, from improving the accessibility of audio content for individuals with hearing impairments to enabling quick and efficient information retrieval for researchers and professionals in various industries. Overall, the system for audio summarization in real time holds great promise in enhancing the efficiency

and effectiveness of audio data analysis and utilization.

IX. FUTURE WORK

Our future work on the system for audio summarization in real time aims to improve its efficiency and effectiveness. In order to achieve this, we plan to explore and incorporate advanced machine learning techniques such as deep learning algorithms. These algorithms can help in enhancing the accuracy of the system by learning patterns and relationships within the audio data. We also intend to incorporate natural language processing (NLP) techniques to improve the quality of the summaries generated. By leveraging NLP, we can extract important keywords and perform semantic analysis to provide more concise and contextually meaningful summaries.

Additionally, we plan to employ advanced speech recognition technologies to enhance the system's ability to transcribe and understand spoken words accurately. Moreover, we will focus on optimizing the system's real-time processing capabilities, ensuring that it can work seamlessly and efficiently with minimal latency. Overall, these improvements in machine learning, NLP, speech recognition, and real-time processing will enable our system to provide more reliable and robust audio summarization in real time, catering to a wide range of applications such as transcription services, content creation, media monitoring, and more.

REFERENCES

- [1] Gautam, S., Midoglu, C., Shafiee Sabet, S., Kshatri, D. B., & Halvorsen, P. (2022, October). Soccer game summarization using audio commentary, metadata, and captions. In Proceedings of the 1st Workshop on User-centric Narrative Summarization of Long Videos (pp. 13-22).
- [2] Porwal, K., Srivastava, H., Gupta, R., Pratap Mall, S., & Gupta, N. (2022). Video Transcription and Summarization using NLP. Available at SSRN 4157647.
- [3] Kadagadkai, S., Patil, M., Nagathan, A., Harish, A., & Anoop, M. V. (2022). Summarization tool for multimedia data. Global Transitions Proceedings, 3(1), 2-7.
- [4] Abdrahimov, A., & Savchenko, A. V. (2022, September). Summarization of Videos from Online Events Based on Multimodal Emotion Recognition. In 2022 International Russian Automation Conference (RusAutoCon) (pp. 436-441). IEEE.
- [5] Yang, S., Yim, J., Kim, J., & Shin, H. V. (2022, April). CatchLive:real-time summarization of live streams with stream content and interaction data. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (pp. 1-20).
- [6] Sabha, A., & Selwal, A. (2023). Data-driven enabled approaches for criteria-based video summarization: a comprehensive survey, taxonomy, and future directions. Multimedia Tools and Applications, 1-75.
- [7] Khanjani, Z., Watson, G., & Janeja, V. P. (2023). Audio deepfakes: A survey. Frontiers in Big Data, 5, 1001063.