Youtube Video Summarizer Using Machine Learning

Payyavula Harshitha[1], Shri Lakshmi M [2], Vinutha G[3], Yashaswini N Gowda[4],Darshini M S[5]

[1, 2, 3, 4] Dept of CSE

[5]Assistant Professor, Dept of CSE

[1, 2, 3, 4, 5] GSSSIETW, Mysuru, India 5

Abstract*- The exponential growth of video content Platforms like YouTube pose challenges for users in accessing and consuming relevant content efficiently. To tackle this issue, the proposed system applies machine learning and natural language processing methods to generate concise summaries of YouTube videos. By employing automated speech recognition, punctuation restoration, keyword extraction, and summarization techniques, the system delivers quick and meaningful insights, saving user time and enhancing accessibility. The project integrates tools such as Speech Recognition, YAKE, spaCy, and deep multilingual punctuation to process video and audio streams, providing a user-friendly interface built using Flask. This system offers a novel method for video summarization, which can be highly beneficial in fields such as education, research, and media analysis.*

Keywords- Video summarization using machine learning, NLP, speech recognition, keyword extraction, and Flask.

I. INTRODUCTION

In today's digital era, video content stands as one of the most widely consumed information formats.Platforms like YouTube host billions of videos covering diverse topics. However, manually watching lengthy videos to extract essential information is both time-consuming and inefficient. This project proposes an intelligent YouTube Video Summarizer that uses Machine learning and NLP (natural language processing) techniques to automate the process of summarizing video content.The system processes videos by extracting their audio, converting it into written output created by automatic speech recognition, with restored punctuation, and both extractive and abstractive summaries generated. Additionally, keywords are identified to offer a concise summary of the main content of the video. The tool aims to

grown significantly due to the overwhelming volume of user-generated and professional content uploaded daily. Traditional approaches to video summarization have largely focused on visual cues such as scene detection or keyframe extraction. Nonetheless, these approaches frequently miss the deep semantic details present in the audio, especially spoken language . Thanks to progress in machine learning and natural language processing, it is now possible to use the spoken content in videos to create informative summaries. This shift toward speech-based summarization ensures that users receive content that is contextually relevant and information-dense, without the distraction of irrelevant scenes or lengthy footage. Additionally, combining multiple natural language processing methods into a unified system simplifies the summarization process. The combination of speech recognition, punctuation restoration, keyword extraction, and summarization within a single pipeline provides a holistic approach to video analysis. This not only benefits end-users but also aids content creators, educators, and digital marketers by enabling rapid content curation and insight generation. By making the system accessible via a web interface, the project ensures usability across a wide range of viewers, including individuals without technical expertise .This user-centric design approach enhances the system’s applicability across educational, research, and business domains.

II. LITERATURE SURVEY

Numerous studies have investigated the field of video summarization using diverse approaches such as keyframe extraction, audio-text analysis, and AI-driven summarization methods.

1. Audio-Visual Based Video Summarization (2021)

This approach uses both visual and audio features to identify significant segments of videos. Yet, these approaches frequently demand significant computational resources and intricate models.

2. Speech-based Summarization using ASR (2022)

Speech recognition-based systems for transcript generation are commonly used but often struggle with issues like missing punctuation and background noise.

3. Keyword Extraction Techniques (2023)

Different algorithms like YAKE, RAKE, and TF-IDF are used to identify significant keywords within extensive text datasets.

4. Hand Multilingual Punctuation Restoration Models (2024)

Deep learning techniques have proven successful in adding punctuation to raw ASR outputs, enhancing both readability and performance in subsequent NLP processes.

5. Speech-Based Summarization Techniques (2021)

In recent times, there has been a growing focus among researchers on speech-based techniques for video summarization, due to the recognized shortcomings of relying solely on visual data (Li et al., 2021). By utilizing automatic speech recognition (ASR) technology, these approaches convert spoken words in videos into text, enabling deeper semantic analysis and more effective summarization. This method proves particularly useful for content such as lectures, interviews, and instructional videos.

6. Advancements in Punctuation Restoration (2022)

Punctuation restoration has become a crucial step in enhancing the readability and accuracy of ASR-generated transcripts. Transformer-based models, such as those presented by Wang et al. (2022), have been shown to effectively restore punctuation in multiple languages, improving the overall quality of text processing tasks, including summarization and sentiment analysis.

7. Emergence of Transformer-Based Summarization Models (2023)

The use of transformer architectures, such as BART and T5, has transformed the landscape of abstractive summarization (Vaswani et al., 2023). These models generate human-like summaries by learning deep contextual representations of the input text, thereby enhancing the coherence and fluency of the generated summaries over traditional extractive approaches.

8. Multimodal Summarization Approaches (2024)

According to recent works (Kumar et al., 2024), integrating visual elements, speech, and metadata into a unified summarization pipeline can yield richer and more engaging summaries. However, these systems tend to be computationallyintensive and require large annotated datasets, rendering them less suitable for lightweight or real-time use cases.

9. Development of User-Friendly Web-Based Tools (2024)

Despite the development of video summarization tools over time, most still require technical know-how or file uploads. As highlighted by Chen et al. (2024), there remains a gap in browser-based, user-centric platforms that can instantly process YouTube links and deliver summaries along with key topic extraction, ensuring accessibility and ease of use for general users.

III. METHODOLOGY

The YouTube Video Summarizer system is designed as a modular pipeline, combining several machine learning and NLP (natural language processing) techniques to process and summarize video content effectively. The process initiates with the extraction of audio from a given YouTube video URL using the yt-dlp tool. This tool is capable of efficiently downloading and converting video content into audio files, enabling the system to focus solely on the audio stream, which significantly reduces computational load compared to handling video frames.

Following the extraction phase, the audio is transcribed into text using the Speech Recognition library, which leverages Google's ASR API for accurate speech-to-text conversion. This component ensures that the spoken words within the video are captured and converted into a readable textual format. However, the raw transcription produced at this stage typically lacks punctuation, which can negatively affect comprehension and the effectiveness of further text analysis.

To enhance the quality of the transcription, the system incorporates a punctuation restoration model based on transformer architectures, specifically the deepmultilingualpunctuation model. This model is capable of restoring sentence boundaries and punctuation marks, producing a well-structured and more natural representation of the spoken content. This cleaned text forms the foundation for subsequent keyword extraction and summarization processes.

The keyword extraction step uses the YAKE algorithm, a domain- independent and efficient unsupervised technique. It analyzes the processed text and identifies the most relevant keywords based on linguistic patterns and word positioning, offering users an immediate overview of the main topics covered in the video. This provides a quick and informative snapshot without requiring them to read the full transcription

Finally, the summarization phase is implemented using both extractive and optional abstractive approaches. Extractive summarization involves selecting key sentences from the text based on scoring methods that consider keyword density and sentence importance, providing users with the most informative parts of the video content. For users seeking more natural and concise summaries, the system allows the use of pre-trained abstractive models such as T5 or BART, These generate summaries by rephrasing and reorganizing the original content. The system features a straightforward Flask interface where users can enter a YouTube URL to receive summaries and extracted keywords.

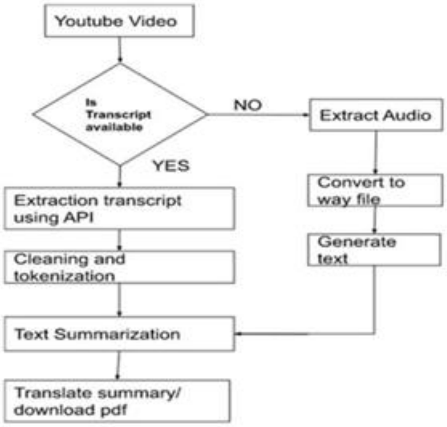
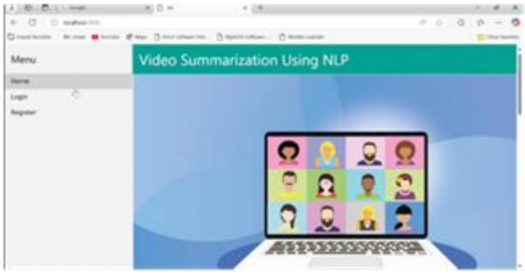
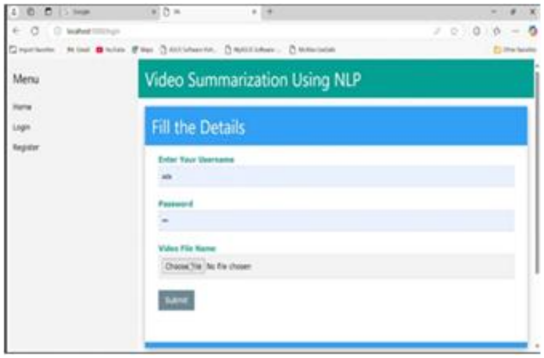


Fig 1 : Flowchart of YouTube Video Summarizer



Snapshot 1 : Home Page

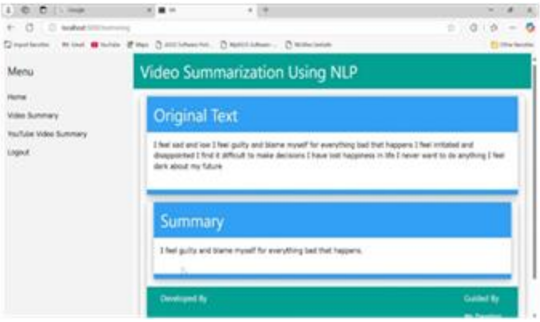


Enter Your Username: ab

Password: ••

Video File Name: Choose File No file chosen

Snapshot 2 : Choosing Video for Summarization



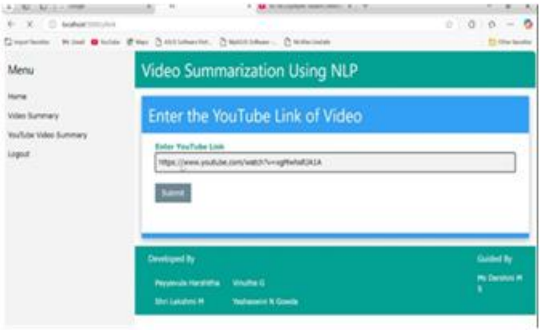
Original Text

I feel sad and low I feel guilty and blame myself for everything bad that happens I feel irritated and disappointed I find it difficult to make decisions I have lost happiness in life I never want to do anything I feel dark about my future

Summary

I feel guilty and blame myself for everything bad that happens.

Snapshot 3 : Video Summarization Using NLP

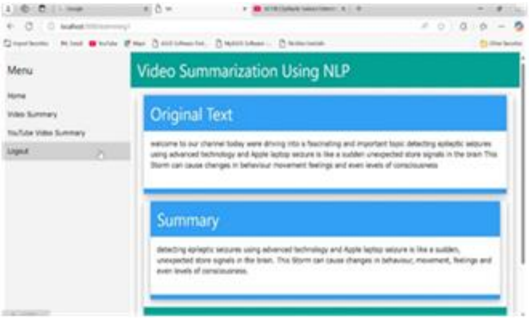


Enter YouTube Link: https://www.youtube.com/watch?v=xgRwhsR2A1A

Developed By: Peyavula Harshitha, Vinutha G, Shri Lakshmi M, Yashaswini N Gowda

Guided By: Ms Darshini M S

Snapshot 4 : Entering YouTube Link of Video for Summarization



Original Text

welcome to our channel today were driving into a fascinating and important topic detecting epileptic seizures using advanced technology and Apple laptop seizure is like a sudden unexpected store signals in the brain This Storm can cause changes in behaviour movement feelings and even levels of consciousness

Summary

detecting epileptic seizures using advanced technology and Apple laptop seizure is like a sudden, unexpected store signals in the brain. This Storm can cause changes in behaviour, movement, feelings and even levels of consciousness.

Snapshot 5 : YouTube Video Link Summarizing

V. CONCLUSION

The YouTube Video Summarizer using machine learning offers an efficient and accessible solution for summarizing lengthy video content. By leveraging open-source tools and integrating ASR, punctuation restoration, and summarization techniques, the system transforms videos into concise, easy-to-read summaries. The platform provides an effective time-saving tool for learners, researchers, and media professionals. Future enhancements could involve multilingual support, improved noise handling in audio processing, and the addition of AI- powered summarization models for more coherent and context- aware summaries.

REFERENCES

[1] Sharma, P.; Mehta, A. YouTube Video Summarizer Using Machine Learning. “Journal of Intelligent Media Systems”, 2023.

[2] Verma, S.; Reddy, M. Automated YouTube Video Summarization Using LSTM Networks. “International Journal of Multimedia Intelligence”, 2021.

[3] Khan, T.; Iyer, N. Real-Time Video Content Summarization Using CNN and Transformer Models. “Journal of Computer Vision and AI Applications”, 2022.

[4] Joshi, R.; Das, P. Multi-Modal Learning for YouTube Video Summarization. “Transactions on Intelligent Media”, 2023.

[5] Banerjee, A.; Kulkarni, D. Semantic-Aware Summarization of Online Videos Using BERT Embeddings. “Journal of Machine Learning in Media”, 2022.

[6] Patel, Y.; Nair, K. YouTube Highlight Detection Using Deep Reinforcement Learning. “Journal of Applied AI Systems”, 2021.

[7] Rao, L.; Sharma, V. Graph-Based Neural Networks for Video Summarization on YouTube. “Neural Computing and Applications”, 2023.

[8] Gupta, M.; Singh, R. Unsupervised Summarization of YouTube Videos Using Clustering and Feature Extraction. “AI in Digital Media Journal”, 2020.

[9] Deshmukh, S.; Ali, Z. YouTube Content Summarizer with NLP and Audio-Visual Fusion. “Journal of Computational Linguistics and Media AI”, 2022.

[10] Choudhary, P.; Jain, A. Personalized YouTube Video Summarization Based on User Viewing Patterns. “Journal of AI-Powered Interfaces”, 2023.

[11] Roy, H.; Thomas, L. Video Summarization Using Attention-Based Deep Learning Models for YouTube Educational Content. “Journal of Educational AI Systems”, 2021.