Video Summarization for Marathi Language

Prof. Indira Joshi[1], Shruti Aurade[2], Nishant Sanap[3], Nisha Kangane[4]

[1]Indira Joshi, Associate Professor, Computer Engineering, NHITM

[2]Shruti Aurade, Computer Engineering, NHITM [3]Nishant Sanap, Computer Engineering, NHITM [4]Nisha Kangane, Computer Engineering, NHITM

---***---

Abstract - The Video Summarization Platform using Python Flask is a comprehensive tool designed to summarize Marathi and English videos while providing summaries in Hindi, Marathi, and English languages. Leveraging machine learning and natural language processing (NLP) techniques, this platform offers a sophisticated solution for efficiently extracting key information from videos. The platform begins by transcribing the audio content of the video into text using automatic speech recognition (ASR) technology. This transcription process ensures that the platform can accurately analyze and summarize the video's content. Next, the text is translated into the target languages, namely Hindi, Marathi, and English, enabling users from diverse linguistic backgrounds to access the summarized content. To generate concise and informative summaries, advanced NLP algorithm is applied. This algorithm analyze the transcribed text to identify the most significant phrases, sentences, and concepts. By considering factors such as keyword frequency, semantic relevance, and context, the platform effectively distils the video's content into digestible summaries. Additionally, machine learning models are employed to classify the type of video content. These models are trained on diverse datasets encompassing various video genres and topics. By recognizing patterns and features within the video content, the platform can accurately categorize videos into distinct types, such as news, interviews, tutorials, or entertainment. The platform's user interface, powered by Python Flask, offers a seamless experience for users to upload videos, select their preferred language for summarization, and receive concise summaries in their chosen languages. The intuitive design ensures accessibility and ease of use, catering to both novice and advanced users. Overall, the Video Summarization Platform serves as a valuable resource for individuals seeking efficient ways to consume multimedia content. Whether for educational, informational, or entertainment purposes, this platform empowers users to access summarized video content in multiple languages, facilitated by cutting-edge machine learning and NLP technologies.

*Key Words***:** Transcription, Marathi-speaking users, Marathi YouTube videos, video content, transcription, summary, translation, Natural Language Toolkit (NLTK), content comprehension, user interaction data, past summaries, recommendation

I. INTRODUCTION

One of the most prevalent ways to get to visual information is presently video. It would take nearly 85 years fair to see each video that is distributed to YouTube each day due to the enormous volume of video information! Subsequently, it is significant to have automated strategies for video substance investigation and comprehension. Programmed video summarization in specific is a significant apparatus for helping human users in browsing video fabric. A compelling video rundown would condense the fundamental focuses of the unique video into a brief, distinguishable overview.

There are numerous ways that video outlines can cut down on the length. Embarking on a journey to revolutionize the way Marathi-language video content is consumed, this project introduces an innovative video summarization system that stands as a beacon of linguistic inclusivity and technological advancement. At its core, the project is driven by the aspiration to distil the essence of Marathi video narratives into succinct summaries, thereby democratizing information and transcending the barriers of time and language. The system is not merely a tool but a gateway to a world where the richness of Marathi visual storytelling is made accessible to all, regardless of their capacity to engage with prolonged content. By harnessing the transformative potential of the TF-IDF vectorizer algorithm and the scikit-learn library, the project transcends traditional summarization methods, offering a unique blend of quantitative analysis and qualitative insight. With the Flask framework as its canvas, the project paints an interactive landscape where users can effortlessly navigate through the realms of video content, uploading links or files with the simplicity of a click. This system is a testament to the symbiotic relationship between human linguistic heritage and artificial intelligence, where each video becomes a chapter in the grand narrative of Marathi culture, waiting to be summarized, understood, and appreciated. In essence, the project seeks to create a harmonious symphony of technology and language, where every frame of a video is not just seen but truly perceived, and every word not just heard but deeply resonated with. It is a step towards a future where the power of video summarization becomes a key to unlocking the treasures hidden within the Marathi language, for scholars and casual viewers alike.

II. LITERATURE SURVEY

They used Latent Dirichlet Allocation (LDA) in Paper [1], which has been shown to be effective in summarization of documents. The suggested LDA summarizing model is divided into three stages. The first phase prepares the subtitle file for modelling by deleting stop words and doing other pre-processing tasks. The subtitles are used to train the LDA model in the second step in order to generate the list of keywords that will be employed to extract relevant sentences. The summary is prepared based on the list of keywords generated in the third phase.

M. Z. Khan et al. [7] assessment of four comparison datasets made up of movies portraying different occasions from both to begin with- and third-person viewpoints shows that our execution is competitive with that of completely administered state-of-the-art methods.

D. Sahrawat et al. [8] with Bit Transient Segmentation (KTS) for shot sections and a worldwide attention-based customized memory organize coupled with LSTM for shot score learning, we provide a clear strategy for abridging videos. Inquire about on information sets such as TV Sum and SumMe uncovers that our procedure performs generally 15 ways better than the state-of-the-art.

G. Yalınız et al. [11] for the unsupervised video summarization issue, an approach combining deep support learning and independently recurrent neural systems is put forward. In this method, the calculation can be created with more layers and steps without experiencing gradient-related issues.

Hussain et al. [10] there are a number of proposed low-level highlights and technique-based soft computing techniques that drop brief of completely utilizing MVS. In this paper, author integrate delicate computing strategies based on profound neural systems into a two-tier framework to accomplish MVS. Target-appearance-based shot segmentation is carried out by the to begin with online layer, which at that point spares the comes about in a lookup table before sending them to the cloud for extra processing.

III. PROBLEM STATEMENT

The problem entails developing a Transcription Web App tailored for Marathi-speaking users. Upon receiving a Marathi YouTube video link, the app fetches the video content, extracts its audio, and transcribes it, providing a summary. Additionally, users can opt to translate the Marathi summary into Hindi and English. Leveraging Natural Language Toolkit (NLTK) and Natural Language Processing (NLP), this multi-functional tool aims to enhance accessibility and comprehension of Marathi content on YouTube while facilitating language understanding and content discovery for users.

IV. OBJECTIVES

- Develop a web application interface that accepts Marathi YouTube video links as video.
- Implement functionality to fetch the video content from the provided link and extract the audio.
- Utilize speech-to-text technology to transcribe the audio into Marathi text.
- Summarize the transcribed text to provide a concise summary of the video content.
- Provide options for users to translate the Marathi summary into Hindi and English.
- Utilize NLTK (Natural Language Toolkit) and NLP (Natural Language Processing) techniques for text summarization.

V. PROPOSED METHODOLOGY

- Data Collection and Preprocessing: Collect diverse video datasets in Marathi and English languages. Preprocess the videos to extract audio content and synchronize it with subtitles or captions.

- Automatic Speech Recognition (ASR) with Google Speech Recognition API: Utilize the Google Speech Recognition API for transcribing the audio content of the videos into text. This API offers accurate transcription for multiple languages, including Marathi and English. Preprocess the transcribed text to remove noise and improve transcription quality.

- Machine Translation: Translate the transcribed text into target languages (Hindi, Marathi, and English) using machine translation models. Google Translate API or other similar services can be used for this purpose.

- Text Summarization with TF-IDF: Implement text summarization using the Term Frequency-Inverse Document Frequency (TF-IDF) technique.

- Video Classification with TF-IDF: Utilize TF-IDF to classify the type of video content. Extract features from the transcribed text using TF-IDF vectors. Train classification models such as support vector machines (SVMs) or random forests on labeled video datasets to classify videos into different categories (e.g., news, interviews, tutorials).

- Web Development with Python Flask: Develop the backend of the Video Summarization Platform using Python Flask. Create APIs for video uploading, language selection, and summary generation. Design frontend components using HTML, CSS, and JavaScript for an intuitive user interface.

- Integration and Deployment: Integrate the ASR, translation, summarization, and classification modules into a cohesive platform. Deploy the platform on a web server, possibly using containerization with Docker for scalability and ease of management

VI. WORKING

Introducing a revolutionary Transcription Web App designed to streamline the process of accessing, summarizing, and translating Marathi YouTube videos. This innovative platform aims to enhance accessibility and comprehension for users, offering a seamless experience from video input to summarized content output. The Transcription Web App operates with simplicity at its core. Users begin by providing a Marathi YouTube video link as input. Leveraging cutting-edge technology, the app swiftly fetches the video content and extracts its audio component. Utilizing advanced transcription algorithms, it accurately converts the audio into text, generating a comprehensive summary of the video's content.

But the functionality doesn't end there. Recognizing the diverse linguistic needs of users, the app offers options for translating the Marathi summary into both Hindi and English. This multi-lingual capability ensures that language barriers are effectively overcome, facilitating broader access to valuable content. Moreover, the Transcription Web App goes beyond mere transcription and translation.

By harnessing the power of Natural Language Processing (NLP) techniques, it analyses the summarized content to provide insightful recommendations. Drawing from past summaries created using NLTK (Natural Language Toolkit) and NLP, the app offers personalized suggestions tailored to the user's preferences. By identifying patterns in the types of videos frequently watched, it intelligently recommends genres that align with the user's interests. By seamlessly integrating transcription, translation, recommendation, and content discovery functionalities, it offers a comprehensive solution that caters to diverse user needs. Whether seeking quick summaries, language translations, or personalized recommendations, users can rely on this innovative platform to elevate their video consumption experience.
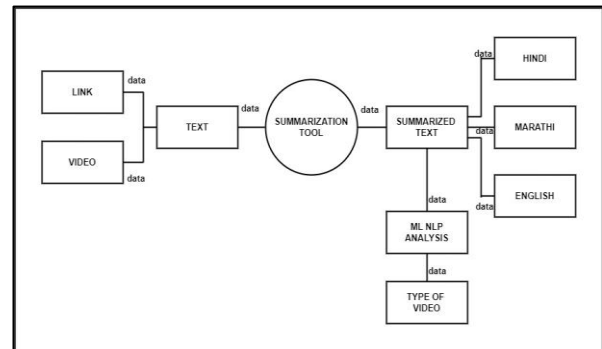
VII. WORKFLOW
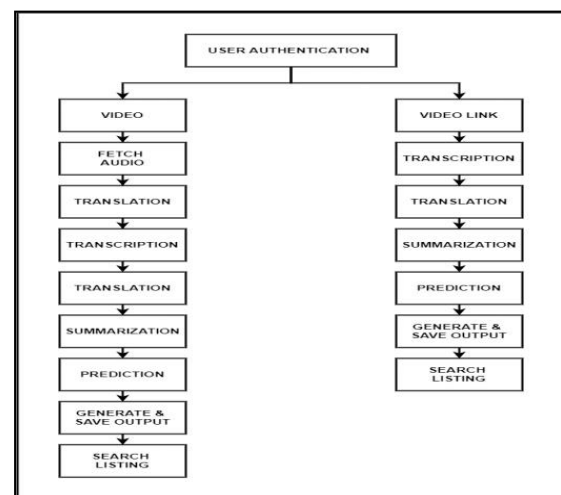

Fig 1: Workflow

Fig 2: Flowchart

VIII. PERFORMANCE EVALUATION:

In addition to visual features extracted from video frames, textual information associated with the videos can also provide valuable context for emotion prediction. To leverage this textual data, we utilized the TF-IDF Vectorizer algorithm to preprocess and extract key textual features from the accompanying metadata or transcripts of the videos. These features were then incorporated alongside visual features into our emotion prediction models. By considering both visual and textual cues, our approach aimed to capture a more comprehensive representation of the emotional content within the videos. In the performance evaluation phase, we conducted experiments to assess the impact of incorporating textual features on the predictive accuracy of our models. This involved comparing the performance of models trained solely on visual features with those incorporating both visual and textual features.
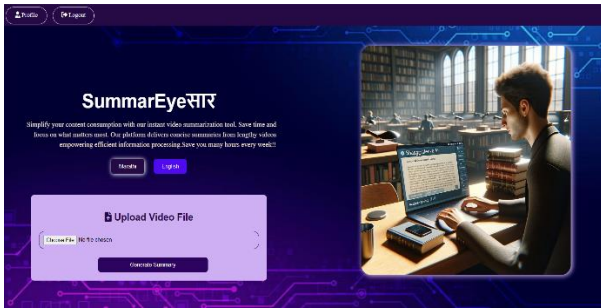
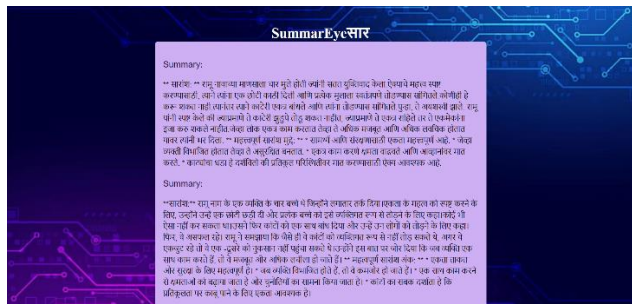## IX. RESULTS

Fig 3: Home Page
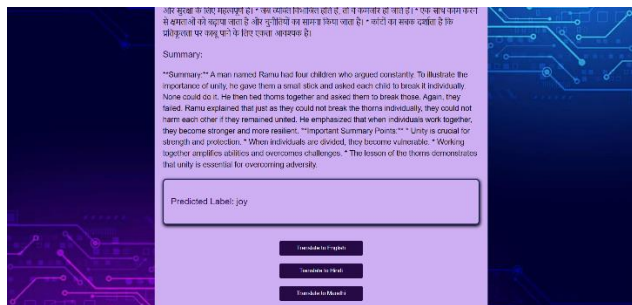

Fig 4: Summary generated

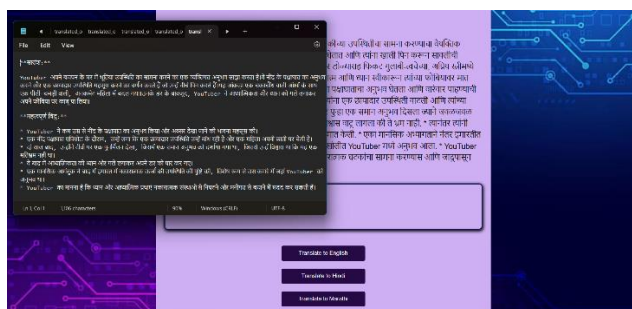**Fig 5: Predicted Video Label**

Fig 6: Download Summary

X. CONCLUSIONS

In conclusion, the Transcription Web App offers a seamless solution for Marathi-speaking users by extracting audio content from Marathi YouTube videos and transcribing them into summaries. Additionally, users can conveniently translate these summaries into Hindi and English. Leveraging NLTK and NLP technologies, the app not only provides efficient

transcription but also offers personalized recommendations based on the user's viewing history. By analyzing past summaries, the app identifies the user's preferred genres, enabling targeted content suggestions. This functionality enhances the user experience by streamlining content discovery, ultimately fostering a more tailored and engaging viewing environment. With its multifaceted capabilities, the Transcription Web App not only facilitates accessibility to Marathi content but also augments user engagement and satisfaction through personalized recommendations and efficient transcription services.

REFERENCES

1. Alrumiah, S. S., Al-Shargabi, A. A. (2022). Educational Videos Subtitles' Summarization Using Latent Dirichlet Allocation and Length Enhancement. CMC-Computers, Materials & Continua, 70(3), 6205–6221.

2. Sangwoo Cho, Franck Dernoncourt , Tim Ganter, Trung Bui, Nedim Lipka, Walter Chang, Hailin Jin, Jonathan Brandt, 11 Sep 2021

3. S. Chopra, M. Auli, and A. M. Rush, ''Abstractive sentence summarization with attentive recurrent neural networks,'' in Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics Hum. Lang. Technol., June 2016, pp. 93–98.

4. Ghadage, Yogita H. and Sushama Shelke. "Speech to text conversion for multilingual languages." 2016 International Conference on Communication and Signal Processing (ICCSP) (2016): 0236-0240.

5. Pravin Khandare, Sanket Gaikwad, Aditya Kukade, Rohit Panicker, Swaraj Thamke, " Audio Data Summarization system using Natural Language Processing, " International Research Journal of Engineering and Technology (IRJET) Volume 06, Issue 09, [September - 2019], e-ISSN: 2395-0056; p-ISSN: 2395-0072.

6. Hugo Trinidad and Elisha Votruba, " Abstractive Text Summarization Methods "

7. Prof. S. A. Aher , Hajari Ashwini M , Hase Megha S, Jadhav Snehal B, Pawar Snehal S, " Generating Subtitles Automatically For Sound in Videos, " International Journal of Modern Trends in Engineering and Research (IJMTER) Volume 03, Issue 03, [March – 2016] ISSN (Online):2349–9745; ISSN (Print):2393-8161

8. Aiswarya K R, " Automatic Multiple Language Subtitle Generation for Videos, " International Research Journal of Engineering and Technology (IRJET) Volume 07, Issue 05, [May - 2020], e-ISSN. 2395-0056, p-ISSN: 2395- 0072.

9. Savelieva, Alexandra & Au-Yeung, Bryan & Ramani, Vasanth. (2020). Abstractive Summarization of Spoken and Written Instructions with BERT.

10. Patil, S. et al. "Multilingual Speech and Text Recognition and Translation using Image." International journal of engineering research and technology 5 (2016): n. Pag.

11. S. Sah, S. Kulhare, A. Gray, S. Venugopalan, E. Prud'Hommeaux and R. Ptucha, "Semantic Text Summarization of Long Videos," 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), 2017, pp. 989-997, doi: 10.1109/WACV.2017.115.

12. A.Dilawari and M. U. G. Khan, "ASoVS: Abstractive Summarization of Video Sequences," in IEEE Access, vol. 7, pp. 29253-29263, 2019, doi: 10.1109/ACCESS.2019.2902507.