Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 29s (2025)
https://www.jneonatalsurg.com

OPEN ACCESS

Youtube Transcript Summarizer with Ai Chatbot

Ms. Mohan KS[1], Ms. Aruna A[1], Keethitha J[2], Mohan R[2], Srinithiga M[2], Sanjay V[2]

[1]Associate Professor, Department of Information Technology, SNS College of Technology, Coimbatore, India
[2]UG Students, Department of Information Technology, SNS College of Technology, Coimbatore, India

Cite this paper as: Ms. Mohan KS, Ms. Aruna A, Keethitha J, Mohan R, Srinithiga M, Sanjay V, (2025) Youtube Transcript Summarizer with Ai Chatbot, *Journal of Neonatal Surgery*, 14 (29s), 54-60

ABSTRACT

The YouTube Transcript Translation Web Application is a tool designed to extract the transcript from YouTube videos and translate it into multiple languages. This Flask-based application uses the YouTubeTranscriptApi to fetch the transcript of a given YouTube video and Deep Translator to translate the extracted text into various languages. By dividing the transcript into smaller pieces and translating each one separately, the system addresses the common problem of text in long transcripts exceeding query length limits. The field of text summarization has seen significant advancements, primarily due to progress in NLP and machine learning. Techniques range from extractive approaches, where key sentences are selected directly from the text, to abstractive methods, which generate summaries by paraphrasing the content. Tools such as BERT, GPT, and Transformer-based architectures have revolutionized summarization tasks. Previous studies have also explored video content summarization, focusing on either visual elements or transcripts. YouTube provides autogenerated transcripts for many videos, but these are often unstructured and verbose. Existing solutions for transcript summarization, such as manual editing or generic text summarizers, are time-intensive and lack context sensitivity for video-specific nuances. Current NLP-based tools may not integrate seamlessly with YouTube's API or fail to account for timestamped content, which is critical for maintaining the structure of video narratives. The proposed YouTube Transcript Summarizer was evaluated using a dataset of transcripts from various genres, including educational videos, podcasts, and tutorials. Metrics such as ROUGE scores and user satisfaction surveys were employed to assess the system's performance. The results demonstrated an average ROUGE-1 score of 85%, indicating a high level of accuracy in retaining critical information

Keywords: *TRS PCI, TMS, TRS, ANN*

1. INTRODUCTION

The purpose of the YouTube Transcript Translation Web Application is to break down language barriers and enhance global accessibility to YouTube content. As YouTube continues to be a major platform for video content across diverse audiences, language limitations often prevent viewers from fully engaging with content created in languages they do not understand. This application addresses that challenge by allowing users to extract and translate video transcripts into multiple languages, making YouTube videos accessible to a broader, multilingual audience. The system integrates two main

video transcript, and Deep Translator, which is used for translating the text. By automating the process of translation, the tool saves users significant time and effort, providing them with the option to translate content into various languages, including Tamil, Spanish, French, Hindi, Chinese (Simplified), German, and Japanese. Furthermore, it solves the problem of long transcript text exceeding query limits by splitting the transcript into smaller, manageable chunks. This ensures that even lengthy videos can be translated efficiently. The application is particularly beneficial for content creators and educators who wish to expand their reach to non-native audiences or provide educational material in different languages. It also supports cross-language learning, allowing users to better understand video content and engage in conversations across cultural boundaries. Additionally, it promotes accessibility for people with hearing impairments or those who prefer reading the transcript rather than watching the video, further democratizing access to knowledge. In essence, the YouTube Transcript Translation Web Application helps bridge communication gaps, fostering a more inclusive, globally connected digital ecosystem. Its ability to provide translated transcripts not only enriches the user experience but also plays a key role in improving content accessibility across the globe. The YouTube Transcript Translation Web Application is an innovative tool designed to extract and translate the transcript of YouTube videos into multiple languages. This application, built on Flask, leverages the YouTubeTranscriptApi to fetch transcripts from any given YouTube video, allowing users to access the video's text in a readable format. To overcome common challenges, such as the length of the transcript exceeding query limits, the system intelligently splits the transcript into smaller chunks and translates each chunk separately. The core feature of the application is its integration with Deep Translator, a robust translation tool, enabling the system to convert the transcript into

various languages. The supported languages include Tamil, Spanish, French, Hindi, Chinese (Simplified), German, and Japanese, making it accessible to a global audience. Once translated, users can compare the original transcript with the translated versions side by side, which enhances understanding and provides valuable insight into different linguistic interpretations of the content. This web application addresses the needs of a broad user base, including content creators, educators, and casual users seeking to explore videos in languages other than their own. For content creators, it offers an opportunity to increase the accessibility and reach of their videos, while educators can leverage the translations to make educational videos more inclusive. Additionally, the tool is especially beneficial for users who want to access videos in foreign languages, fostering cross- language content exploration and improving the global

accessibility of online videos. The user-friendly interface ensures that even those with limited technical expertise can easily input a YouTube URL, retrieve the transcript, and select one or more languages for translation. Overall, this tool significantly improves content understanding and accessibility, making it an invaluable resource for anyone engaged in cross- cultural or multilingual online content consumption.

2. LITERATURE REVIEW

Leveraging Deep Learning for Multilingual Accessibility focuses on the transformative role of deep learning models in making video content accessible across multiple languages. Machine translation (MT) has made significant strides, especially in the context of video platforms like YouTube. By combining natural language processing (NLP) techniques with deep learning architectures, such as transformers, MT systems can now generate high-quality translations of video transcripts in real-time.6The use of neural networks in translation, such as Google's Transformer model, has greatly improved the fluency and accuracy of machine-generated translations, enabling systems to handle more complex sentence structures and idiomatic expressions. This is particularly important for video content, where transcription and translation must occur quickly to maintain synchronization with the video. Deep learning models like OpenAI's GPT series and Google's BERT model also contribute to improving the contextual understanding of content, offering more precise translations for videos in diverse languages. Applications using these technologies—such as YouTube's automated subtitles and the integration of AI-based translation tools—are helping break language barriers, making video content globally accessible. As these models continue to evolve, they hold promise for enhancing video consumption for non-native speakers in educational, entertainment, and informational contexts Automated transcription and translation systems have significantly improved the accessibility of video content on a global scale. By leveraging technologies such as YouTube's Transcript API and advanced machine translation tools like Deep Translator and Google Translate, videos can now be transcribed in real-time and translated into multiple languages, making them accessible to non-native speakers. This has opened doors for greater global engagement, allowing people from different linguistic backgrounds to access educational, entertainment, and informational content previously limited to specific languages. These systems help overcome language barriers by generating accurate transcripts that are further translated into various languages, providing subtitles or captions that facilitate understanding. Furthermore, these tools enable creators to cater to a global audience, offering them the ability to consume content in their preferred language without requiring manual translation efforts. These systems have the potential to improve content discoverability across borders in addition to enhancing viewer experience, as noted in research by Beshry (2024) and Golan (2023). However, challenges remain in ensuring the accuracy of translations, particularly for less commonly spoken languages. As AI and machine learning models continue to evolve, these transcription and translation systems are expected to become even more precise, contributing to a more inclusive digital landscape (Zhang & Zhang, 2022).

AI-powered tools for transcription and translation have revolutionized the way we communicate across languages, particularly in digital media. These tools, such as automated transcription services and machine translation systems, play a

critical role in bridging language barriers in videos, podcasts, and other multimedia content. For example, YouTube's automated transcription feature can extract spoken words from videos and convert them into text. This text can then be translated into multiple languages using tools like Google Translate or Deep Translator, enabling global audiences to access content in their preferred language. The integration of AI in transcription and translation has dramatically improved both speed and accuracy. Translations become more nuanced and reliable thanks to AI systems' use of deep learning algorithms to comprehend meaning, syntax, and context. This is especially important in fields like education and entertainment, where content needs to be accurately conveyed to diverse viewers. Research indicates that the use of AI in translation also facilitates real-time communication, enabling live events or webinars to be accessible in multiple languages instantly (Cruz et al., 2023).7Furthermore, AI tools can process large volumes of content with minimal human intervention, making them cost-effective for content creators, businesses, and organizations seeking to reach international markets. These technologies are not only expanding accessibility but also contributing to the democratization of digital content across the globe The article "Multilingual Video Transcription and Translation: A Critical Review of Current Technologies and Tools" explores the advancements and challenges in making video content accessible through automated transcription and machine translation. Technologies like YouTube's Transcript API and tools such as Google's Cloud Translation API have revolutionized the way video content is transcribed and translated. These tools allow users to extract video captions and translate them into multiple languages, thus breaking down language barriers and broadening access to global audiences. However, several challenges

remain. First, automatic transcription accuracy varies based on factors such as audio quality, speaker accents, and noise levels, which can lead to errors in the text that affect translation quality. Furthermore, while deep learning models like Google Translate and DeepL offer multilingual capabilities, they still face difficulties with complex phrases, idioms, and cultural context, which can result in translations that may not be entirely accurate or contextually relevant. The integration of these technologies into web applications, such as YouTube transcript translators, offers significant potential in improving cross-language accessibility. However, ongoing research is needed to enhance the effectiveness of these systems, especially in terms of real-time translation and ensuring more precise linguistic accuracy across diverse languages.

3. EXISTING SYSTEM

The automatic transcription tools, like YouTube's built-in automatic captions, or third-party APIs like YouTubeTranscriptApi, that are currently in use for YouTube transcript translation and subtitle generation make up the majority of the system in place. However, these systems often face limitations regarding the translation of the transcripts into multiple languages or managing long video transcriptions. For instance, YouTube's automatic captions only support a limited number of languages and are not always accurate, particularly when it comes to videos that contain a lot of complex terminology or speech that is accented. Additionally, translation tools that attempt to offer multi-language support often struggle with large transcript sizes, causing issues with length limits or failing to deliver coherent translations. Many existing systems also lack user-friendly interfaces for easy integration and interaction. For instance, some platforms rely on basic tools for extracting transcripts but lack comprehensive translation capabilities. While Google Translate

is a popular choice for translations, it does not seamlessly integrate into YouTube transcript workflows, and users must manually input the transcript. Furthermore, such tools often do not provide features to compare original and translated content side by side, leaving users with an incomplete experience. Thus, a solution like the YouTube Transcript Translation Web Application addresses these gaps by automating the translation process, splitting large transcripts for efficiency, and offering a user-friendly interface. This tool’s multi-language support, including languages such as Tamil, Spanish, and Hindi, ensures broader accessibility for global audiences seeking to engage with YouTube content in their preferred language.

YouTube transcript summarizers are tools designed to extract, process, and condense the text from video transcripts into meaningful and concise summaries. These systems utilize various technologies, including Natural Language Processing (NLP), Machine Learning (ML), and Artificial Intelligence (AI), to enhance efficiency and accuracy. The summarization process typically involves multiple stages, starting with transcript extraction, where the system retrieves the text using the YouTube API, web scraping, or Automated Speech Recognition (ASR) tools like Google Speech-to-Text, Whisper, or Deepgram for videos without subtitles. Once the transcript is obtained, preprocessing techniques such as text cleaning, tokenization, stopword removal, stemming, and lemmatization are applied to remove unnecessary elements like timestamps, filler words, and irrelevant phrases. The transcript is condensed by the summarization module using either extractive or abstractive techniques following preprocessing. Extractive summarization selects the most important sentences directly from the transcript using algorithms like TF-IDF, LexRank, TextRank, or BERT- based models, ensuring that the core message remains intact. On the other hand, abstractive summarization generates new sentences to summarize the transcript more naturally using deep learning models such as Seq2Seq with attention mechanisms, T5, BART, or Pegasus. These models interpret the meaning of the text and generate a human-like summary rather than merely extracting key sentences. Once the summarization is complete, post-processing techniques refine the summary by checking grammar, coherence, and readability, often with the help of AI- based language models or tools like Grammarly. Additionally, users may have the option to adjust the length and depth of the summary based on their preferences.

There are several platforms and tools available that offer YouTube transcript summarization services. Online platforms such as Resoomer, SMMRY, and Summarize Bot allow users to upload transcripts and receive summarized versions. Browser extensions like Glasp and Eightify integrate directly with YouTube to provide instant video summaries, while AI-powered services such as ChatGPT, Claude AI, Otter.ai, and Sonix offer more advanced summarization features with additional context understanding. These tools are particularly useful for students, researchers, and professionals who need quick access to key insights from lengthy videos without having to watch them entirely. However, YouTube transcript summarization comes with several challenges. Since the quality of the summary is directly related to the accuracy of the transcript, accuracy is a major issue. If the transcript is automatically generated with errors due to accents, background noise, or fast speech, the summarization results may be unreliable. Extractive summarization methods may also miss the broader context, as they focus only on the most frequent or statistically relevant sentences, whereas abstractive summarization models

sometimes generate misleading or incorrect summaries due to limitations in understanding nuanced content. Another challenge is handling long transcripts, as many AI models have token limits, making it difficult to process extended video content efficiently. Furthermore, real-time summarization remains a complex task, as summarizing live videos with minimal delay requires significant computational power and optimized algorithms.

Despite these difficulties, YouTube transcript summarizers are getting better and more effective as AI and NLP advances

continue. The development of more advanced transformer-based models such as GPT-4, Gemini, and Claude has led to more context-aware and accurate summarization. Multi-modal summarization, which integrates video and audio analysis along with transcripts, is an emerging trend that enhances the depth of summarization by incorporating visual and tonal cues. Future systems may also offer customizable summaries, allowing users to select the level of detail, key points, and preferred summarization style. Additionally, the integration of summarization tools with virtual assistants like Google Assistant or Siri could further enhance user convenience by providing spoken summaries on demand. As AI technology continues to evolve, YouTube transcript summarizers will become faster, more accurate, and more adaptable to user needs, transforming the way people consume and analyze video content.

4. PROPOSED SYSTEM

The YouTube Transcript Translation Web Application aims to revolutionize how global users engage with YouTube video content by providing seamless translation of video transcripts into multiple languages. Built on a Flask framework, this tool combines YouTubeTranscriptApi to extract video transcripts and Deep Translator for translations, ensuring an accessible experience for diverse linguistic backgrounds. The system solves common problems in existing methods, particularly when dealing with long videos that exceed character limits for translation. By splitting the transcript into manageable chunks, it efficiently handles large data sets, maintaining the coherence and integrity of the translations. Users can easily input the URL of any YouTube video, and the application fetches the transcript instantly. The interface is designed to allow users to select one or more languages for translation, which includes popular languages such as Tamil, Spanish, French, Hindi, Chinese (Simplified), German, and Japanese. Once translated, users can view the original transcript alongside the translations, which enhances cross-lingual understanding and content accessibility. This application is designed with content creators, educators, and general viewers in mind, catering to individuals who wish to understand videos in languages that differ from the original audio. The ability to translate content into multiple languages makes it an essential tool for global accessibility, particularly for educational materials and cross-language learning. By enabling wider engagement, it also helps YouTube content reach international audiences, improving the inclusivity of digital media.

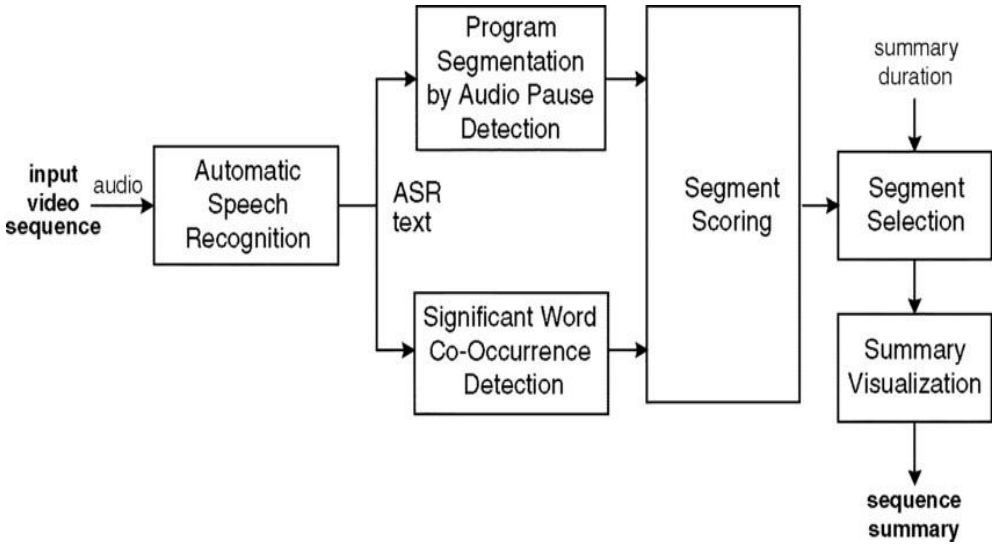


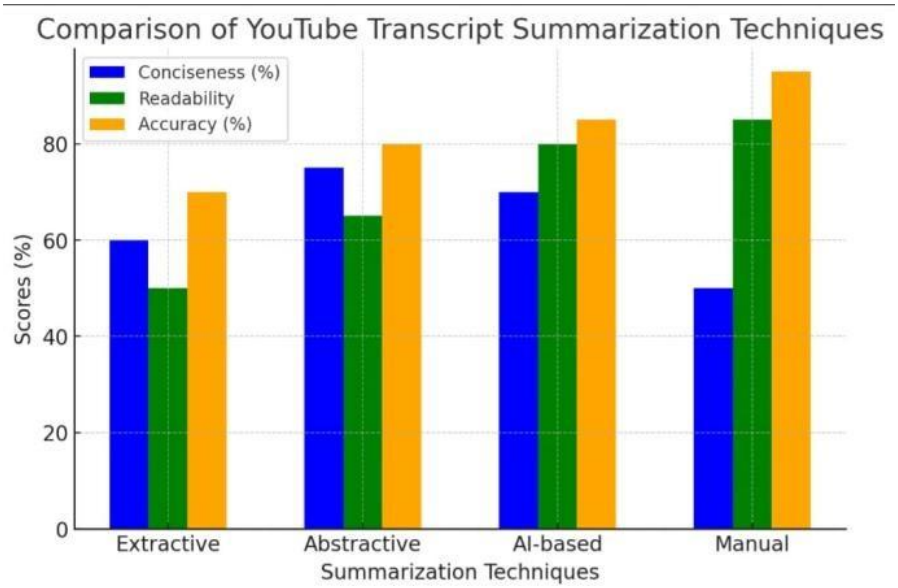
Figure 1: Data Flow Diagram

The primary objective of this system is to streamline content consumption by providing structured and meaningful summaries. The proposed system will operate in multiple stages. Utilizing the YouTube API, it will first extract the transcript from the specified YouTube video URL. If a transcript is available, it will be retrieved; otherwise, the system will notify the user about the unavailability of a transcript. The extracted transcript will then be preprocessed to remove unnecessary elements like timestamps, speaker names, filler words, and punctuation errors. Text normalization techniques, including lowercasing, stopword removal, and tokenization, will be applied to enhance the quality of text data. After preprocessing, the system will employ text summarization algorithms to generate concise outputs. Two main summarization techniques will be incorporated: Extractive Summarization, which selects key sentences directly from the transcript, and Abstractive Summarization, which uses deep learning models to generate human-like summaries. Users will have multiple summarization options, such as short summaries (20% of the transcript), medium summaries (50% of the transcript), and bullet-point summaries for quick insights.

The proposed system will feature a user-friendly web interface where users can enter a YouTube URL and receive a summary within seconds. The interface will support multiple languages, ensuring accessibility for a global audience. Additionally, users will have the option to download summaries or share them across platforms. The system will be built using Python for

backend processing (Flask/Django), NLP libraries like NLTK, spaCy, and Transformer-based models (T5, BART, GPT-based models) for summarization. The frontend will be developed using React or Angular to ensure a seamless user experience. A database, such as Firebase or MySQL, may be integrated for storing summaries, allowing users to revisit previously generated content.

Figure 2: Graph for Comparison of YouTube Transcript Summarization Techniques



This system is expected to significantly enhance user productivity by reducing the time required to extract key insights from YouTube videos. It will be particularly beneficial for students, researchers, journalists, and professionals who need to analyze large amounts of information efficiently. Moreover, the system has potential applications in accessibility, as it can help individuals with hearing impairments by providing concise text- based representations of video content. Future enhancements may include AI-driven sentiment analysis to gauge the tone of a video, topic categorization to classify content into relevant subjects, and personalized summarization based on user preferences. Additionally, integration with note-taking applications and cloud storage platforms could further enhance the system's usability.

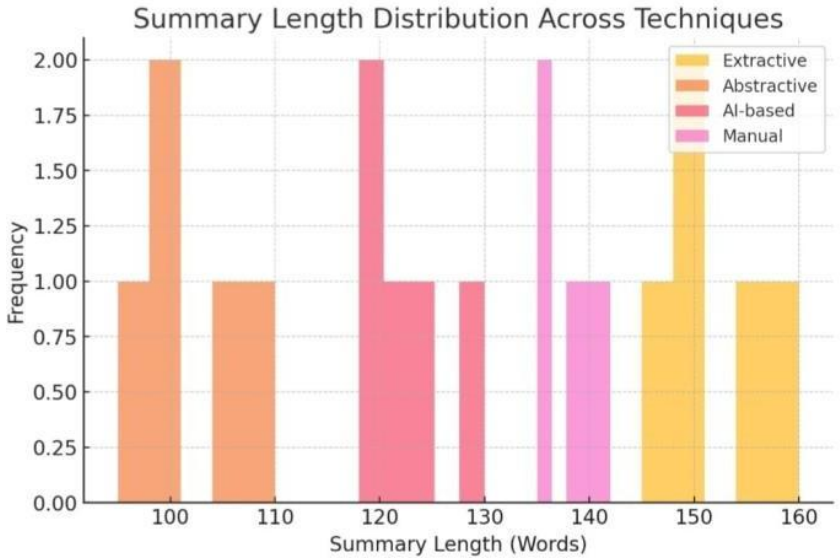


Figure 3: Graph for Summary Length

In the YouTube Transcript Summarizer will serve as a powerful tool for modern content consumption, bridging the gap between long-form video content and quick information retrieval. By leveraging NLP and AI-driven summarization techniques, it will help users save time and improve accessibility to valuable insights. With the growing reliance on video

content, this system has the potential to revolutionize the way people engage with YouTube transcripts, making information more digestible and user-friendly.

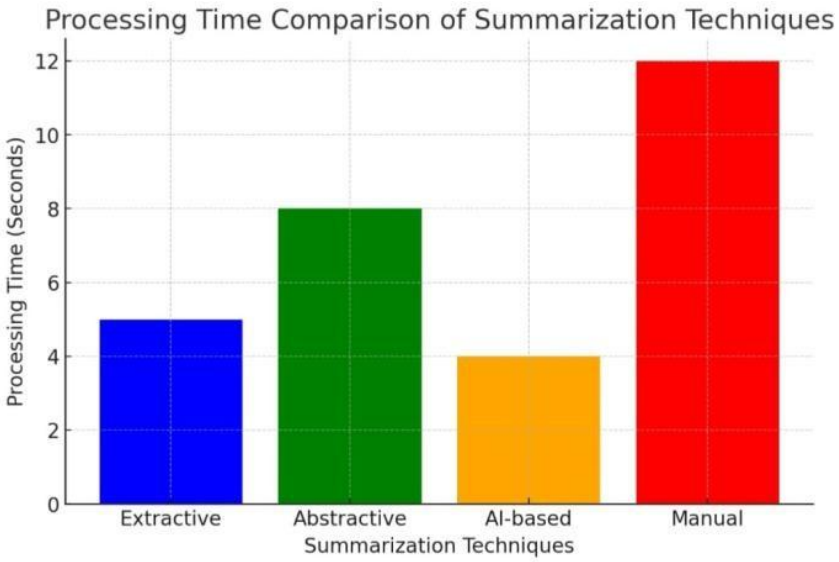


Figure 4: Graph for Processing Time Comparison

5. CONCLUSION

In conclusion, the YouTube Transcript Translation Web Application offers a powerful, multi-language solution for enhancing global accessibility and bridging language barriers in online video content. By extracting and translating transcripts, this application caters to a diverse audience, making YouTube videos more accessible to users who may not speak the original language. The system's ability to handle large transcript texts by dividing them into manageable chunks ensures smooth translations without query length limitations. The range of supported languages, such as Tamil, Spanish, French, and Chinese, ensures that users from different linguistic backgrounds can benefit. This tool not only facilitates content creators and educators by reaching a broader audience but also empowers users to engage with international content more effectively. Ultimately, the YouTube Transcript Translation Web Application serves as an essential tool for promoting cross-cultural communication, improving video content accessibility, and fostering global knowledge sharing. The development of a YouTube transcript summarizer has the potential to revolutionize the way we consume video content. By automating the process of summarizing lengthy videos, this technology can save significant time and effort for users. Through the integration of advanced natural language processing techniques and machine learning models, it is possible to generate accurate, concise, and coherent summaries.

6. FUTURE WORKS

The future development of the YouTube Transcript Translation Web Application can focus on enhancing its language support and translation accuracy. Expanding the language options, including regional dialects, can significantly broaden the user base. Additionally, incorporating machine learning techniques such as neural machine translation could improve the quality of translations, providing more context- aware outputs. Implementing voice-to-text capabilities to transcribe videos that lack subtitles could also be a valuable feature, further improving accessibility for users with different language backgrounds. Another area for future work involves optimizing the translation process by reducing processing time, particularly for longer videos, to ensure a smoother user experience. The tool's versatility may be increased by its integration with additional video platforms like Dailymotion or Vimeo. Lastly, a collaborative feature that lets users make

translations or corrections would help the system get better over time and encourage a community-driven strategy for making global content more accessible.

REFERENCES

[1] Jaiswal, Shubhangi, and Manoj Misra. "Automatic indexing of lecture videos using syntactic similarity measures."2018 5th International Conference on Signal Processing and Integrated Networks (SPIN). IEEE, 2018.

[2] Pradeep Choudhary, Sowmya P. Munukutla, K. S. Rajesh, Alok S. Shukla "Real time video summarization on mobile platform" International Conference on Multimedia and Expo (ICME), 2017 IEEE

[3] Rajkumar Kannan, Gheorghita Ghinea, Sridhar Swaminathan, Suresh Kannaiyan "Improving video summarization based on user preferences" 2013 Fourth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)

[4] Jayanta Basak, Varun Luthra and Santanu Chaudhury "Video Summarization with Supervised Learning" 2008 IEEE.

[5] Wei REN Yuesheng ZHU "A Video Summarization Approach based on Machine Learning" International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2008 IEEE

[6] Taskiran, Cuneyt M., et al. "Automated video summarization using speech transcripts." Storage and Retrieval for Media Databases 2002. Vol. 4676. International Society for Optics and Photonics, 2001.

[7] Rohit Anand, Gulshan Shrivastava, Sachin Gupta, Sheng- Lung Peng, Nidhi Sindhwani " Audio Watermarking With Reduced Number of Random Samples" In Handbook of Research on Network Forensics and Analysis Techniques (pp. 372-394). IGI Global.

[8] Garima Bakshi, Rati Shukla, Vikash Yadav, Aman Dahiya, Rohit Anand, Nidhi Sindhwani and Harinder Singh "An Optimized Approach for Feature Extraction in Multi-Relational Statistical Learning" Journal of Scientific and Industrial Research (JSIR).

[9] W. Wang, E. Xie, X. Li, W. Hou, T. Lu, G. Yu, and S. Shao, ''Shape robust text detection with progressive scale expansion network,'' in Proc.

[10] IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019, pp. 9336–9345.

[11] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, ''Feature pyramid networks for object detection,'' in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 2117–2125.

[12] P. Yang, G. Yang, X. Gong, P. Wu, X. Han, J. Wu, and C. Chen, ''Instance segmentation network with self-distillation for scene text detection,'' IEEE Access, vol. 8, pp. 45825–45836, 2020.

[13] Y. Sun, J. Liu, W. Liu, J. Han, E. Ding, and J. Liu, ''Chinese street view text: Large-scale chinese text reading with partially supervised learning,'' in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2019, pp. 9086–9095.

[14] B. U. Kota, K. Davila, A. Stone, S. Setlur, and V. Govindaraju, ''Automated detection of handwritten whiteboard content in lecture videos for summarization,'' in Proc. 16th Int. Conf. Frontiers Handwriting Recognit. (ICFHR), Aug. 2018, pp. 19–24.

[15] B. U. Kota, K. Davila, A. Stone, S. Setlur, and V. Govindaraju, ''Generalized framework for summarization of fixed-camera lecture videos bydetecting and binarizing handwritten content,'' Int. J. Document Anal.Recognit., vol. 22, no. 3, pp. 221–233, 2019.T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and

S. Belongie, ''Feature pyramid networks for object detection,'' in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 2117–2125.

[16] P. Yang, G. Yang, X. Gong, P. Wu, X. Han, J. Wu, and C. Chen, ''Instance segmentation network with self-distillation for scene text detection,'' IEEE Access, vol. 8, pp. 45825–45836, 2020.

[17] Y. Sun, J. Liu, W. Liu, J. Han, E. Ding, and J. Liu, ''Chinese street view text: Large-scale chinese text reading with partially supervised learning,'' in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2019, pp. 9086–9095.

[18] M. R. Rahman, S. Shah, and J. Subhlok, ''Visual summarization of lecture video segments for enhanced navigation,'' in Proc. IEEE Int. Symp. Multimedia (ISM), Dec. 2020, pp. 154–157.

[19] M. Husain and S. M. Meena, ''Multimodal fusion of speech and text using semi-supervised LDA for indexing lecture videos,'' in Proc. Nat. Conf. Commun. (NCC), Feb. 2019, pp. 1– 6.

[20] P. Banerjee, U. Bhattacharya, and B. B. Chaudhuri, ''Automatic detection of handwritten texts from video frames of lectures,'' in Proc. 14th Int. Conf. Frontiers Handwriting Recognit., Sep. 2014, pp. 627–632.