VIDEO TRANSCRIPT SUMMARIZER

Ilampiray P[1*], *Naveen Raju* D[1], *Thilagavathy* A[1], *Mohamed Tharik* M[1], *Madhan Kishore* S[1], *Nithin* A.S[1], *Infant Raj* I[2]

[1]Department of Computer Science and Engineering, R.M.K. Engineering College Thiruvallur, Tamil Nadu, India.
[2]Department of Computer Science and Engineering, K.Ramakrishnan College of Engineering, Trichy - 621112, Tamil Nadu, India.

ABSTRACT. In today's world, a large number of videos are uploaded in everyday, which contains information about something. The major challenge is to find the right video and understand the correct content, because there are lot of videos available some videos will contain useless content and even though the perfect content available that content should be required to us. If we not found right one it wastes your full effort and full time to extract the correct usefull information. We propose an innovation idea which uses NLP processing for text extraction and BERT Summarization for Text Summarization. This provides a video main content in text description and abstractive summary, enabling users to discriminate between relevant and irrelevant information according to their needs. Furthermore, our experiments show that the joint model can attain good results with informative, concise, and readable multi-line video description and summary in a human evaluation.

1 Introduction

NLP, a subset of Artificial Intelligence, is a field that concentrates on the interaction between machines and human languages. Its primary objective is to enable machines to understand, interpret, and generate natural language text or speech.. Video summarization, which involves generating concise and accurate summaries of longer videos, is an important application of NLP. The goal is to produce short and coherent summaries that capture the key information of the original video. This technology can be valuable in situations where time is limited, or when there is a need for a quick overview of the video content. Video summarization typically involves a combination of techniques such as text extraction, audio analysis, and image processing, among others. Our main idea is to able to find the short summary of YouTube video and present it in a textual format. As short summarization of text methods will very helpful for users because we can read the important content in short period of time. In the field of NLP we generate summaries of transcripts and produce human-readable outputs. Nowadays from child to older people are easy to YouTube video for many

* Corresponding author: ilampiray.vp@gmail.com

purposes like educational, entertainment and many other kinds of genre, it is very necessary to find out exact content we required. In the Internet we can see that so many are videos are long but not even contain good useful information which wastes our time. We can directly move to the main to content of YouTube video by removing the useless information of the videos. The main goal of our this paper is to increase work efficiency and save the time of a user. There are many users just seeing the thumbnail of the Youtube video and catchy title turns eager into the video, wastes the time by watching useless information. Students often search for YouTube videos before exams, but due to time constraints, they may watch them at double speed, which can lead to confusion about the subject matter. Having access to recorded sessions and transcripts of meetings can be helpful in obtaining a summary of the video content, saving time and effort. The main focus of our paper is to extract the most important information from the transcript and present it in a concise paragraph. Our objective is to save users' time by providing them with relevant and useful information on their desired topic. Automatic text summarization techniques have been developed in recent years to facilitate this process. Text summarization involves transforming text into a condensed version that conveys the main message to users. Although text summarization is a challenging task due to the limitations of machines in understanding human language and knowledge, it offers various benefits such as content organization, summarization, data retrieval, and question answering. While earlier studies primarily focused on simplifying and summarizing single document texts, advancements in technology have paved the way for more efficient and rapid text summarization methods.

2 Literature survey

The research paper introduced Latent Dirichlet Allocation (LDA) as an effective technique for document summarization. Their proposed LDA summarizing model consists of three stages [1]. In the initial phase, the subtitle file undergoes pre-processing, including the removal of stop words and other tasks. The LDA model is then trained using the subtitles in the second step to generate a list of keywords, which will be utilized to extract relevant sentences. Finally, in the third phase, the summary is generated based on the extracted keywords. Comparatively, the quality of the summaries produced by the LDA-based approach surpasses that of TF-IDF and LSA summaries. The paper also presented Stream Hover, a platform designed for explaining and summarizing transcripts of live streamed videos [2]. They explored a neural extractive summarization model that learns vector representations of audio files and extracts significant observations from subtitles. These observations are utilized to construct summaries using a vector quantized autoencoder. Additionally, the paper proposed a system that can generate subtitles for movies in English, Hindi, or Malayalam, depending on user preference. The system comprises three components: audio extraction, voice recognition, and subtitle generation. For audio extraction, the FFMPEG platform is employed to convert audio files of any format into .wav (Waveform Audio) format [3]. The .wav file obtained from the audio extraction process is utilized to generate subtitles in the form of a .srt file. The content of the audio, obtained through the Google Translate API, is processed for speech attention. In the Subtitle creation module, the synchronized lyrics from the .srt file are combined with the video using the Moviepy video enhancement library in Python [4]. The research paper introduces a system that generates abstractive summaries for videos covering various topics such as cooking, cuisine, software configuration, and sports. To expand the vocabulary, the model is pre-trained on large English datasets using transfer learning. Transcript pre-processing is also performed to improve sentence structure and punctuation in ASR system results [5]. The evaluation of the results on the How2 and WikiHow datasets involves the use of ROUGE and Content-F1 scoring metrics. The paper categorizes different Text Summarization Methods, with a focus on giving more importance

to abstractive text summarization [6]. The authors express their belief that abstractive summarization, despite being more challenging and computationally intensive than extractive summarization, holds greater potential for generating more natural and human-like summaries. This suggests that there may be further advancements in this field, offering new perspectives from computational, cognitive, and linguistic standpoints. The study recommends the ASoVS model, a hybrid end-to-end approach, for generating video descriptions and text summaries in an abstractive manner [7]. The model incorporates a deep neural network and captures various aspects of the video, such as people's traits (gender, age, emotion), scenes, objects, and behaviors, to provide a multi-line description. The utilization of OCR technology enables the conversion of images into text. The paper outlines the strategies employed for subtitle generation, which involves three modules: Audio Extraction, responsible for converting MPEG-compliant input files to .wav format; Speech Recognition, which utilizes Hidden Markov Models (HMMs) to recognize extracted speech using language and acoustic models; and Subtitle Generation, which produces synchronized .txt/.srt files. This approach is particularly beneficial for individuals who are deaf, have reading difficulties, or are learning to read [8]. The study also presents a multilingual speech-to-text conversion method that involves feeding human voice utterances into a Speech-To-Text (STT) system, utilizing Mel-Frequency Cepstral Coefficient feature extraction, Minimum Distance Classifier, and Support Vector Machine techniques for voice classification [9]. The evaluation of text summaries in the paper is done using ROUGE Metrics (Recall-Oriented Understudy for Gisting Evaluation), which compare the computer-generated summary with human-written ideal summaries by measuring the overlapping units such as n-grams, word pairs, and word sequences [10]. The systems discussed in the papers [11] are not applicable to videos that lack readily available subtitles. Additionally, these systems are limited to processing English videos only. Furthermore, the absence of a media player in the proposed system [11] requires the entire video to be uploaded for subtitle generation. In contrast, the system described in the work [12] generates video descriptions solely based on the visual content, without considering the audio. This system is particularly suitable for generating descriptions for CCTV footages. However, it should be noted that the system in [12] is designed for audio files and is not compatible with video files. The system presented in the article [3] exclusively works with text input, lacking the capability to extract subtitles from videos or generate subtitles for videos. In the work [13], the focus is solely on translation rather than text summarization. The generated output in this system is the translated version of the text obtained through speech recognition. Moreover, the paper [14] only addresses the extraction of subtitles from videos, without providing a summarized version of the text. The vast amount of information available on the internet can be overwhelming and emotionally draining for individuals. To address this issue, summaries are employed to condense texts into a more concise form while retaining the essential information. The main goal of a summary is to effectively convey the key information in a concise manner. However, generating a useful summary can be challenging, particularly when the original document contains repetitive sentences. In such cases, reducing the document size would result in a loss of content [15]. This challenge is known as automatic text summarization, which aims to create a compact and effective summary that captures the crucial information and overall meaning of the original document. Automated text summarization is a complex task as computers lack the linguistic understanding and knowledge possessed by humans. Consequently, automated text summarization requires advanced technologies and techniques and is a time-consuming process. Nonetheless, despite the challenges involved, text summarization is gaining importance in fields such as journalism, research, and content creation, where the ability to extract key information quickly is vital.

3 Methodology

This paper majorly focus on providing clean, clear and correct summary of the YouTube videos that users don't need to waste their time at. This paper use more popular python libraries each for each purpose. They are YouTube Transcipt api for transcript extraction, BERT is for summarization library it is combination of GPT and BART, google translate api for translation and Flask framework is for Backend Connection in Fig.1.

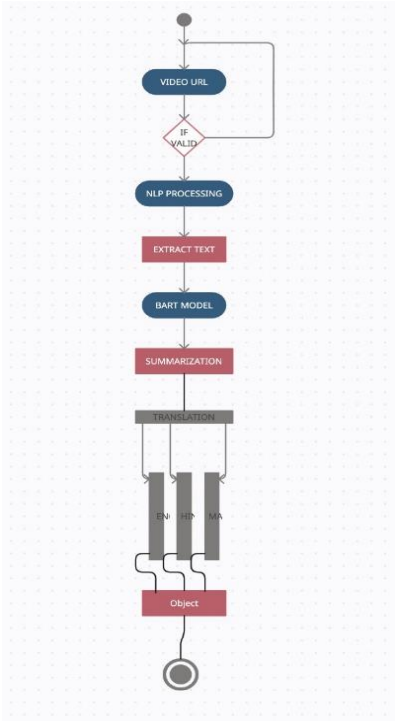

Fig. 1. System Architecture

3.1 Getting URL

A Uniform Resource Locator (URL) is a reference to an internet resource that provides the location of that resource on a computer network. It serves as a means to retrieve information from the internet. When a user enters a YouTube URL in the search box, the system processes the URL in the background to obtain the necessary information. Then check whether it is valid URL or NOT, If it is valid URL then shortened the link process will happened in the backend. Then required YouTube video will get from the link. It will process for the Video To Text Extraction. In this user will enter the URL, we will check whether the URL is valid or not. If the URL is invalid it will return the error message. If the URL is valid, then it will pass correct URL in next step that it will pass the URL to NLP model.

3.2 Video to text extraction

After the required YouTube video is attained, then we will pass the URL from NLP model file to utubeextract.py file. Then by using the youtube_transcript_api we will get the transcript for required YouTube video. There are three different ways to transcript one is

automatically transcript generated, second-one is manually transcript generated and third-one is videos that doesn't contain transcript. Videos without transcripts are being excluded from our selection. Once we obtain the transcript of a video using an API, we proceed to process the text for subsequent procedures. This involves removing commas, punctuation marks, and full stops, as they are essential in determining sentence boundaries. This process can done by using the python library "punctuator. Then our next step is to apply the text preprocessing method for extracted transcript. This above method is like removing punctuations, stopword, exclamatory expression, stemming and for scanning purposes. The purpose of above task is to shortening a wide length of paragraph into a short summary. That summary will contain important information in the video. In NLP text summarization we tend to use the Extractive Summarization. This Extractive summarization delivers the summary that contains only the important phrases and sentences. The site will continuously monitor the data in the user given web page and if there is any change in the amount specified, an immediate alert is passed to the user or customer. Once the user provides the required information, our website will scrap the required data from the product's web page. For scrapping the data our website uses beautiful soup module. It is a Python module, which is basically used to scrap the data from website in Fig.2.

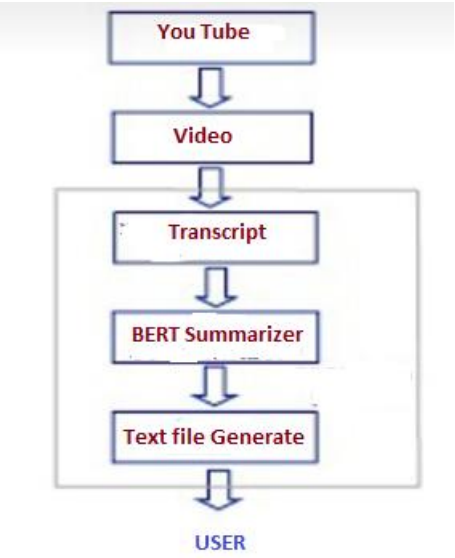

Fig. 2. Text Extraction

3.3 NLP processing

This Process is for Text Summarization. NLP contains many method for text summarization. We will talk about the text summarization is process of shortening the big paragraph into short summary. If the paragraph contains lots of lines, we need more time to cover the content, we have time scarcity so we want only a main report of that text. We can able to convert the large text into to small text by removing unimportant information. The process of breaking down lengthy text into digestible paragraphs or sentences is known as NLP Text Summarization. This above process will retrieve the important information in Fig.3.

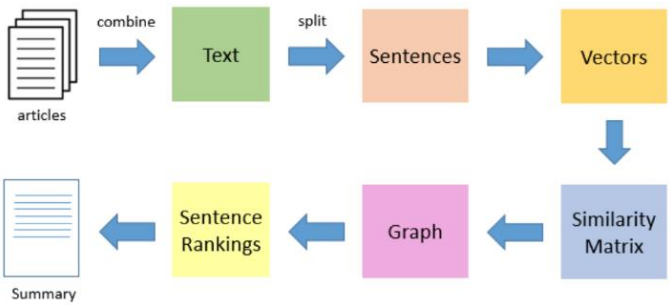

Fig. 3. NLP Processing

3.4 BERT summarizer

BERT is Bidirectional Encoder Representations from Transformers. This BERT introduces new advanced approach to deliver NLP tasks. BERT is one of the important algorithm of Natural Language Processing Models. The important information can be retained and extracting the relevant information by Extractive Text Summarization. This Extractive Summarization is more challenging. In this progress we using superior embeddings that provided by encoder models like BERT. By taking two supervised approaches and use the BERT sentence embeddings to build an extractive summarizer. The first considers only embeddings and their derivatives. Based on the internal structure of the article we can able to good summarizer can parse meaning and select correct meaningful sentences. The Unsupervised *Text-Rank* model is the baseline approach. Other than approaches are incorporates the sequential information and advantage of well known particular to new corpuses. In fact, many publishers have been deployed this strategy. Baseline for the second approach is Lead. Rouge-1 and Rouge-L F1 metric are important things in supervised models outperform in Fig.4.

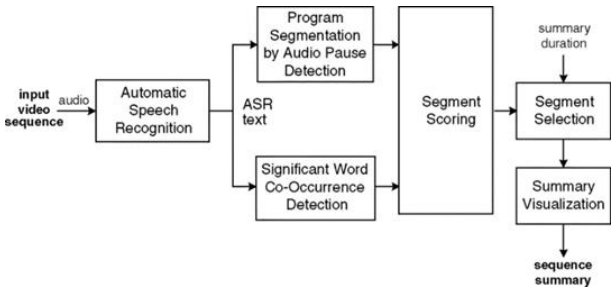

Fig. 4. BERT Summarization Diagram

3.5 Language translation

A language translator is a very handy application that helps us to communicate in different languages by translating our language to the desired language. In earlier times, when there were no Language Translation Applications, it was very difficult for people to communicate with people coming from different parts of the world. In this we can create our own language translation project using python. Our objective is to create a Language Translator which would help us translate a word, sentence or even a paragraph to another language. We will try to incorporate as many languages as possible. We will be using T-kinter Module to build our GUI for the project and google-trans library to present us with a number of languages

that are a part of it.

4. Results and discussions

Initially the verifying the valid URL or not, if valid we will check whether the transcript is available or not. Then we will go for fetching the required YouTube video and using YouTube transcript api we will fetch the transcript of YouTube video. Videos lacking transcripts are being filtered out. Once we acquire the video transcript through an API, we proceed to process the text for subsequent steps. This involves eliminating commas, punctuation marks, and full stops, which are crucial for identifying sentence boundaries. This process can done by using the python library "punctuator. Then our next step is to apply the text preprocessing method for extracted transcript in Fig. 5.

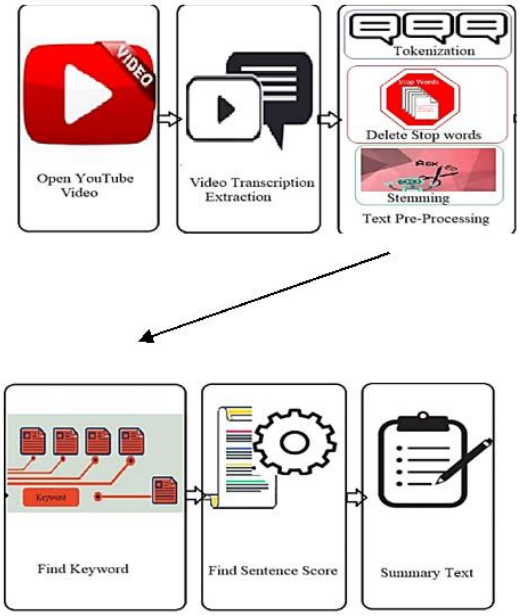

Fig. 5. Block Diagram of Proposed System

This paper is used for following functions
Crash-Course: Students can able to find out a particular topic of the subject in the you-tube video and precisely able to get the quick read of the youtube video.
Education in the online era has paved the way for generating summaries and allowing students to create their own notes based on video class transcripts. Similarly, in video and audio conferences, transcripts can be used to outline discussions held during team conferences. In the context of patent research, the extraction of crucial claims from patents and research papers can be facilitated using figures such as Fig. 6, Fig. 7, and Fig. 8.

Fig. 6. Extractive Summarizer

Video Title	Total time of video (minutes)	Summary time requested by the user (minutes)	Processing time (minutes)	Memory usage (MB)
How to have constructive conversations | Julia Dhar	11	3	3	190.23
How I Built 5 Income Sources That Make $42,407 Per Month	17	6	10	170.03
Self-Taught Programmer vs Coding Bootcamp vs Computer Science Degree	17	5	4	150.63

Fig. 7. Metric Table of Our Results

Video Link	Total time of video (minutes)	Summary time requested by the user (minutes)	Summary time formed by our algorithm (minutes)
https://www.youtube.com/watch?v=BFZsNN6eNvQ	11	3	2 min 55 sec
https://www.youtube.com/watch?v=2xSKCAtWIyo	17	6	5 min 50 sec
https://www.youtube.com/watch?v=LN06tzw7mb0	15	4	3 min 55 sec

Fig. 8. Metric Table of Processing time and memory Usage

4.1 Screenshots

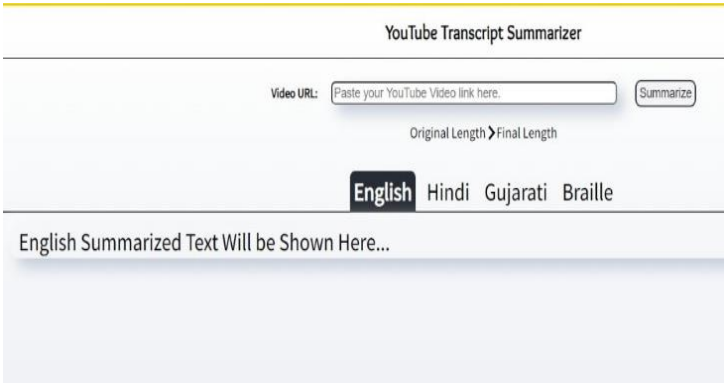

Fig. 9. Home Page (In this Home page, the user paste URL in video URL box, The URL which is copied from the YouTube video)

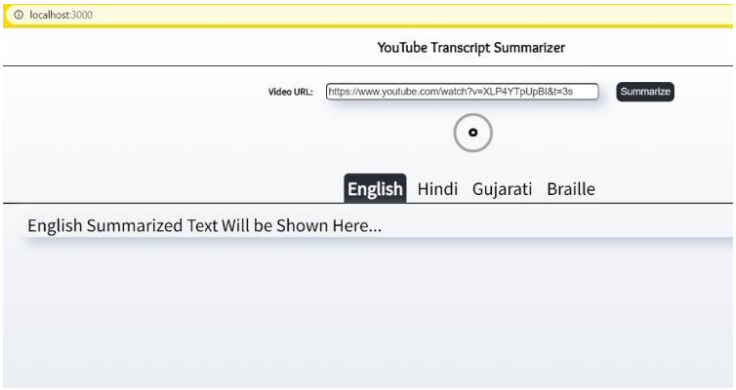

Fig. 10. Home page (after paste the link, in backend required video will process and it takes some time)

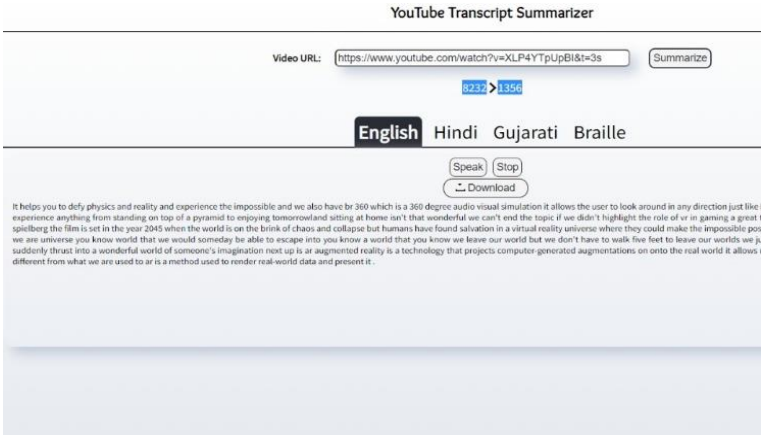

Fig. 11. Output Page (here the summarized text in English, It also display the comparison between word count of before and after summarization)

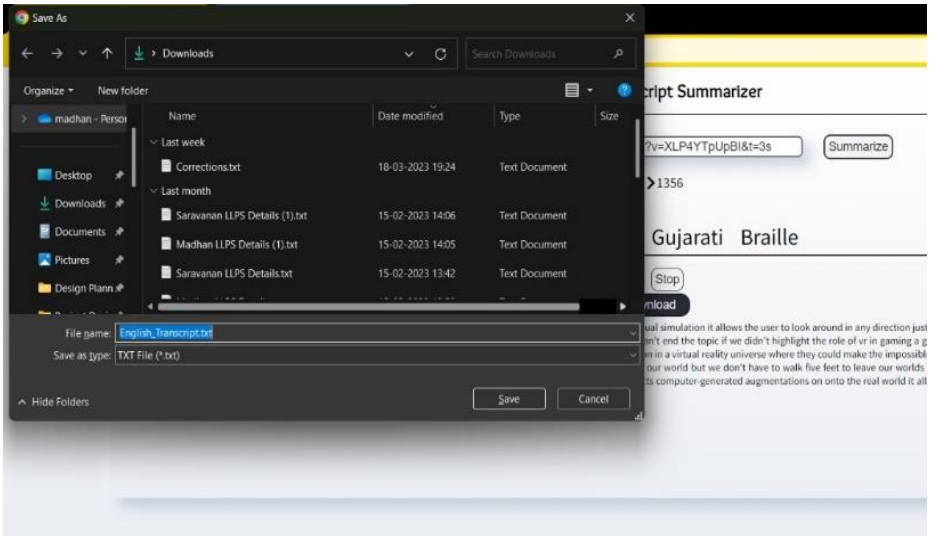

Fig. 12. Here we can able to download summarized text and save it as text file.

5 Future works

In the future, we intend to extend our work in the based of extensions. This extension need to be available in all the browser, social media, and all other Video website like(YouTube, Share-Chat etc). User only purpose is need to add the extension and need to select the required video It will automatically fetch the link and display the required shortened summary. browser and machine learning and artificial intelligence technology.

6 Conclusions

In conclusion, our website can save time for the user .Instaed of Seeing the whole buffer waste content of the video we will prefix see the what is main content of YouTube Video and Know which video will perfectly for us. It save the time and effort of the user. By using our website their burden will be reduced for search the right YouTube Video. Our website will also provide Multi-Language Summarization and made the availability to Text-to-Speech Process also. We are confident that our paper will effectively address the needs of users by saving their time and efforts. Our approach aims to provide users with only the relevant and useful information on the topics that interest them, eliminating the need to watch lengthy videos. This time saved can be utilized for further knowledge acquisition and exploration.

REFERENCES

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

2 Sangwoo Cho, Franck Dernoncourt , Tim Ganter, Trung Bui, Nedim Lipka, Walter Chang, Hailin Jin, Jonathan Brandt, Hassan Foroosh, Fei Liu, "StreamHover: Livestream Transcript Summarization and Annotation", (2022).

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., (2016).

4 Ghadage, Yogita H. and Sushama Shelke. “Speech to text conversion for multilingual languages.” 2016 International Conference on Communication and Signal Processing (ICCSP) (2016).

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), **6**, 9 (2019).

6 S.M.Mahedy Hasan, Md. Fazle Rabbi, Arifa Islam Champa, Md. Asif Zaman, “An Effective Diabetes Prediction System Using Machine Learning Techniques”, 2nd International Conference on Advanced Information and Communication Technology (ICAICT), (2020).

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), **3**, 3 (2016).

8 Aiswarya K R, “ Automatic Multiple Language Subtitle Generation for Videos, ” International Research Journal of Engineering and Technology (IRJET), **7**, 5 (2020).

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

10 Patil, S. et al. “Multilingual Speech and Text Recognition and Translation using Image.” International journal of engineering research and technology **5** (2016).

11 S. Sah, S. Kulhare, A. Gray, S. Venugopalan, E. Prud'Hom meaux and R. Ptucha, "Semantic Text Summarization of Long Videos," IEEE Winter Conference on Applications of Computer Vision (WACV), (2017).

12 A. Dilawari and M. U. G. Khan, "ASoVS: Abstractive Summarization of Video Sequences," in IEEE Access, **7** (2019).

13 Lin, Chin-Yew, “ROUGE: A Package for Automatic Evaluation of Summaries,” In Proceedings of 2004, Association for Computational Linguistics, Barcelona, Spain.

14 A. E. B. Ajmal and R. P. Haroon, “Maximal marginal relevance based malayalam text summarization with successive thresholds,” International Journal on Cybernetics and Informatics, **5**, 2 (2016).

15 M. Allahyari et al., “Text summarization techniques: A brief survey,” International Journal of Advanced Computer Science and Applications, **8**, 10 (2017).