iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** III **Month of publication:** March 2024

DOI: https://doi.org/10.22214/ijraset.2024.59243

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

Text and Video Summarization Using ML

A. Karthik[1], Sumit Chelluru[2], Y. Sanjana[3], Dr. S. Kirubakaran[4]
[1, 2, 3]*UG Student, Department of CSE, CMR College of Engineering & Technology, Hyderabad, Telangana*
[4]*Professor, Department of CSE, CMR College of Engineering & Technology, Hyderabad, Telangana*

Abstract: ***The Text and Video Summarization tool is a revolutionary approach to manage time and immense amount of data; one must go through. Current text summarization tools struggle to summarize large text files accurately and consistently, especially when dealing with different document formats and file lengths. This paper efficiently handles condenses large amounts of text to meet the growing need for more efficient information consumption accessible to all with a user-friendly interface. By integrating innovative methodologies, advanced NLP techniques, and a diverse dataset enriched with human insights, this project aims to advance the field of automated text summarization, contributing to more effective information condensation. The tool streamlines text summarization as well as video summarization capabilities into one umbrella, providing its users option and choice of even summarizing audio as well, this system offers a seamless and responsive solution. It contributes to a more efficient use of existing resources by producing summaries, making time management efficient and effective.***

Keywords: Natural Language Processing, wav2vec2, BERT, Python, Summarization, Extractive

I. INTRODUCTION

A Text and video summarization tool is a comprehensive and streamlined application designed to facilitate the summarization of various formats of data, be it audio, video, and even conventional ways such as pasting texts from research papers. This application serves as a centralized hub where individuals, whether they are employees, customers, citizens, or other stakeholders, can submit their sources of information and can expect a generalized summary within a matter of few seconds depending on the size of the text or the video file's length. The primary goal of a Text and Video summary using ML system portal is to enhance the efficiency of individuals by generating efficient summaries by extraction and with context preservation approach, thus saving the hassle of going through multiple documents and resources before getting meaningful information. The paper [1], presents a summarization procedure based on the application of transformers. the paper [6], present some summarization techniques like feature based, clustering, short selection, event based, etc. in 2023 paper [5] discussed various video summarization techniques like using mean shift key based video summarization, deep learning, LSTM, faster RCNN, etc.

The objective of a Text and Video summarization is to provide direction to its users in terms of the document or resources before they could choose one to follow. This paper focuses on the use of various possibilities of summarization tool. Our objective is to develop a Text and video summarization application that streamlines the entire process of selecting the resources for study by providing context aware summaries for various resources, the support for different forms like recorded videos of MP4 formats, prerecorded audio in MP3 format, YouTube URLs, or simply by pasting text into the text box submission.

II. LITERATURE SERVEY

A. Text Summarization using NLP Technique

In this paper they proposed a framework of the algorithm, comprises four key phases: data pre-processing, training of the transformer T5 model, generating the summary, and evaluating the results using the ROUGE score.

Data pre-processing begins with a thorough analysis of the dataset to identify and remove any missing or null values. This ensures data integrity and reliability throughout subsequent phases. Tokenization, a crucial step, involves converting input sentences and words into tokens. Additionally, the maximum and minimum lengths of both articles and summaries are calculated, providing insights into the distribution of text lengths within the dataset. Mean lengths for each category are also computed, aiding in understanding the typical length characteristics of articles and summaries.

The transformer architecture, a cornerstone of modern NLP, is employed for sequence-to-sequence tasks, effectively managing long-range dependencies. Comprising multiple attention layers, the transformer network excels at recognizing intricate word patterns through positional encoding. The model has two main components: encoder and decoder. The encoder, with N=6 homogeneous layers, utilizes multi-head self-attention and fully connected feed-forward networks, incorporating residual connections and layer normalization for improved performance.

©IJRASET: All Rights are Reserved

Conversely, the decoder also consists of N=6 layers and incorporates multi-head attention to the output of the encoder stack. Attention mechanisms play a vital role in the transformer model, mapping input queries to key-value pairs and producing weighted sums of values based on compatibility functions.

The T5Base approach, or Text-to-Text Transfer Transformer, represents a significant advancement in transformer-based architectures. Operating on a text-to-text paradigm, T5 is a pre-trained encoder-decoder model trained on both unsupervised and supervised tasks. Leveraging multi-task learning, T5 transforms each task into a text-to-text format, enhancing its versatility and performance across various tasks.

Attention masks facilitate the self-attention operation in transformer models, ensuring that the input and output sequences maintain identical lengths during processing. The encoder-decoder transformer, comprising two layers, enables the creation of new output sequences from input sequences, leveraging fully visible attention masks to facilitate self-attention mechanisms.

B. From Video Summarization to real time video Summarization in Smart City

The paper divides into various video summarization techniques, recognizing the need for tailored approaches given the diverse user preferences and application domains. Summary-based methods encompass static, dynamic, hierarchical, multi-view, image, and text summaries. Static summarization, akin to keyframing, presents storyboard-like sequences, while dynamic summarization, such as Video Skimming, selects short dynamic sections or the most relevant shots. Hierarchical summaries offer multi-level abstractions, ranging from few keyframes at higher levels to more detailed representations at lower levels. Multi-View Summary (MVS) captures diverse perspectives simultaneously. Text summaries, generated via NLP techniques, provide paragraph-length textual summaries without audio or visual elements.

Preference-based summarization categorizes approaches into domain-specific, query-based, semantically-based, event-based, and feature-based strategies. Domain-based summarization further divides into pixel and compressed domains. Pixel-based methods extract information directly from frame pixels, while compressed domain approaches decode partially compressed video to extract features.

Information source-based summarization classifies methods as internal, external, or hybrid, depending on the sources of information utilized. Algorithms abstract semantics from video content and extract audio-visual cues from various phases of the video lifecycle.

Training strategy-based approaches aim to improve summary quality through effective feature extraction and model selection. Deep-learning-based algorithms fall into supervised, unsupervised, or semi-supervised categories. Supervised approaches leverage labelled data for training, while unsupervised methods use unlabelled data. Semi-supervised approaches combine both labelled and unlabelled data for training, enhancing model performance.

III. METHOD AND EXPERIMENTAL DETAILS

The proposed method for summarization we are going to use Gradio, Pytube, transformers, and others. Audio processing is a key component, utilizing libraries like Pyaudioconvert, transformers for Wav2Vec2.0, and the Bert summarizer. The code facilitates the extraction of information from YouTube videos by utilizing the Pytube library and processing the resulting audio with the assistance of MoviePy and Pydub. the text as summarization capabilities using the Bert-extractive-summarizer library.

We can convert the summary of text into pdf using Fpdf. Ffmpeg is a command-line tool and a multimedia framework that can be used to record, convert, and stream audio and video. The Gradio library is used to create a user-friendly web interface, enabling users to interact with the tool by providing YouTube video URLs, recorded video names, audio names, or text for summarization. The code is organized into modular functions and packages, enhancing readability and maintainability. the project adopts a systematic approach, making use of pre-trained models and algorithms to carry out tasks such as text summarization, punctuation, capitalization, and audio-to-text conversion.

Using the pytube the video in the link will be downloaded in background and the captions will be extracted, using that captions the text will be summarized using bert library. Overall, the implemented code showcases an integrated solution for AI-powered note generation with a user-friendly interface.

A. Taking the Input

The user can access the interface which is created using gradio and give inputs like text, video, etc. Gradio is an open-source python package used for web application. It passes the text or video link to the source code. It an easy-to-use interface. It also displays the results of the summary after summarization.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

B. Method used for text Summarization

BERT, a Bidirectional Encoder Representations from Transformers, summarises user text using dynamic connections between input and output elements. It's pretrained on masked language modelling (MLM) and next sentence prediction (NSP) tasks, learning from unlabelled text like Wikipedia and the Brown Corpus. This bidirectional approach enables BERT to continuously improve, even during practical applications like Google search. With its foundation in transformers, BERT effectively captures complex linguistic patterns, enhancing natural language understanding and processing.

C. Method used for Video Summarization

User-provided videos are summarized using BERT, pytube for downloading, and wav2vec2 for caption extraction. Wav2Vec2, a self-supervised NLP model, excels in understanding unlabelled speech data. Its architecture adeptly captures subtle language nuances, enriching the comprehension of spoken language. This transformative model demonstrates the power of self-supervised learning in unlocking intricate linguistic patterns, standing as a beacon of innovation in NLP. Collaborating with BERT, it facilitates comprehensive video summarization, leveraging advanced NLP techniques to condense video content effectively and efficiently.

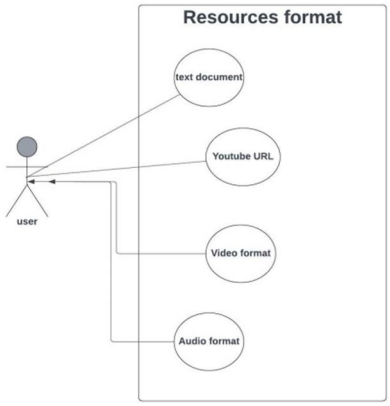


Fig 1 : Choosing the input format

The user can select any type of input format like text document, youtube url, video format and audio format. These type of formats are taken and summarized.

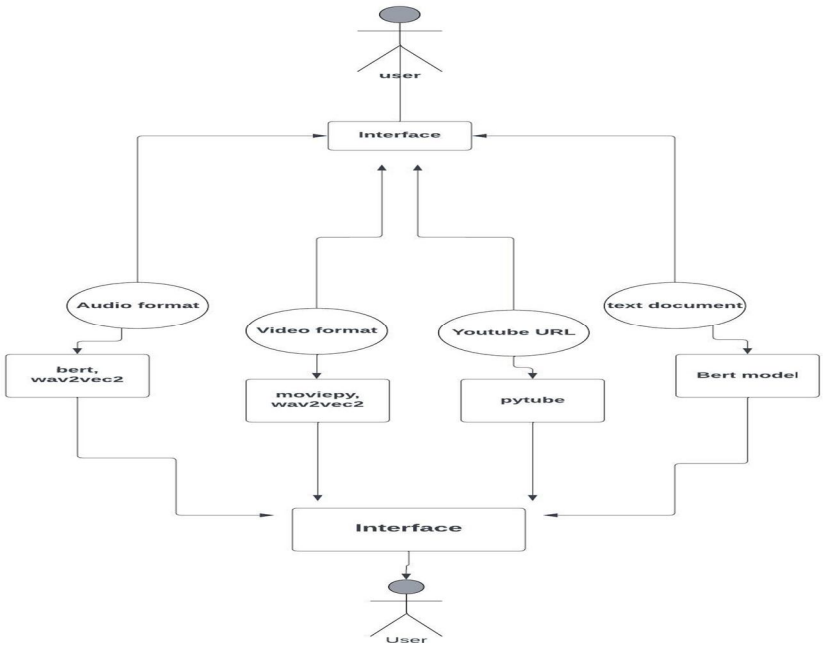


Fig 2 :Flow chart of text and video summarization tool using ml

This is the complete architecture of the model where different types of libraries are used to do summarization. Bert is used to summarize the text. Wav2vec2 is used to extract the captions in video.

IV. RESULTS AND DISCUSSION

The text summarization project starts by collecting a diverse set of documents and preparing them for analysis. Utilizing a machine learning model, such as a Transformer-based architecture, the system learns to extract key content and generate concise summaries. Users interact with the system through a user-friendly application, capturing text input via mobile or PC. The model processes the input and produces abstractive summaries, preserving the essential information from the original text.

The system incorporates checks for specific summarization features, adapting to different document types and styles. It continuously improves through updates and user feedback, ensuring relevance in the face of evolving language patterns.

Ethical considerations are paramount, with the system designed to adhere to legal standards and privacy regulations. Quality control mechanisms review generated summaries to confirm accuracy and maintain high standards of output.

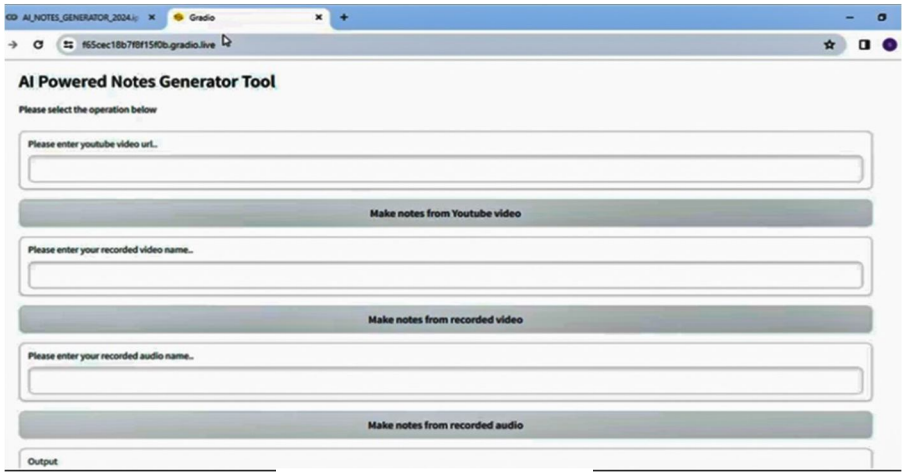


Fig 3 : Unified user Interface

This is the user interface where the user can interact and submit the data in it and do summarization. It is the interface created using gradio.

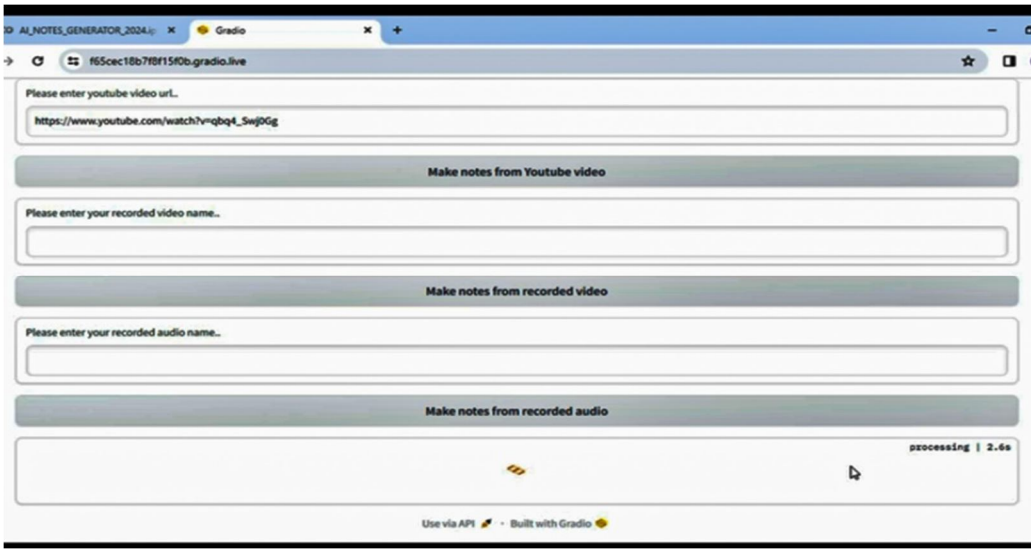


Fig 4: YouTube video summarization

This is where the user can submit the youtube video link by pasting in the box and by clicking on make notes from youtube video button they can get the summary display in it.

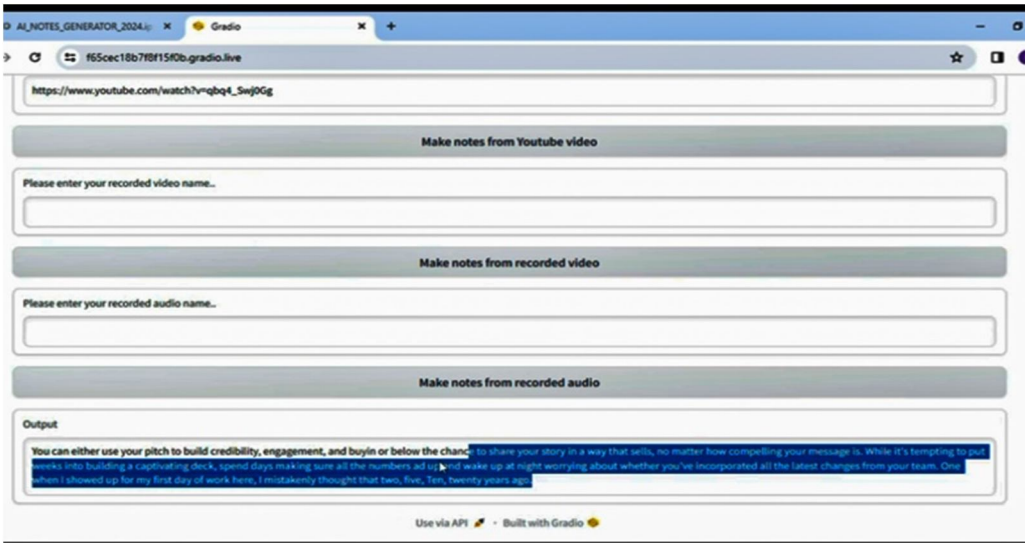


Fig 5 : video summary generated using URL of a YouTube video

This is the display of the summary of the youtube video link provided. In the output box the people can see the summary of the video.

V. CONCLUSION

With the huge amount of data available on the internet, extracting the top data as a concept brief would be useful for some users. Since there is a huge amount of data deployed on the internet, there is always a need to find a way to reduce the length of the text and provide a clear summary.

Summarizing the text is still a work in progress in several research areas and needs to be further researched and developed in the summary of the texts. Due to the huge increase in the amount of data available online, extracting the top information as a conceptual summary will be useful for some researchers.

In this paper we have drawn the reader's attention to the latest and main data problems as well as the need to summarize the texts and explained how short texts can be useful while keeping the original texts intact. The optimization way can also be used to solve different problems.

REFERENCES

[1] Balaji, N. & N, Megha & Kumari, Deepa & P, Sunil & Bhavatarini, N. & A, Shikah. (2022). Text Summarization using NLP Technique. 30-35. 10.1109 /DISCOVER 55800. 2022.9974823.

[2] Bedi, Parminder & Bala, Manju & Sharma, Kapil. (2023). Extractive text summarization for biomedical transcripts using deep dense LSTM-CNN framework. Expert Systems. 10.1111/exsy.13490.

[3] Onah, Daniel & Pang, Elaine & El-Haj, Mo. (2022). A Data-driven Latent Semantic Analysis for Automatic Text Summarization using LDA Topic Modelling. 10.1109/BigData55660.2022.10020259.

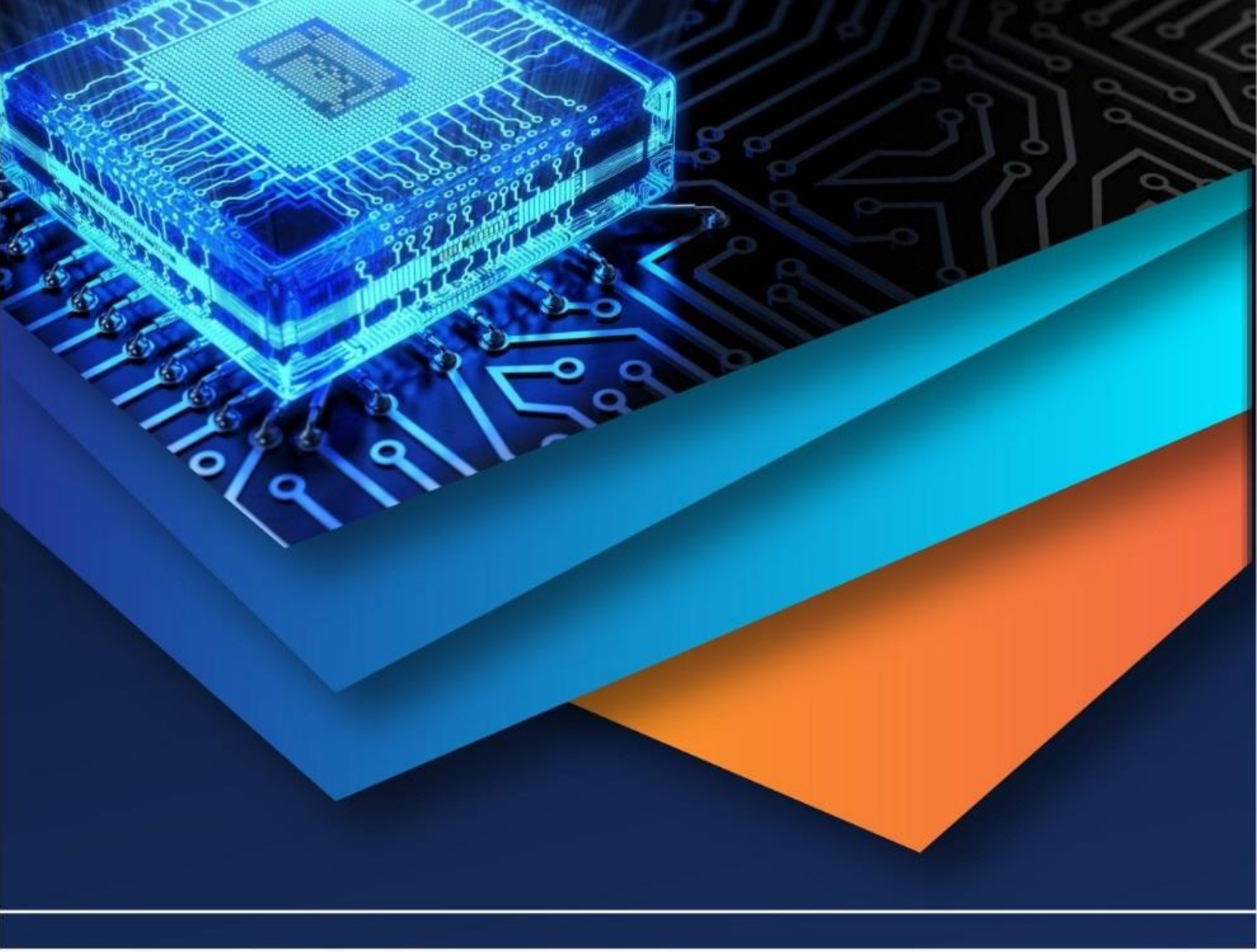
[4] Shambharkar, Prashant & Goel, Ruchi. (2023). From video summarization to real time video summarization in smart cities and beyond: A survey. Frontiers in Big Data. 5. 1106776. 10.3389/fdata.2022.1106776.

[5] Hendi, Sajjad & Hussein, Karim & Taher, Hazeem. (2023). Digital Video Summarization: A Survey. International Journal of Innovative Computing. 13. 67-71. 10.11113/ijic.v13n1-2.421.

[6] Saini, Parul & Kumar, Krishan & Kashid, Shamal & Saini, Ashray & Negi, Alok. (2023). Video summarization using deep learning techniques: a detailed analysis and investigation. Artificial Intelligence Review. 56. 1-39. 10.1007/s10462-023-10444-0.

[7] K. Prudhvi, A. B. Chowdary, P. S. R. Reddy, and P. L. Prasanna, "Text summarization using natural language processing," in Intelligent System Design. Springer, 2021, pp. 535–547.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)