ABSTRACTIVE TRANSCRIPT SUMMARIZATION OF YOUTUBE VIDEOS

[1] G. KIRAN KUMAR
Assistant Professor
Computer Science & Engineering
Anurag University
garakirankumar512@gmail.com

[2] P. SRI CHARAN , [3] L. SREE VIDHYA, [4] K. SHREYA REDDY

[2,3,4] UG Student
Computer Science & Engineering
Anurag University

Abstract -There are tons of videos on YouTube, and it's hard to keep up with all the information in them. Watching each and every video takes a lot of time and we need a quicker way to understand what's in the videos without watching them all. The project aims to create a smart computer program that reads and summarizes what is said in YouTube videos. Advanced models like Seq2Seq RNN-LSTM and PEGASUS are used to make summaries. Here, we are not just using one program but combining a few to make the summaries even better. Additionally, a powerful model BERT is tested to see if it can make the summaries even more efficient. A system is proposed where it quickly summarizes what is in the YouTube video, without having to watch the whole video.

Keywords – Deep Learning, YouTube Video Summarization, Seq2Seq RNN-LSTM, PEGASUS, Ensemble Method, BERT.

I. INTRODUCTION

Since the advent of digital multimedia, a vast array of content has flooded social platforms daily, including news, documentaries, movies, talk shows, and sports events. This abundance of diverse content demands significant processing time and memory. To address this, condensing videos becomes crucial, enabling consumers to obtain maximum information in minimal time. Watching videos is passive, unlike reading, making it challenging to absorb all information actively. Hence, condensing video content becomes essential for effective consumption. Artificial text summarization technologies offer a solution by extracting pertinent details from the video more efficiently. This not only reduces the volume of video data but also facilitates quicker navigation. Users can glance at summaries of YouTube videos to determine if the content aligns with their interests, potentially saving them time.

This study aims to advance abstractive transcript summarization for YouTube videos by employing an ensemble approach, combining Seq2Seq RNN-LSTM, PEGASUS, and BERT for enhanced accuracy and robustness. The investigation builds upon a base paper, extending the use of ensemble methods and introducing BERT to optimize summarization efficacy, ultimately refining the extraction of key information from diverse video sources.

The study addresses the challenge of improving abstractive transcript summarization for YouTube videos. Despite existing methods like Seq2Seq LSTM and PEGASUS, there's a need for enhanced accuracy and robustness. The investigation explores an ensemble approach, incorporating BERT, to optimize performance and extract key information effectively from diverse video sources, aiming to refine summarization in the dynamic realm of YouTube content.

II. PROPOSED ALGORITHM

2.1 Bidirectional Encoder Representations from Transformers (BERT)

BERT (Bidirectional Encoder Representations from Transformers) is a groundbreaking NLP model from Google AI Language that applies bidirectional training to language modeling using a Transformer encoder, allowing it to learn contextual relations between words in a text sequence more effectively than previous approaches. Key innovations include Masked Language Model (MLM), where 15% of words in a sequence are randomly masked during training and the model predicts the original values of those words, teaching it to utilize context from surrounding words; and

Next Sentence Prediction (NSP), where BERT learns if one sentence follows another in the original text by training on sentence pairs. BERT can be fine-tuned for tasks like classification, question answering, and named entity recognition by adding task-specific layers to its output, and while most hyperparameters remain consistent, some require adjustment during fine-tuning. Available in two sizes, BERT_base uses 110 million parameters, and BERT_large uses 345 million, providing superior performance on many tasks due to its larger scale. BERT's bidirectional approach allows it to achieve state-of-the-art results on many NLP tasks, though MLM slows down model convergence, which is offset by its increased context awareness. Overall, BERT represents a major advance in NLP by offering a way to train models that deeply understand language context, enhancing its versatility and practical applications.

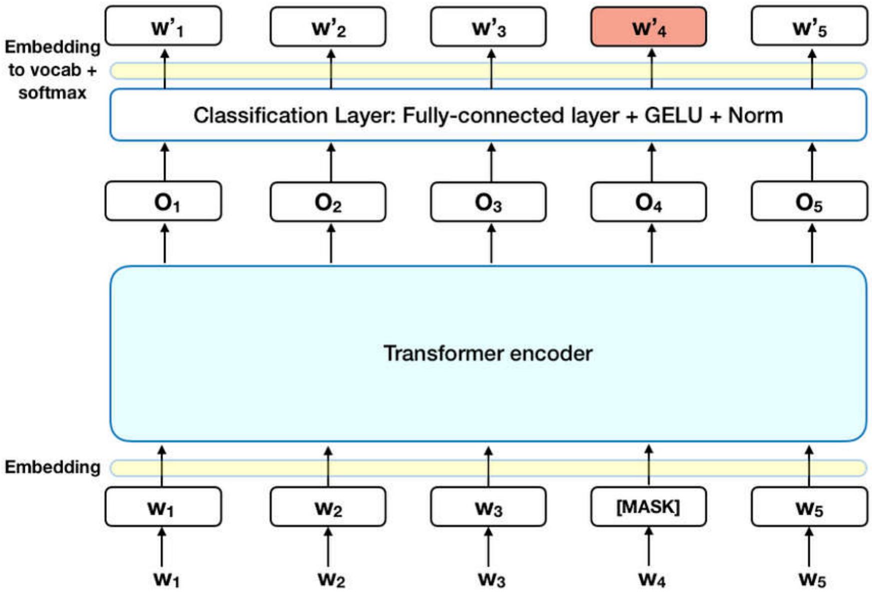

Fig 2.1. Masked LM (MLM)

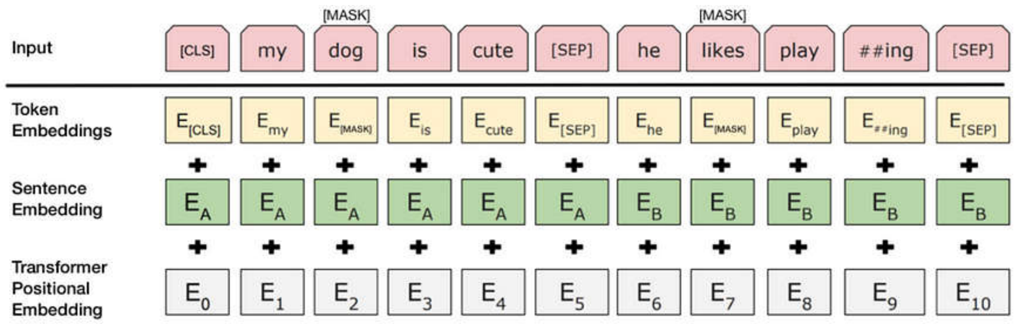

Fig 2.2. Next Sentence Prediction (NSP)

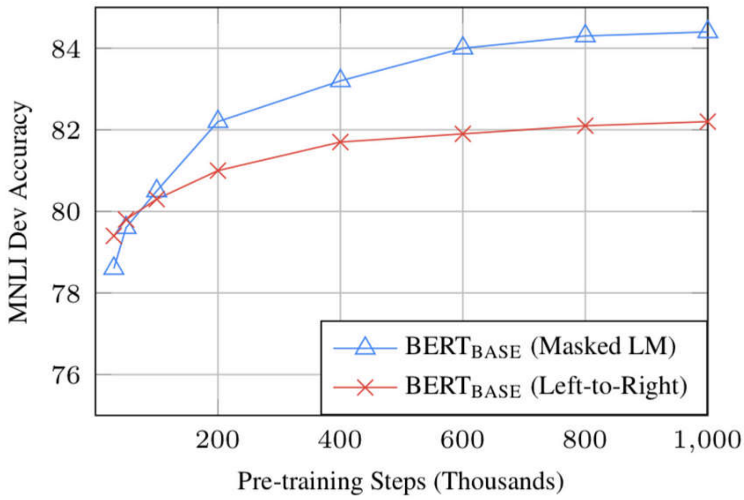

Fig 2.3. BERT Takeaways

	Training Compute + Time	Usage Compute
$BERT_{BASE}$	4 Cloud TPUs, 4 days	1 GPU
$BERT_{LARGE}$	16 Cloud TPUs, 4 days	1 TPU

Fig 2.4. Compute Considerations (Training and Applying)

2.2 PEGASUS

Pegasus is a cutting-edge large-scale transformer-based pre-trained model designed for abstractive text summarization. Developed by Google Research, Pegasus (Pre-training with Extracted Gap-sentences for Abstractive Summarization) leverages a novel pre-training method where it masks and reconstructs multiple consecutive sentences (gap-sentences) in a document. This approach enables Pegasus to learn robust contextual understanding and generate coherent summaries. During pre-training, the model learns to predict the masked sentences from the rest of the document, enhancing its ability to capture broader context and produce more natural, fluent summaries. Pegasus has demonstrated state-of-the-art performance across a range of summarization tasks and benchmarks, making it a valuable tool in natural language processing for applications such as news summarization, document compression, and other text generation tasks.

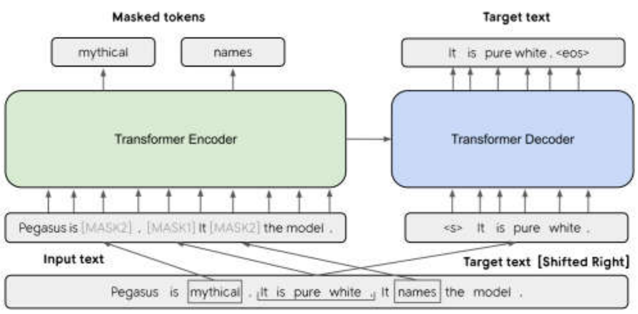

Fig 2.5. Working of PEGASUS

2.3 Sequence 2 Sequence (seq2seq)

Sequence to Sequence (seq2seq) models, utilizing Encoder-Decoder architectures, are pivotal in solving complex natural language processing tasks such as machine translation, question answering, and text summarization. The model consists of an encoder, typically an LSTM or GRU, which processes the input sequence and encodes it into a context vector that captures its information. This vector is then used by the decoder, another LSTM, which generates the output sequence based on the encoded information. Despite their efficacy, seq2seq models face challenges with long sequences due to limited memory and potential vanishing gradients. These limitations prompt exploration of advanced models like Attention and Transformers for more robust solutions. Nonetheless, seq2seq models continue to be integral in NLP, powering applications like Google Translate and voice-enabled devices. Further research aims to enhance their efficiency and applicability across a broader range of tasks.

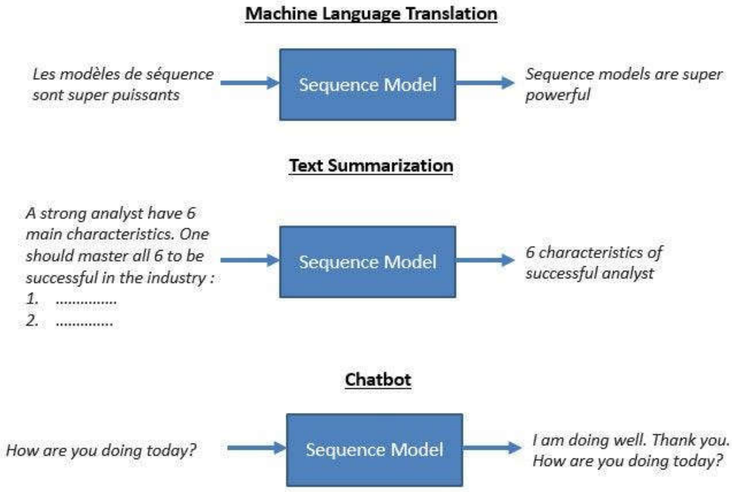

Fig 2.6. Sequence model

III. Experiment and Result

3.1 Experiment findings

1. **Improved Summarization:** The system successfully summarized YouTube video transcripts into concise and understandable summaries.
2. **Enhanced Efficiency:** By using advanced models like Seq2Seq RNN-LSTM and PEGASUS, the system was able to efficiently generate accurate summaries without requiring users to watch the entire video.
3. **Better Performance with Ensemble Method:** Combining multiple summarization methods through the ensemble method resulted in even more accurate and comprehensive summaries.
4. **BERT Integration:** Integrating BERT into the system further improved the quality and speed of summarization, making it even more effective at capturing the main points of the videos.
5. **Ease of Use:** Overall, the system provided a user-friendly solution for quickly understanding the content of YouTube videos, making it easier for users to access relevant information without spending excessive time watching videos.

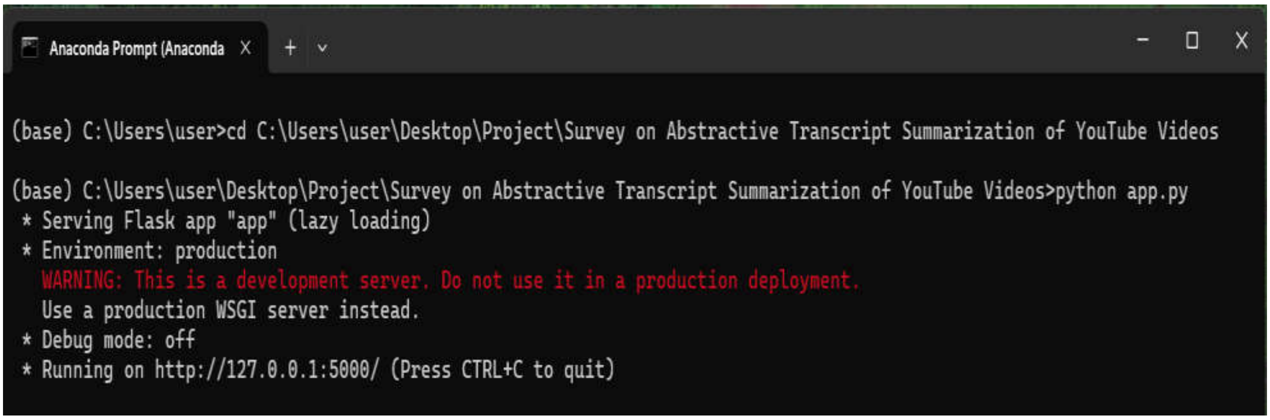

Fig 3.1. Command prompt

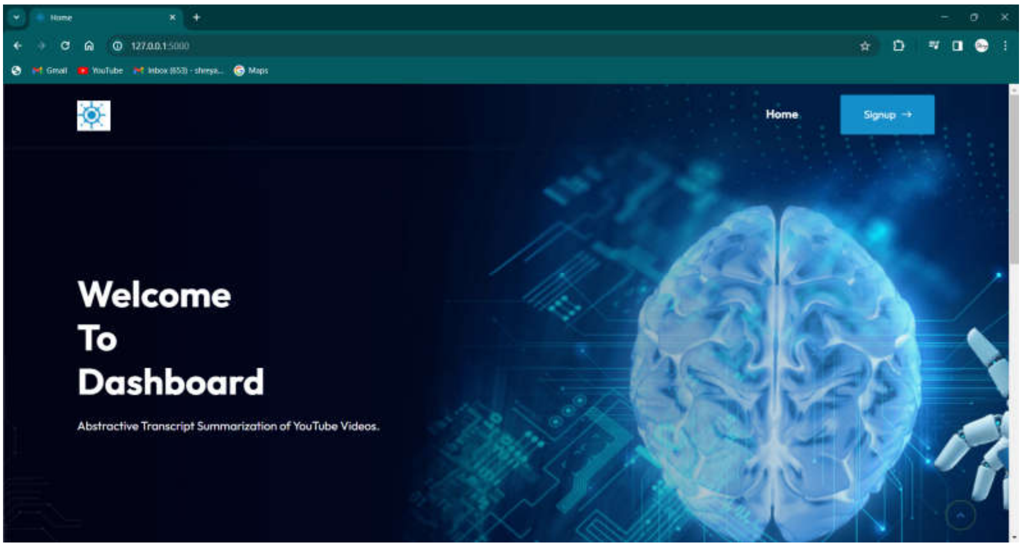

Fig 3.2. Landing page

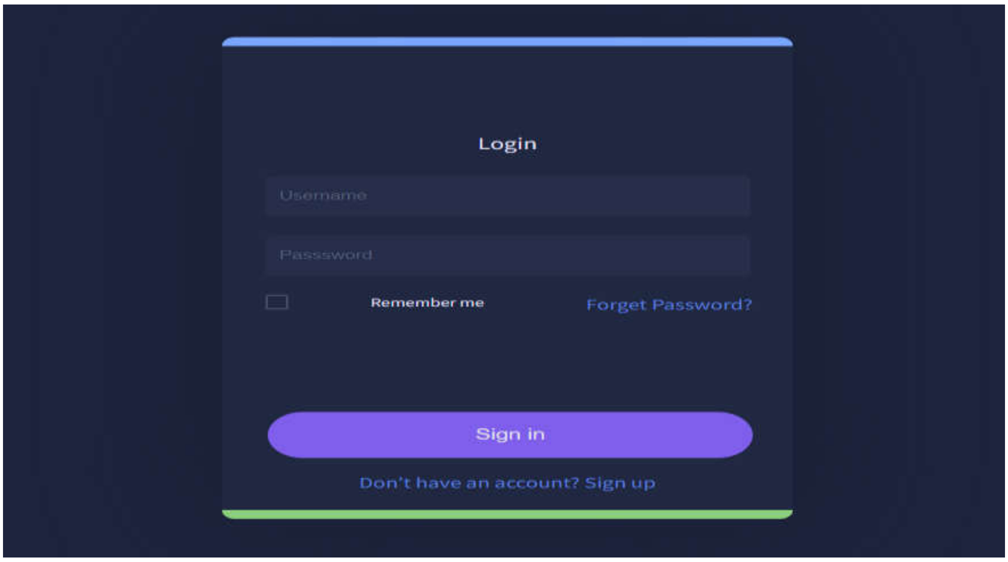

Fig 3.3. Login page

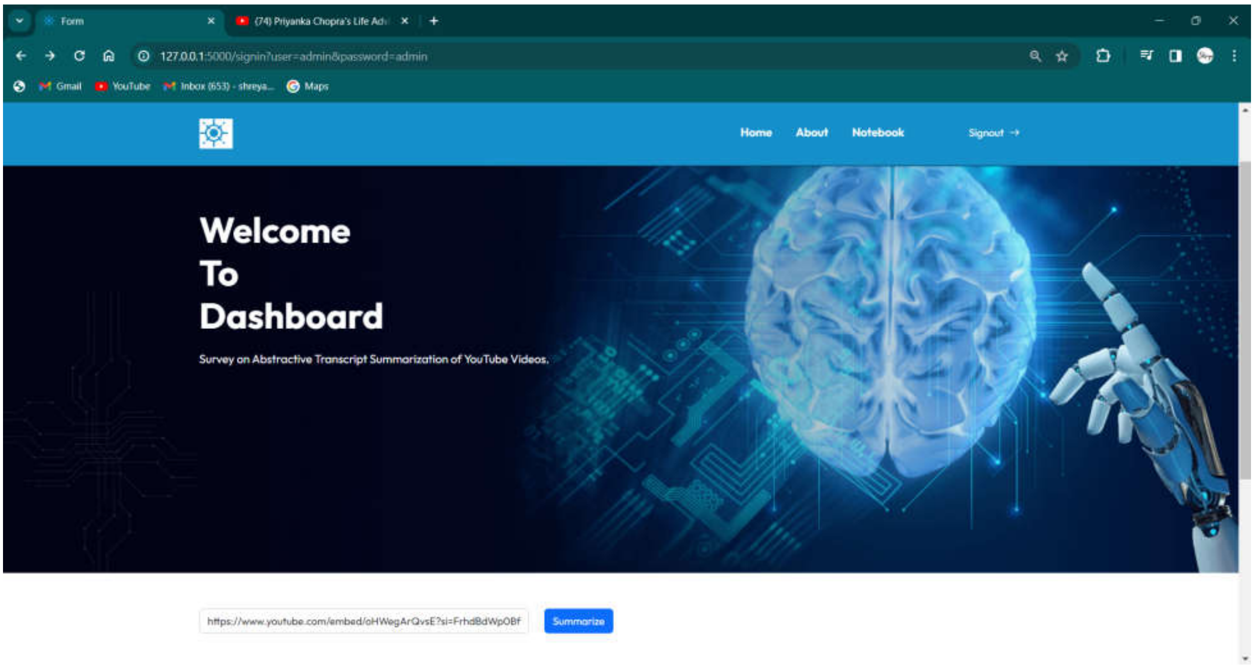

Fig 3.4. Dashboard

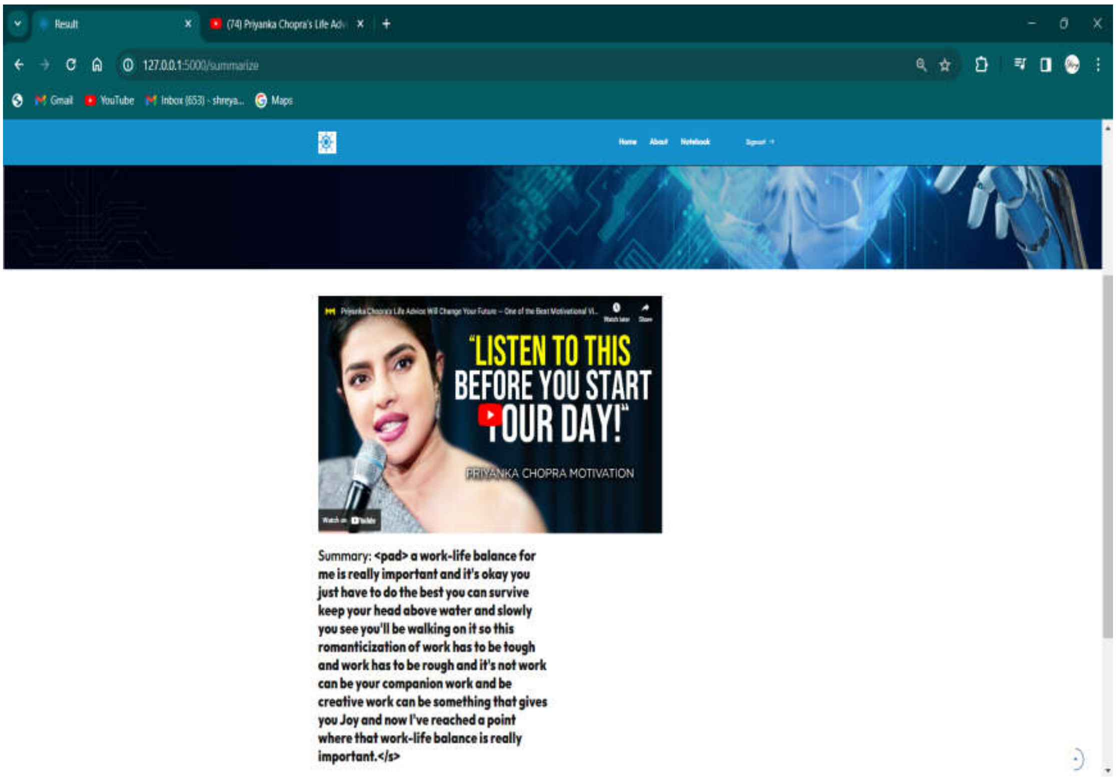

Fig 3.5. Result

IV.CONCLUSION

In conclusion, this study pioneers advancements in abstractive summarization of YouTube videos by synergistically leveraging the strengths of established techniques such as Seq2Seq LSTMs and PEGASUS. The introduction of an ensemble model, amalgamating predictions from these models, enhances accuracy and resilience, marking a notable contribution to the field. Moreover, advocating for the integration of BERT further elevates performance, showcasing a novel approach that represents a significant leap in summarizing YouTube content effectively. The comprehensive nature of the ensemble model holds immense potential to revolutionize video summarization, addressing the dynamic and diverse content landscape on the YouTube platform. This research not only builds upon existing methodologies but also introduces a novel framework that combines their strengths, signaling a promising direction for the future of video summarization. The presented approach opens avenues for improved content understanding and accessibility, catering to the evolving needs of users in the digital era.

REFERENCES

[1] Nayu Liu; Xian Sun; Hongfeng Yu; Fanglong Yao; Guangluan Xu; Kun Fu, "Abstractive Summarization for Video: A Revisit in Multistage Fusion Network With Forget Gate", published in IEEE Transactions on Multimedia (Volume: 25), **DOI:** 10.1109/TMM.2022.3157993.

[2] Surabhi Bandabe, Janhavi Zambre, Pooja Gosavi, Roshni Gupta, Prof. J. A. Gaikwad, “Youtube Transcript Summarizer Using Flask”, published in International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 11 Issue IV Apr 2023- Available at www.ijraset.com

[3] Prof. S. H. Chaflekar, Achal Bahadure, Hosanna Bramhapurikar, Ruchika Satpute, Rutuja Jumde, Sakshi A. Bakhare, Shivani Bhirange, “YouTube Transcript Summarizer using Natural Language Processing”, published in International Journal of Advanced Research in Science, Communication and Technology, DOI:10.48175/ijarsct-3034

[4] Siri Dharmapuri, Sashank Desu, Karthik Alladi, Harika Gummadi, Harshit Gupta, S. Noor, Mohammad Shareef, “An Automated Framework for Summarizing YouTube Videos Using NLP”, published in E3S Web of Conferences, 2023, DOI:10.1051/e3sconf/202343001056

[5] Jitender Kumar, Ritu Vashistha, Roop Lal, Dhrumil Somanir, “YouTube Transcript Summarizer”, published in International Conference on Computing Communication and Networking Technologies, 2023, DOI:10.1109/ICCCNT56998.2023.10308325

[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.

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

[14] 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.

[15]. 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” , arXiv : 2109.05160v1 [cs.CL] 11 Sep 2021

[16]. 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.

[17]. 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.

[18]. 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.

[19]. P. S. Silpa *et al*., "Designing of Augmented Breast Cancer Data using Enhanced Firefly Algorithm," *2022 3rd International Conference on Smart Electronics and Communication (ICOSEC)*, Trichy, India, 2022, pp. 759-767, doi: 10.1109/ICOSEC54921.2022.9951883.

[20]. Mallikarjuna Reddy, A.,Venkata Krishna, V. and Sumalatha, L.” Face recognition approaches: A survey” International Journal of Engineering and Technology (UAE), 4.6 Special Issue 6, volume number 7 , 117-121,2018.
[21]. A. Mallikarjuna Reddy, V. Venkata Krishna, L. Sumalatha," Face recognition based on stable uniform patterns” International Journal of Engineering & Technology, Vol.7 ,No.(2),pp.626-634, 2018,doi: 10.14419/ijet.v7i2.9922 .

[22]. Naik, S., Kamidi, D., Govathoti, S. et al. Efficient diabetic retinopathy detection using convolutional neural network and data augmentation. Soft Comput (2023). https://doi.org/10.1007/s00500-023-08537-7.

[23] V. NavyaSree, Y. Surarchitha, A. M. Reddy, B. Devi Sree, A. Anuhya and H. Jabeen, "Predicting the Risk Factor of Kidney Disease using Meta Classifiers," *2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon)*, Mysuru, India, 2022, pp. 1-6, doi: 10.1109/MysuruCon55714.2022.9972392.

[24] A. Mallikarjuna Reddy, V. Venkata Krishna, L. Sumalatha, “Efficient Face Recognition by Compact Symmetric Elliptical Texture Matrix (CSETM)”, Jour of Adv Research in Dynamical & Control Systems, Vol. 10, 4-Regular Issue, 2018.

[25] Mallikarjuna Reddy, A., Rupa Kinnera, G., Chandrasekhara Reddy, T., Vishnu Murthy, G., et al., (2019), “Generating cancelable fingerprint template using triangular structures”, Journal of Computational and Theoretical Nanoscience, Volume 16, Numbers 5-6, pp. 1951-1955(5), doi: https://doi.org/10.1166/jctn.2019.7830.

[26] Mallikarjuna A. Reddy, Sudheer K. Reddy, Santhosh C.N. Kumar, Srinivasa K. Reddy, “Leveraging bio-maximum inverse rank method for iris and palm recognition”, International Journal of Biometrics, 2022 Vol.14 No.3/4, pp.421 - 438, DOI: 10.1504/IJBM.2022.10048978.