Video Metadata Generation & Classification

[1]Dr. Y. Chitti Babu, [2]Damisetty Bhagya Sri,[3]Madasani Meghana, [4]Immadisetty Mohana Lakshmi

[1]Associate Professor, Dept of Computer Science and Science, St. Ann's College of Engineering & Technology, Chirala-523187, India.

[2,3,4]U. G Student, Dept of Computer Science and Engineering, St. Ann's College of Engineering & Technology, Chirala-523157, India.

ABSTRACT-*The Video Metadata Generation and Classification project is a web-based application developed to automatically generate metadata and classify videos based on their content. With the rapid growth of video content, manually generating metadata has become a time-consuming process. To address this problem, the proposed system automates the process of video analysis using speech recognition and keyword-based classification techniques. Users can upload a video through a Flask-based web interface, where the uploaded video is first previewed. The system then extracts the audio using FFmpeg and converts the speech into text using Google Speech Recognition, which implements Automatic Speech Recognition (ASR). The extracted text is analyzed to generate relevant keywords and classify the video into predefined categories such as Sports, Food, Education, and General using a Keyword-Based Classification Algorithm. The system is developed using Python, Flask, HTML, CSS, FFmpeg, and the SpeechRecognition library. It provides a simple, efficient, and cost-effective solution for automated video organization and retrieval, making it suitable for educational platforms, digital libraries, media organizations, and multimedia content management systems.*

Keywords: Video Metadata Generation, Video Classification, Automatic Speech Recognition (ASR), FFmpeg, Google Speech Recognition, Flask, Keyword-Based Classification, Python.

INTRODUCTION

The rapid growth of digital video content has increased the need for efficient video organization and retrieval. Manual metadata generation is time-consuming and often inconsistent. The proposed Video Metadata Generation and Classification system automates this process by extracting audio using FFmpeg, converting speech into text through Google Speech Recognition (ASR), generating keywords, and classifying videos using a Keyword-Based Classification Algorithm, providing an efficient solution for generating metadata from video.

RELATED WORK

Several researchers have proposed techniques for video metadata generation and classification. Early approaches relied on manually assigning metadata such as titles, descriptions, and categories, which was time-consuming and often inconsistent. And another research may rely on using deep learning techniques and some use machine learning algorithms which becomes expensive, and in efficient to do. However, these approaches often require large datasets, high computational resources, and complex model training.

LITERATURE SURVEY

Previous studies have explored different approaches for this project one of the research papers is 'A Deep Learning Approach for Video Metadata Generation'. It is based on Deep learning approach which extracts audio and video using trained model. But the major disadvantage is it requiring large datasets and high computational resources. Another research Paper 'Metadata Generation Process for Video Action Detection'. It detects metadata by detecting human actions in video sequences, but it doesn't give accurate data. The Another research paper is 'Video Categorization using Feature and Context Extraction'. It classifies videos based on visual extraction and contextual information. However, this method involved complex feature extraction techniques.

EXISTING SYSTEM

The Existing System is based on Deep Learning techniques to automatically generate metadata and classify video content by analysing both visual and audio information. This system extract metadata such as description, keywords and video categories, by training the model with large data sets. Training the model with complex patterns improves accuracy of video classification. Although it has high accuracy but it has also several limitations like the deep learning model requires large datasets, training the model is more time-consuming process. The performance may decrease when processing videos outside the trained categories. It becomes expensive due to deep learning model training and feature extraction.

PROPOSED METHOD

The proposed Video Metadata Generation and Classification provides an automated approach for analysing video content based on its spoken content. This system is developed by using Flask Based wed interface, it extracts audio from video using FFmpeg and the extracted audio is converted into text using Google Speech Recognition, the video is categorised based

on Keyword-Based Classification Algorithm.

The Proposed System offers several advantages, including reduced manual work, improved efficiency and the cost is also effective when compared to other systems. This system is easy to use and User friendly. Unlike training a model it can extract audio using FFmpeg and classify videos using Keyword-Based Classification algorithm.

SYSTEM ARCHITECTURE

This system has Five Modules first module is Video Upload Module, where user upload video then the system displays a video preview. The second module is Audio Extraction Module uses FFmpeg. This module separates the audio from video. Then the extracted audio processed in the Speech Recognition module. This module uses Google Speech Recognition to converts the extracted audio into text by recognizing the spoken words in the video. The extracted text is processed by the Keyword-Based Classification Module. In this module, the recognised text is compared with keywords to different categories. Finally, the result Module displays the result to the user.

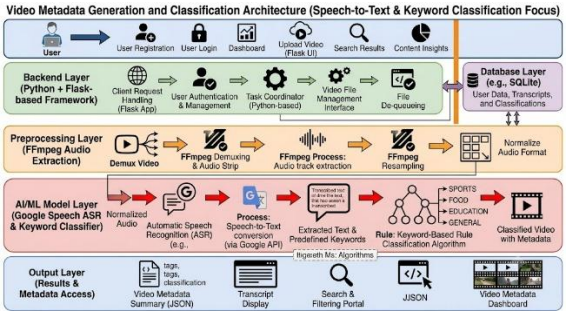

Fig-1 System Architecture

METHODOLOGY DESCRIPTION

The proposed methodology begins with video upload through a Flask-based web interface, followed by video preview and audio extraction using FFmpeg. The extracted audio is converted into text using Google Speech Recognition and Automatic Speech Recognition (ASR). The text is then pre-processed to identify meaningful words and generate relevant keywords. A Keyword-Based Classification Algorithm analyses the extracted content and classifies the video into predefined categories. Finally, the generated metadata, keywords, category, and extracted text are displayed to the user.

RESULTS & DIMENSIONS

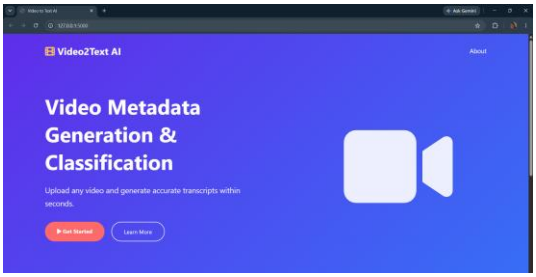

Fig-2 Home Page

This is the home page of this system, in this page we can click on get started or learn more they can redirect to respective pages.

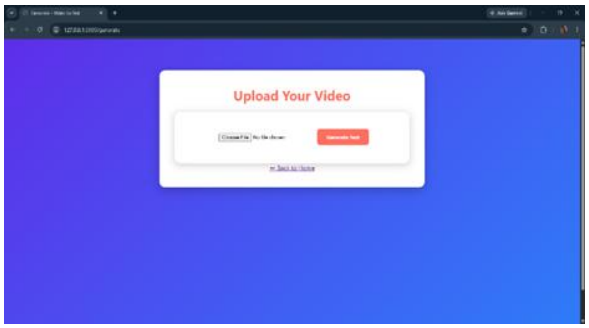

Fig-3 Generate Page

In this page user can upload a video to extract text from the video by clicking the generate text button.

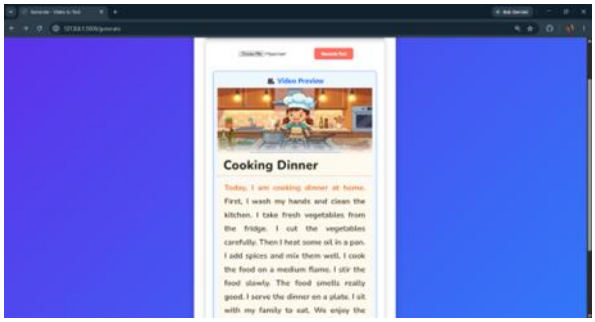

Fig-4 Video Preview

In this window the uploaded video can be previewed by the user.

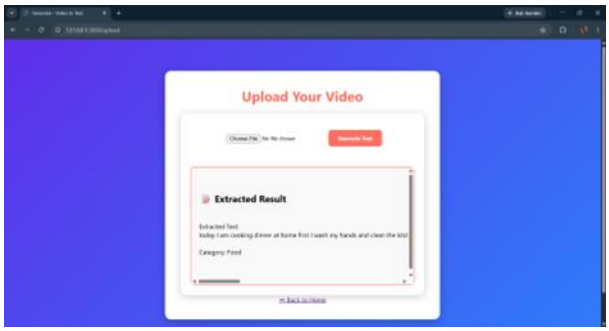

Fig-5 Result Page

The Result Page displays the extracted text from the video and the category of the video.

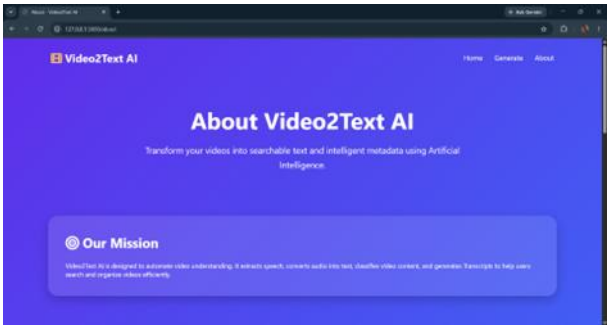

Fig-6 About Page

The above two images describe about the About page of this project it contains the detailed explanation of this project.

CONCLUSION

The Video Metadata Generation and Classification system has been designed successfully to automate the process of analysing video content. The system provides a simple and user-friendly web interface. The Proposed system utilizes FFmpeg, Google speech Recognition, Keyword Based Classification algorithms. Overall, the proposed work makes an effective solution for automatically generate data.

FUTURE SCOPE

The Future improvements may include the integration of Artificial Intelligence (AI)

and Natural Language Processing (NLP) Techniques. The system can be extended to support multilingual speech recognition, Adding features like classification of videos in multiple languages.

REFERENCES

[1] D. P. Harini, "Electronic Health Record Stage Identification using Principal Component Analysis," *Machine Intelligence Research,* vol. 18, no. 01, pp. 864- 873, 17 8 2024.

[2] D. P. Harini, "Two Level Intrusion Detection For Detecting Intruders in Multitier Web Applications," *International Journal of Engineering & Science Research,* vol. 3, no. 9, pp. 472-478, 2013.

[3] D. P. Harini, "Analyze and Predict of Human Cyber Attackers using Artificial Neural Network," *Journal of Basic Science and Engineering,* vol. 21, no. 1, pp. 1529 - 1536, 8 7 2024.

[4] R. Ahmad, O. Dalvi, M. Gupta, D. Totre, and B. Sonawane, "Video Categorization Using Feature and Context Extraction," in Proc. IEEE International Conference, 2023. [Online]. Available: https://ieeexplore.ieee.org/document/10392421

[5] N. Agarwal, R. Gupta, S. K. Singh, and V. Saxena, "Metadata Based Multi-Labelling of YouTube Videos," in Proc. IEEE International Conference, 2017. [Online]. Available: https://ieeexplore.ieee.org/document/7943219

[6] B. Vishwakarma, "Video Classification and Metadata Generation," 2023. [Online]. Available: https://www.scribd.com/document/703777755/Video-Metadata-Generation-and-Classification-1

[7] M. Liu, "Video Classification Technology Based on Deep Learning," in Proc. IEEE International Conference, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9258519

[8] Z. Wu, T. Yao, Y. Fu, and Y.-G. Jiang, "Deep Learning for Video Classification and Captioning," arXiv preprint arXiv:1609.06782, 2016. [Online]. Available: https://arxiv.org/abs/1609.06782

[9] R. Shetty and J. Laaksonen, "Video Captioning with Recurrent Networks Based on Frame- and Video-Level Features and Visual Content Classification," arXiv preprint arXiv:1512.02949, 2015. [Online]. Available: https://arxiv.org/abs/1512.02949

[10] K. Schulte and M. Mes, "Revolutionizing the Danish Broadcasting Industry: A Comprehensive Study of the Influence of AI on Video Metadata Generation," Copenhagen Business School, 2024. [Online]. Available: https://research.cbs.dk/en/studentProjects/r

evolutionizing-the-danish-broadcasting-industry-a-comprehensive-

[11] J. Scherer, A. Scherp, and D. Bhowmik, "Event and Entity Extraction from Generated Video Captions," arXiv preprint arXiv:2211.02982, 2022. [Online]. Available: https://arxiv.org/abs/2211.02982

[12] R. Pasunuru and M. Bansal, "Multi-Task Video Captioning with Video and Entailment Generation," in Proc. 55th Annual Meeting of the Association for Computational Linguistics (ACL), 2017. [Online]. Available: https://aclanthology.org/P17-1117

[13] M. Tian and G. Li, "Rethink Video Retrieval Representation for Video Captioning," Pattern Recognition, vol. 160, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0031320324004953

[14] X. Ren et al., "From Visual Features to Key Concepts: A Dynamic and Static Concept-Driven Approach for Video Captioning," Pattern Recognition Letters, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0167865525001394

[15] N. Aafaq et al., "Spatio-Temporal Dynamics and Semantic Attribute Enriched Visual Encoding for Video Captioning," arXiv preprint arXiv:1902.10322, 2019. [Online]. Available: https://arxiv.org/abs/1902.10322

[16] B. Günsel, A. M. Ferman, and A. M. Tekalp, "Video Indexing Through Integration of Syntactic and Semantic Features," in Proc. IEEE International Conference, 1998. [Online]. Available: https://research.itu.edu.tr/en/publications/video-indexing-through-integration-of-syntactic-and-semantic-feat

[17] G. M. Awad, C. Snoek, A. Smeaton, and G. Quénot, "TRECVid Semantic Indexing of Video: A 6-Year Retrospective," National Institute of Standards and Technology (NIST), 2016. [Online]. Available: https://www.nist.gov/publications/trecvid-semantic-indexing-video-6-year-retrospective

[18] J. Jang, "CoreCaption: Core Caption Based Text-to-Video Retrieval," in Proc. IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2026. [Online]. Available: https://openaccess.thecvf.com/content/WACV2026/html/Jang_CoreCaption_Core_Caption_based_Text-to-Video_Retrieval_WACV_2026_paper.html

[19] X. Pang, L. Gao, and L. Zhang, "A Video Summary Generation Model Based on Hybrid Attention Using Multimodal Features," IEEE Access, 2025. [Online].

Available: https://ieeexplore.ieee.org/document/10902170

[20] Y. Yuan, J.-Y. Shen, and Q.-B. Song, "A New Rule-Based Video Classification Approach," in Proc. IEEE International Conference, 2004. [Online]. Available: https://ieeexplore.ieee.org/document/1264476