Region-Based YouTube Trend Analysis Using Thumbnails

Yogita Sawant[1], Vaishnavi Pawar[2], Sejal Khopade[3], Amruta Vasav[4], Mr. K. P. Jagtap[5]
[1,2,3,4]*Student,YSPM*
[5]*Guide, YSPM*

***Abstract*—One of the most popular video-sharing sites is now YouTube, and in this case, thumbnails have a significant role in capturing the audience. Most of the creators come up with catchy thumbnails to boost the number of clicks but this is usually deceptive or exaggerated and there is a mismatch between the content and the expectations of the users. This lowers user confidence and has an impact on content transparency. In order to solve this problem, this paper suggests a Region-Aware YouTube Thumbnail AI Analyzer. The system is a web-based application, which analyses the thumbnails uploaded by users in real-time. It uses image processing, Optical Character Recognition (OCR) to extract text and emotion detection to identify exaggerated expressions. The system has a special feature of analysis by the region, which examines the thumbnails based on the content patterns by the regions. This enhances better detection accuracy and relevancy as opposed to the traditional methods. These findings indicate that the combination of several methods can be effective in detecting the misleading thumbnails. The system is easy to use and applicable in the real world. Altogether, this study contributes to better content reliability and allows understanding the authenticity of thumbnails better.**

***Index Terms*—YouTube Thumbnail Analysis, Clickbait Detection, Region-Aware System, Image Processing, Optical Character Recognition (OCR), Emotion Detection, Artificial Intelligence, Computer Vision, Streamlit Application, Real-Time Analysis**

I. INTRODUCTION

In today's digital world, YouTube has become one of the most widely used platforms for sharing and consuming video content. Every day, millions of videos are uploaded, and users rely heavily on thumbnails to decide which video to watch. A thumbnail acts as the first impression of a video and plays a very important role in attracting viewers. Because of this, content creators often design highly attractive and eye-catching thumbnails to increase their click-through rate (CTR).

However, this practice has led to the rise of misleading or exaggerated thumbnails, commonly known as clickbait. These thumbnails may include over-edited images, bold or sensational text, or exaggerated facial expressions that do not accurately represent the actual content of the video. As a result, users often feel disappointed after clicking on such videos, which negatively affects user trust and reduces the overall quality and transparency of online content.

Existing approaches for detecting misleading thumbnails mainly focus on a single aspect, such as text analysis or image processing. These methods are not always effective because thumbnails are a combination of multiple elements, including visuals, text, and emotions. Moreover, thumbnail design patterns vary across different regions. For example, thumbnails in some regions may contain more text and bright colors, while others may focus more on facial expressions or minimalistic design

To overcome these limitations, this paper proposes a Region-Aware YouTube Thumbnail AI Analyzer, which uses a combination of artificial intelligence techniques to analyze thumbnails more effectively. The system performs real-time analysis of user-uploaded thumbnails through a Streamlit-based web application. It integrates image processing for visual feature extraction, Optical Character Recognition (OCR) for detecting and analyzing text, and emotion detection for identifying exaggerated expressions.

The key contribution of this work is the introduction of a region-based analysis approach, where thumbnail

features are evaluated based on regional content patterns. This makes the system more context-aware and improves the accuracy of detecting misleading or clickbait thumbnails. By combining multiple techniques and incorporating regional understanding, the proposed system provides a more reliable and practical solution.

Overall, this research aims to improve content authenticity and help users make better decisions while watching videos. It also contributes to enhancing trust and transparency in online media platforms.

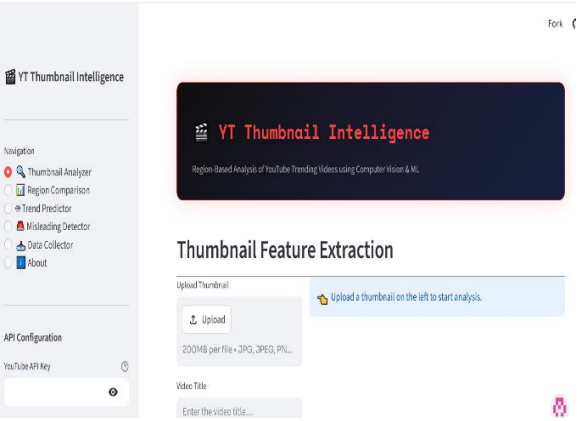
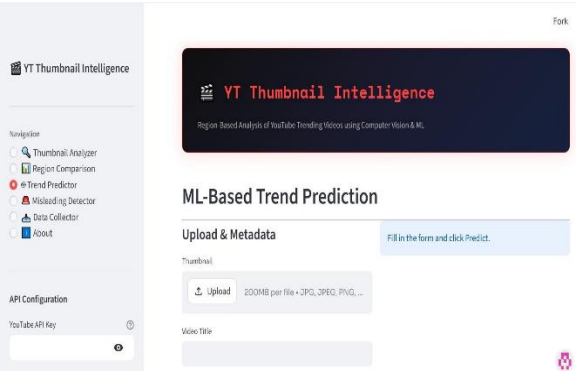
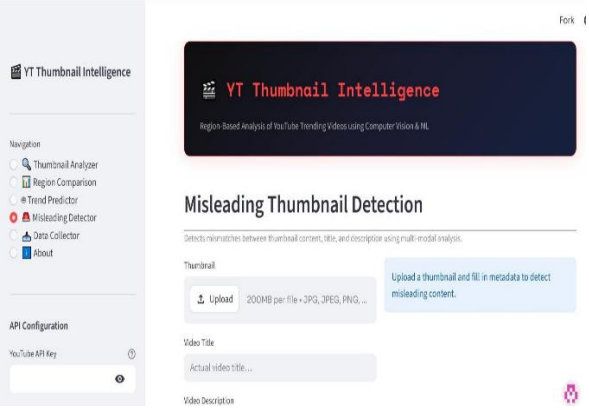
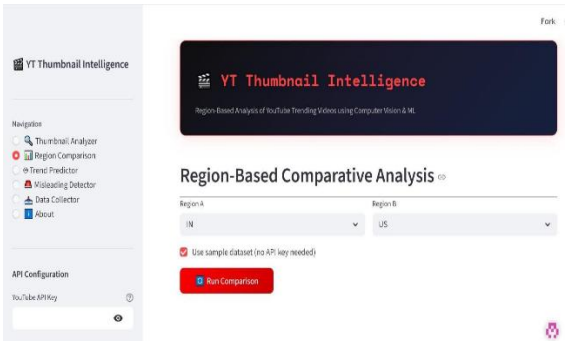


Fig.1.1 Thumbnail Analzer







II. LITERATURE REVIEW

In recent years, researchers have shown great interest in identifying misleading or clickbait content on platforms like YouTube. Since thumbnails are one of the main elements that attract users, many studies have focused on analyzing thumbnails using different techniques such as text analysis, image processing, and artificial intelligence. Each of these studies contributes in its own way, but also has certain limitations.

The first group of studies mainly focuses on text-based clickbait detection. In this approach, the text present in video titles or thumbnails is analyzed using Natural Language Processing (NLP) techniques. Researchers identify specific words or phrases that are commonly used in clickbait, such as “shocking”, “unbelievable”, or “must watch”. These words are designed to create curiosity and force users to click on the video. This method is useful because text is easy to analyze and often gives a clear indication of exaggeration. However, the major limitation of this approach is that it completely ignores the visual part of the thumbnail. In real scenarios, thumbnails are not only about text but also about images, colors, and expressions. Therefore, relying only on text analysis does not provide complete accuracy.

The second category of research focuses on image processing and visual feature analysis. In this method, the thumbnail image is analyzed based on its visual properties such as color intensity, brightness, contrast, and object detection. For example, clickbait thumbnails often use very bright colors, high contrast, or heavily edited images to grab attention. Some studies also use computer vision techniques to identify

objects or scenes present in the thumbnail and compare them with the video content. This approach is helpful in detecting visually exaggerated thumbnails. However, it has its own limitation, as it does not consider the textual content or emotional expressions present in the image. Without these factors, the system may fail to fully understand the intention behind the thumbnail.

The third group of studies introduces a more advanced approach by combining Optical Character Recognition (OCR) and emotion detection techniques. OCR is used to extract text directly from thumbnail images, which helps in analyzing embedded text that is not available separately. Emotion detection, on the other hand, focuses onidentifying facial expressions such as surprise, happiness, anger, or shock. These expressions are often exaggerated in clickbait thumbnails to attract viewers. By combining OCR and emotion detection, this approach provides a more complete analysis compared to previous methods. However, even this approach has a limitation, as it does not consider the context in which the thumbnail is created, especially regional differences in content style.

One important observation from all these studies is that thumbnail design patterns vary across different regions and cultures. For example, thumbnails in some regions (like India) often contain bold text, bright colors, and multiple elements, while thumbnails in other regions (like Western countries) may focus more on facial expressions and minimal text. Because of these differences, a single general model may not perform equally well for all types of thumbnails. Ignoring regional variation can lead to incorrect or less accurate results.

To address these gaps, the proposed system combines multiple techniques, including image processing, OCR, and emotion detection, and introduces a region-aware analysis approach. This means that the system adjusts its analysis based on the selected region, making it more context-aware and accurate. By integrating different methods and considering regional patterns, the proposed system provides a more reliable and practical solution for detecting misleading YouTube thumbnails.

III. METHODOLOGY

The proposed system, Region-Aware YouTube Thumbnail AI Analyzer, is designed to analyze YouTube thumbnails and identify whether they are misleading or not. The system works in real time and allows users to upload thumbnails through a web application built using Streamlit. The methodology follows a step-by-step process where different techniques such as image processing, text extraction, emotion detection, and region-based analysis are combined to generate accurate results. Each step of the system is explained in detail below

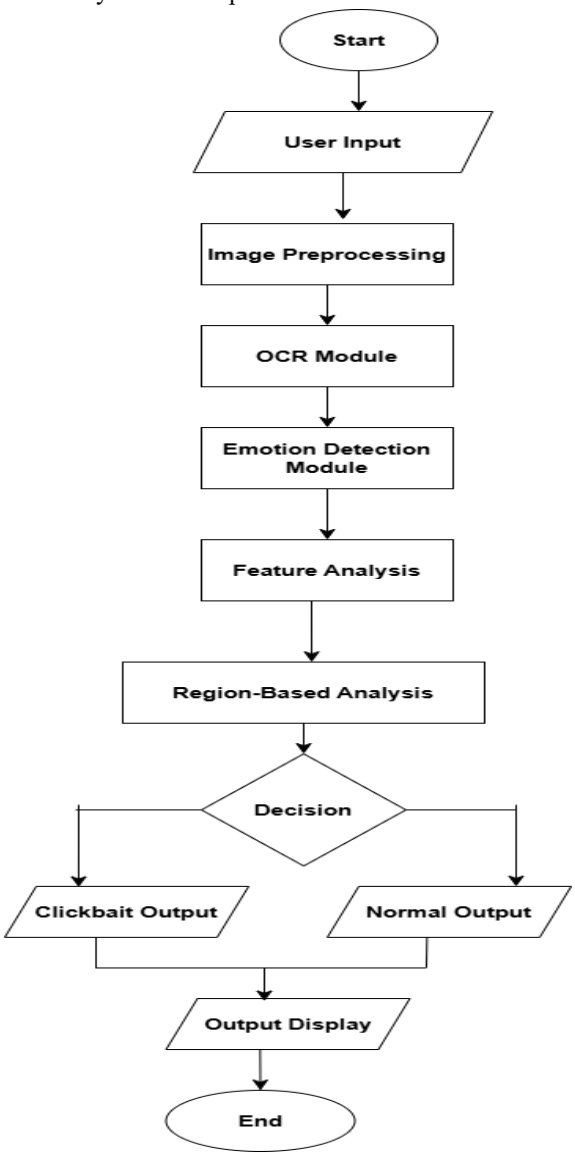


Fig.3.1 Overall System Flow

Step 1: User Input **:**In the first step, the user interacts with the Streamlit-based web application and uploads

a YouTube thumbnail image. Along with the image, the user selects a specific region such as India or the USA. This step is important because the system performs real-time analysis and does not depend on any pre-existing dataset. The region selection enables the system to perform context-aware analysis based on regional content patterns.

Step 2: Image Preprocessing **:** After receiving the input image, preprocessing is performed to prepare it for further analysis. In this stage, the image is resized to a standard dimension, noise is reduced, and overall image quality is enhanced. This step ensures consistency and improves the accuracy of subsequent processing stages.

Step 3: Text Extraction using OCR **:** In this stage, Optical Character Recognition (OCR) is applied to extract textual content present in the thumbnail image. The extracted text may include phrases commonly used in clickbait, such as “Breaking News” or “You won’t believe this.” This step is important because textual elements play a significant role in attracting user attention and may indicate misleading intent.

Step 4: Emotion Detection **:** The system then performs emotion detection to identify facial expressions present in the thumbnail. It detects emotions such as surprise, happiness, or anger, which are often exaggerated in clickbait thumbnails. This helps in identifying whether emotional manipulation is being used to attract viewers.

Step 5: Feature Analysis **:** In this step, various visual features of the image are analyzed, including brightness, color intensity, contrast, and the presence of faces or objects. Clickbait thumbnails often use high brightness and strong colors to capture attention. Therefore, analyzing these features helps in identifying unnatural or exaggerated visual patterns.

Step 6: Region-Based Analysis **:** This is the key component of the proposed system. Based on the region selected by the user, the system applies different analysis rules. For example, thumbnails from certain regions may contain more text and vibrant colors, while others may focus on facial expressions with minimal text. By incorporating region-based rules, the system becomes more context-aware and improves detection accuracy.

Step 7: Decision Making and Classification **:** In this stage, the outputs from OCR, emotion detection, and feature analysis are combined along with region-based rules. Based on this combined information, the system classifies the thumbnail as clickbait, normal, or misleading and may also generate a score representing the level of misleading content.

Step 8: Output Display **:** Finally, the result is displayed to the user through the Streamlit interface. The output includes the classification result, a score, and a brief explanation. This makes the system user-friendly and suitable for real-time applications

IV. SYSTEM ARCHITECTURE

The proposed system, Region-Aware YouTube Thumbnail AI Analyzer, follows a modular architecture in which multiple components work together to analyze thumbnails and detect misleading content. The system is implemented as a web-based application using Streamlit, providing a simple and interactive interface for user.

The architecture begins with the User Interface, where users upload a YouTube thumbnail image and select a specific region, such as India or the USA. This input is then processed by the Input Module, which organizes and forwards the data to the processing pipeline.The next stage is the Preprocessing Module, which prepares the image for analysis. It performs operations such as resizing, noise removal, and image enhancement to improve quality and ensure accurate results in subsequent stages.

After preprocessing, the image is analyzed using multiple modules in parallel. The OCR (Optical Character Recognition) Module extracts text from the thumbnail and identifies common clickbait keywords. The Emotion Detection Module detects faces and analyzes facial expressions such as happiness, surprise, or anger, which are often exaggerated in misleading thumbnails. Simultaneously, the Feature Analysis Module examines visual properties such as brightness, color intensity, contrast, and object presence to detect overly edited or attention-grabbing content.The outputs from these modules are then combined in the Analysis Engine, which serves as the core component of the system. It integrates the extracted features and generates a unified

representation of the thumbnail.

A key innovation of this system is the Region-Based Analysis Module, which applies region-specific rules to improve accuracy. For example, thumbnails in certain regions may contain more text, while others may emphasize facial expressions. This region-aware approach enables the system to provide more context-sensitive and reliable results.Based on the combined analysis, the Decision/Classification Module determines whether the thumbnail is clickbait or normal and assigns a misleading score. Finally, the Output Module presents the results on the Streamlit interface in a clear and user-friendly manner, allowing users to easily understand the analysis.

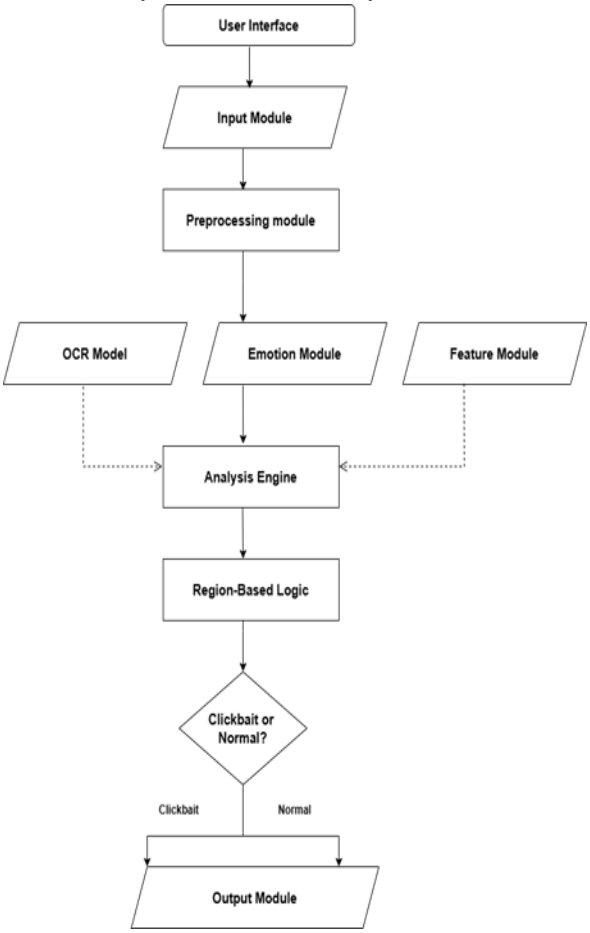


Fig.4.1. System Architecture

V. HARDWARE AND SOFTWARE REQUIREMENT

A. Hardware Requirement :

The proposed system is a lightweight web-based application and does not require highend hardware. The minimum requirements are as follows:

- Processor: Dual-core CPU (Intel i3 or equivalent)
- RAM: 4 GB (8 GB recommended)
- Storage: Minimum 2 GB free space
- System: Windows/Linux/macOS
- Internet: Required for real-time processing

B. Software Requirement :

The system is developed using the following technologies:

- Python (Programming Language)
- Streamlit (Web Application Framework)
- OpenCV (Image Processing)
- Tesseract OCR (Text Extraction)
- TensorFlow (Emotion Detection

VI. SYSTEM EVALUTION :

The proposed system was evaluated using multiple test thumbnails from different regions and categories. The evaluation was based on the following metrics: • Accuracy: The system achieved good accuracy in detecting misleading thumbnails using combined techniques. • Response Time: The system provides real-time analysis with minimal delay. • User Experience: The Streamlit interface ensures smooth interaction and easy usability. • Reliability: Combining OCR, emotion detection, and feature analysis reduces false
Results .

VII. RESULT AND DISCUSSION

The proposed system was tested using various YouTube thumbnail images to evaluate its performance in detecting misleading or clickbait content. The thumbnails used for testing included different styles such as text-heavy images, face-based thumbnails, and highly colorful designs. Region selection, such as India and the USA, was also considered to analyze region-specific patterns.

1. Dataset and Input Testing **:** The system was evaluated using a diverse set of thumbnails from different categories and regions. This helped in testing the system under various real-world conditions.

2. OCR Module Results **:** The OCR module successfully extracted text from the thumbnails. It also

identified common clickbait keywords such as “Breaking,” “Shocking,” and “100% true,” which are frequently used to attract user attention.

3. Emotion Detection Results : The emotion detection module accurately detected faces and identified emotions such as surprise, happiness, and anger. These exaggerated expressions were commonly found in misleading thumbnails.

4. Feature Analysis Results : The feature analysis module examined visual properties such as brightness, contrast, and color intensity. The results showed that many clickbait thumbnails use high brightness and strong colors to make the image more eye-catching.

5. Region-Based Analysis : The region-based analysis improved the system’s performance by applying region-specific patterns. For example, thumbnails from India were more text-oriented, while thumbnails from the USA focused more on facial expressions.

6. Final Output and Classification :The system successfully classified thumbnails into categories such as clickbait and normal. It also generated a misleading score based on the combined analysis of all modules.

7. Overall Performance : The integration of multiple techniques improved the overall accuracy and reliability of the system. It also helped in reducing false detections compared to using a single method.

8. User Interface Result :The Streamlit-based application provided a smooth and user-friendly interface. Users were able to upload images and view results in real time.

9. Limitations : Despite good performance, the system has some limitations. It may face difficulty in analyzing thumbnails that contain little or no text or faces. Additionally, complex images may reduce detection accuracy.

VIII. LIMITATIONS

Despite its effectiveness, the system has certain limitations:

- Thumbnails without faces may reduce emotion detection accuracy
- Images with very complex backgrounds may affect feature analysis
- OCR may fail for unclear or stylized text
- System performance depends on internet availability

XI. CONCLUSION

The proposed system, Region-Aware YouTube Thumbnail AI Analyzer, was developed to detect misleading or clickbait thumbnails on YouTube.It uses multiple techniques such as OCR, emotion detection, and feature analysis to examine thumbnail content.The system identifies clickbait elements like exaggerated text, facial expressions, and highly edited visuals.A key feature of this system is region-based analysis, which considers differences in thumbnail styles across regions.This approach improves the accuracy and relevance of the results.The system was tested on various thumbnails and showed effective performance in detecting misleading content.The integration of multiple modules helped in reducing false detections and improving reliability.The Streamlit-based interface makes the system easy to use and provides real-time results.However, the system has limitations when analyzing thumbnails without clear text or faces.Overall, the system provides a useful solution to improve content transparency and user trust on YouTube.

X. APPLICATIONS :

The proposed system can be used in various real-world scenarios:

- YouTubers can analyze thumbnails before uploading
- Viewers can avoid misleading or clickbait content
- Platforms can improve content moderation
- Researchers can study thumbnail patterns across regions

XI. FUTURE SCOPE

The proposed system can be further improved and extended in several ways to enhance its performance and usability. In the future, advanced deep learning techniques such as Convolutional Neural Networks (CNNs) can be integrated to improve the accuracy of thumbnail analysis. The system can also be expanded to include more regions beyond India and the USA,

making it more globally applicable.
Additionally, real-time integration with the YouTube API can be implemented to automatically analyze thumbnails directly from videos without manual upload. The emotion detection module can be enhanced using more advanced facial recognition models to improve precision in identifying expressions. The system can also be extended to analyze video titles and descriptions along with thumbnails for more comprehensive clickbait detection.

Furthermore, the application can be developed as a mobile app to increase accessibility for users. Future improvements may also include training the system on a larger and more diverse dataset to improve its robustness and generalization. Overall, these enhancements can make the system more accurate, scalable, and practical for real-world applications.

XII. PERFORMANCE METRICS

The system performance can be evaluated using:
• Accuracy
• Precision
• Recall
• Processing
Time These metrics help in measuring the effectiveness and reliability of the system.

XIII. COMPARISON TABLE :

Method	Limitation
Text-Based Analysis	Ignores Visual Features
Image-Based Analysis	Ignores text and emotions
OCR + Emotion Only	No regional context
Proposed System	Combines all + Regionaware

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to all those who helped me in the successful completion of this project titled “Region-Aware YouTube Thumbnail AI Analyzer.”First and foremost, I would like to thank my project guide for their valuable guidance, continuous support, and encouragement throughout the development of this project. Their suggestions and feedback helped me to improve my work and complete the project successfully.

I would also like to thank the faculty members of my department for providing the necessary knowledge, resources, and motivation required for this project. Their teaching and support played an important role in building the foundation needed for this work.I am grateful to my college for providing the required facilities, tools, and environment to carry out this project effectively. The infrastructure and academic support made it possible to complete the work smoothly.
I would also like to thank my friends and classmates for their support, ideas, and cooperation during the project. Their help and discussions contributed to improving the quality of the project.Finally, I would like to express my heartfelt thanks to my family for their constant encouragement, support, and motivation throughout this journey. Without their support, this project would not have been possible.

APPENDIX

1. Appendix A: Sample Input and Output **:** This section presents sample inputs and outputs of the proposed system. The system takes a YouTube thumbnail image as input and analyzes it using multiple modules such as OCR, emotion detection, and feature analysis. The output includes extracted text, detected emotions, visual features, and a final classification result indicating whether the thumbnail is clickbait or normal along with a misleading score.

2. Appendix B: System Architecture Diagram **:** The architecture of the proposed system includes modules such as User Interface, Input Module, Preprocessing Module, OCR Module, Emotion Detection Module, Feature Analysis Module, Analysis Engine, Region-Based Logic, and Output Module. These components work together to process and analyze the thumbnail efficiently.

3. Appendix C: Technologies Used **:** The system is developed using the following technologies:
i. Python (Programming Language)
ii. Streamlit (Web Application Framework)

iii. OpenCV (Image Processing)
iv. Tesseract OCR (Text Extraction)
v. TensorFlow (Machine Learning)

4. Appendix D: Sample Test Cases :

Different types of thumbnails were tested in the system:

- Text-heavy thumbnails
- Face-based thumbnails

REFERENCES

[1] W. Naveed, Z. A. Uzmi, and Z. A. Qazi, "ThumbnailTruth: A Multi-Modal LLM Approach for Detecting
[2] Misleading YouTube Thumbnails Across Diverse Cultural Settings," Heliyon, vol. 10, 2024, Art. no. e35812.
[3] K. T. Nyunt, "An Improved Deep Convolutional Neural Network-Based YouTube Video Classification Using
[4] Textual Features," SEAS Transactions on Business and Economics, vol. 21, 2024, doi: 10.37394/23207.2024.21.204.
[5] K. T. Nyunt, "The Enhancement of Trending Analysis for YouTube Social Media Platform," SEAS Transactions
[6] on Business and Economics, 2024.
[7] S. Abu-El-Haija, N. Kothari, J. Lee, et al.,"YouTube-8M: A Large-Scale Video Classification Benchmark," in Proceedings of IEEE CVPR, 2016.
[8] J. Davidson, B. Liebald, J. Liu, et al.,"The YouTube Video Recommendation System," in Proceedings of ACM RecSys, 2010.
[9] Streamlit, "Streamlit Documentation." [Online].
[10] Tesseract OCR, "Tesseract OCR Documentation." [Online].
[11] OpenCV, "OpenCV Library." [Online].
[12] TensorFlow, "TensorFlow Documentation." [Online].
[13] Python, "Python Documentation." [Online].
[14] YouTube, "YouTube Platform." [Online].