International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

# A Framework for Identifying Clickbait

Adithya A J, Krishna J, Rona Siby, Roshan Babu, Mary Priyanka K S
Dept. of CSE, College of Engineering Kidangoor, Kottayam, Kerala, India aadianeesh@gmail.com, kj@163705@gmail.com, ronasiby03@gmail.com, roshanbabu086@gmail.com, marypriyanka@ce-kgr.org

***Abstract*—The widespread use of clickbait on YouTube, fueled by sensational headlines, eye-catching thumbnails, and mislead- ing descriptions, has become a growing concern due to its deceptive nature and detrimental impact on user experience and content credibility. This manipulation of viewer attention undermines the trustworthiness of digital platforms and has prompted extensive research into automated detection techniques. In response, a variety of machine learning and deep learning approaches have been developed to identify clickbait effec- tively. These include natural language processing (NLP) methods for analyzing textual metadata, convolutional neural networks (CNNs) and vision transformers for detecting visual patterns in thumbnails, and sentiment analysis to evaluate the emotional tone of video titles and descriptions. Advanced transformer models such as BERT and GPT have shown significant success in capturing complex patterns across both text and image data, enhancing the accuracy of classification. Despite these technological advancements, evolving clickbait strategies continue to challenge detection systems. Future research should aim to improve the robustness, scalability, and real-time performance of these frameworks, with an emphasis on multimodal learning and cross-linguistic adaptability to better preserve the integrity and reliability of content-sharing platforms.**

***Index Terms*—Clickbait detection, Machine Learning, Deep Learning, Natural Language Processing, Sentiment Analysis, Image Analysis, Generative AI, YouTube Video Analysis, Transformer Models, Ensemble Learning.**

## I. INTRODUCTION

The widespread use of clickbait, particularly in online platforms such as YouTube, has raised significant concerns regarding user experience quality and content integrity. Clickbait involves misleading or exaggerated headlines, descriptions, and thumbnails designed to attract views, often at the expense of content accuracy. This phenomenon not only misguides users toward irrelevant or false information but also presents challenges in content moderation and data credibility. As clickbait tactics evolve, so too must the methods used to identify and combat them through advanced technological approaches.

Several research efforts have been dedicated to addressing this issue by leveraging machine learning and deep learning techniques. Churi and Patil (2020) propose machine learn- ing algorithms that utilize text-based analysis for clickbait detection, while Shaikh et al. (2021) explore deep learning models through a comparative approach to identify misleading content. Rony et al. (2021) analyze the prevalence of clickbait across various topics and its impact on audience engagement. Additionally, Chakraborty et al. (2021) investigate ensemble learning strategies for detecting clickbait in news media, and Varshney and Vishwakarma (2021) introduce a cognitive evidence-based approach to identifying clickbait in YouTube videos.

Various techniques, including sentiment analysis, image recognition, and metadata analysis, have been employed to detect misleading content. Machine learning models such as Support Vector Machines (SVM), Random Forest, and Extreme Learning Machines (ELM) have been compared for their effectiveness in clickbait detection. Furthermore, ensemble learning has demonstrated potential in enhancing accuracy by combining multiple classifiers. Despite significant advancements, challenges persist, particularly in developing more robust models capable of adapting to the continuously evolving nature of clickbait strategies.

This paper reviews the existing literature on clickbait detection, comparing various approaches and evaluating their effectiveness. By analyzing the strengths and limitations of these techniques, we aim to provide a comprehensive understanding of current clickbait detection research and propose future directions for improving automated detection systems.

## II. LITERATURE SURVEY

This literature survey explores a range of machine learning and deep learning approaches developed to detect and reduce clickbait in digital media. Researchers have analyzed various aspects of clickbait—including its language, structure, and emotional appeal—using different classification techniques to improve detection accuracy and support automated moderation systems.

Chakraborty et.al [1]analyzed the language and structure of clickbait headlines, using machine learning models such as Support Vector Machines (SVMs) alongside neural networks. Their research showed that blending deep learning with traditional machine learning techniques leads to better detection accuracy. They recommend embedding clickbait detection tools directly into media platforms to reduce the circulation of misleading headlines.

In a similar vein, Churi et al. experimented with SVM, Logistic Regression, and Decision Trees for classifying clickbait. Their results showed that Logistic Regression delivered the highest precision (97%). Their work highlights the importance of effective feature engineering in building accurate detection models.

Ghanem et.al [2] investigated how emotions relate to misinformation, introducing an Emotionally-Infused Network (EIN) model that combines emotional signals with linguistic features to enhance clickbait detection. Their findings show that emotions like surprise and fear are closely linked to clickbait content. In a related study, they also analyzed the linguistic characteristics of clickbait and evaluated deep learning models using confusion matrices, emphasizing the importance of balanced datasets and the need for real-time detection systems. Rony et al. [3] examine the widespread use of clickbait in both mainstream and unreliable media by analyzing an extensive dataset of 1.67 million Facebook posts. Their study highlights that clickbait is especially prevalent in entertainment and lifestyle-related content, where it substantially enhances user engagement. However, this surge in interaction comes at the expense of the media's credibility, as such headlines often prioritize sensationalism over accurate reporting.

Sisodia et al. [4]apply ensemble learning techniques, including Random Forest and Gradient Boosting, achieving a detection accuracy of 91.16%. Their research highlights the importance of linguistic features such as punctuation, emotional words, and numerical cues in distinguishing clickbait from non-clickbait.

Mowar et.al [5] extend clickbait detection to YouTube by analyzing video titles with ensemble learning models, recommending their integration into the platform's moderation system to reduce misleading content. Similarly, Ahmad et al. find that ensemble methods like Random Forest and ELM perform well in classification tasks, reinforcing their effectiveness for clickbait detection.

Collectively, these studies demonstrate that machine learning and deep learning techniques—especially ensemble methods and emotion-aware models—significantly enhance clickbait detection accuracy. Future research should prioritize improving real-time detection capabilities, ensuring multilingual adaptability, and addressing ethical concerns surrounding automated content moderation to create more responsible and transparent digital media environments.

## III. METHODOLOGY

### A. *System Architecture*

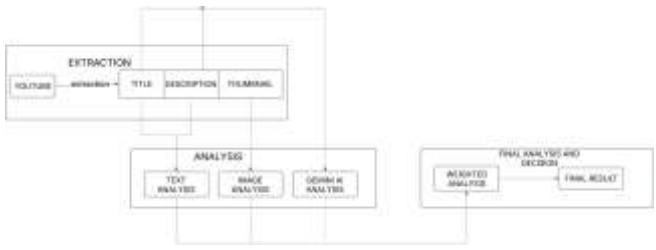


Fig. 1. *System Architecture*

The methodology of this system presents a well- structured, multi-phase framework for developing an AI- powered YouTube video analysis tool. It integrates natural language processing (NLP), image analysis, and generative AI to extract, process, and analyze critical video metadata such as titles, descriptions, and thumbnails. The process is organized into three main phases: design, implementation, and processing.

In the design phase, the system architecture is defined, appropriate machine learning models are selected, and comprehensive data collection strategies are formulated. This phase ensures that NLP is effectively used for analyzing textual elements, image analysis is applied to evaluate thumbnails, and generative AI is incorporated for synthesizing content- based insights.

The implementation phase involves the actual development and integration of these components. NLP algorithms are employed to classify video content and identify linguistic patterns, image analysis models examine thumbnails for their visual and emotional relevance, and generative AI is used to produce comparative insights that enrich the analysis.

The processing phase automates the entire workflow—from metadata extraction to content analysis. It performs sentiment analysis, classifies content types, and generates contextual insights using reasoning mechanisms that enhance interpretability and transparency.

Altogether, this comprehensive methodology facilitates a deep and structured evaluation of YouTube videos, enabling meaningful trend analysis and the generation of insightful, well-reasoned conclusions about the relevance and impact of the content

### B. *Data Extraction*

The system extracts relevant data from YouTube, including:

- **Title**: Used for NLP-based classification.
- **Description**: Supports content analysis.
- **Thumbnail**: Utilized for image-based analysis.

The extraction process guarantees that the collected data is organized and prepared for subsequent processing.

### C. *Data Analysis*

The extracted information undergoes a multi-faceted analysis:

*1) NLP and Sentiment Analysis:*

- The title and description of the content are processed using Natural Language Processing (NLP) techniques.
- Stopwords are removed, followed by feature extraction using methods such as Bag of Words or TF-IDF.
- Sentiment analysis is conducted to evaluate the emotional tone expressed in the content.
- Machine learning models, including SVM, Naive Bayes, and Random Forest, are used to classify the processed text.

*2) Image Analysis:*

- The video thumbnail is analyzed using computer vision techniques.
- Features such as color composition, object recognition, and aesthetic evaluation contribute to classification.

*3) Generative AI Analysis:*

- Google's **gemini-pro** model is used for in-depth content understanding.
- The generative AI examines video-related data to produce insights that go beyond simple classification.

*D. Processing*

After individual analysis components generate their results, a processing step integrates them:

*1) Comparative Analysis:*

- The results from NLP, sentiment analysis, image analysis, and generative AI are then compared and analyzed.
- Weighting techniques are used to determine the signifi- cance of each result.

*2) Final Decision Making:*

- The system synthesizes findings to produce a final clas- sification and reasoning.
- The output offers an explanation for why the video is classified into a specific category.

*E. System Output*

A final report is generated, including:

- Classified category.
- Sentiment and topic breakdown.
- Visual insights from the thumbnail analysis.
- AI-generated explanation supporting the result.

## IV. RESULTS AND DISCUSSION

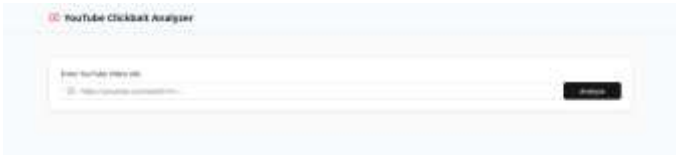


Fig. 2.


Fig. 3.


Fig. 4.

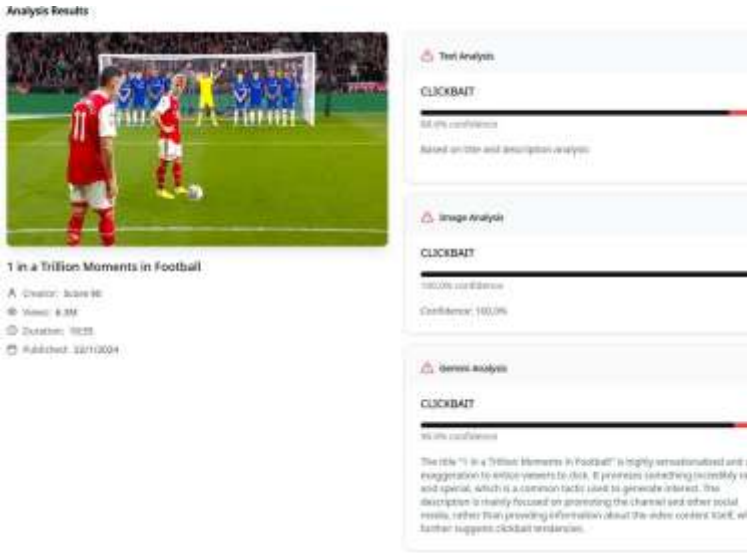


Fig. 5.


Fig. 6.

The **YouTube Clickbait Analyzer** is designed to assess the likelihood of a YouTube video employing clickbait techniques using a multi-modal AI-based approach. The system processes various aspects of a video, including its title, description, and

thumbnail, to determine its classification. The results are generated through a combination of *Natural Language Processing (NLP)*, *Computer Vision*, and *Generative AI analysis*.

### A. Overview of the Analysis

The system evaluates input videos based on three primary components:

- **Text Analysis (NLP and Sentiment Analysis)**: Ex- amines the title and description for linguistic patterns associated with clickbait.
- **Image Analysis (Computer Vision-based Evaluation)**: Assesses the video's thumbnail for exaggerated visual elements.
- **Generative AI Analysis**: Provides an advanced con- textual understanding of the video by evaluating the coherence between textual and visual content.

Each component contributes independently to the classification, and a **weighted decision-making approach** is used to synthesize the final result.

### B. Analysis of a Sample Video

A test case was analyzed using the system, which classi- fied a YouTube video based on the extracted metadata. The breakdown of the results is as follows:

*1) Text-Based Analysis:* The system processed the title and description using NLP techniques. After stopword removal and feature extraction via *TF-IDF (Term Frequency-Inverse Document Frequency)*, classifiers such as **Support Vector Machines (SVM), Na¨ıve Bayes, and Random Forest** were applied. The results showed:

- **Classification:** Not Clickbait
- **Confidence Score:** 85.0%

This suggests the title and description do not exhibit misleading patterns typically associated with clickbait.

*2) Image-Based Analysis:* The thumbnail was analyzed using computer vision techniques to detect:

- High color contrast and saturation.
- Emotionally exaggerated facial expressions.
- Presence of large, attention-grabbing text.

The image analysis classified the video as **clickbait with 70.7% confidence**, indicating strong visual cues related to clickbait.

*3) Generative AI Analysis:* The *gemini-pro* model performed advanced contextual analysis by assessing both textual and visual coherence. It classified the video as clickbait with 90.0% confidence, further supporting the use of attention-grabbing techniques.

### C. Weighted Classification and Decision-Making

The final classification was determined using a weighted approach:

- **Text Analysis (Low Weight)**: 85.0% confidence ⇒ Not Clickbait
- **Image Analysis (Medium Weight)**: 70.7% confidence ⇒ Clickbait
- **Generative AI Analysis (High Weight)**: 90.0% confi- dence ⇒ Clickbait

Since both the image and generative AI models identified the video as clickbait, and their combined influence outweighed the text analysis, the system ultimately classified the video as Clickbait.

### D. Interpretation and System Effectiveness

The results highlight the importance of multi-modal analysis. Relying solely on textual information can lead to misclassification since many clickbait strategies rely on misleading imagery rather than deceptive wording.

### E. Key Strengths of the System

- **Multi-Modal Analysis**: Considers multiple data points rather than just one.
- **Weighted Decision-Making**: Reduces the risk of misclassification from a single feature.
- **Interpretability**: Confidence scores for each module im- prove transparency.

### F. Limitations and Future Enhancements

Despite its accuracy, the system has some limitations:

- **Borderline Cases**: Subtle clickbait techniques can be harder to detect.
- **Weighting Optimization**: Predefined weights may need adaptive learning.
- **Dataset Expansion**:Expanding the dataset to include more categories could enhance the system's ability to generalize.

*1) Proposed Future Improvements:*

- **Fine-tuning Model Weights**: Refining based on real- world feedback.
- **Expanding Training Data**: Including more complex clickbait styles.
- **User Feedback Integration**: Allowing users to contest classifications for better learning.

### G. Conclusion

The **YouTube Clickbait Analyzer** successfully integrates NLP, Computer Vision, and Generative AI to classify clickbait videos. By leveraging a multi-modal approach with weighted decision-making, the system provides improved accuracy compared to traditional methods. Future enhancements will focus on refining classification weights, expanding datasets, and incorporating adaptive learning mechanisms to further improve performance.

## V. C ON C LU S I ON

The reviewed body of research highlights significant advancements in clickbait detection across various digital platforms, driven by the adoption of innovative machine learn- ing, deep learning, and ensemble-based techniques. Methods such as Convolutional Neural Networks (CNNs), Bidirectional Long Short-Term Memory networks (BiLSTMs), and Random Forest classifiers have played a crucial role in achieving

high levels of accuracy when detecting clickbait in both textual and video-based content. The emergence of multi- modal frameworks—capable of integrating textual, visual, and cognitive signals—has further expanded the effectiveness of detection systems, particularly for platforms like YouTube. These frameworks address complex challenges posed by misleading thumbnails, attention-grabbing titles, and even audio elements that are crafted to mislead viewers.

Ensemble learning approaches have gained prominence by combining the strengths of multiple classifiers, allowing systems to overcome the limitations of individual models. In particular, Random Forests have repeatedly outperformed standalone algorithms in identifying subtle and nuanced clickbait patterns, underscoring the importance of combining diverse learning strategies to enhance performance. Emotional analysis has also emerged as a valuable addition to detection models, offering insight into how clickbait exploits psychological triggers such as curiosity, surprise, and fear to manipulate user behavior and increase engagement.

A comprehensive table summarizing the techniques employed in the reviewed studies has been provided, detailing the methodologies, datasets, and performance metrics utilized across various experiments. This comparative analysis offers a consolidated view of the effectiveness and versatility of different approaches, serving as a valuable reference for identifying future research directions. Additionally, the development and deployment of specialized datasets such as BollyBAIT and the Misleading Video Dataset (MVD) have enabled researchers to build more resilient and context-aware models. These datasets play a pivotal role in facilitating both the detection and prevention of clickbait, contributing to the broader goal of promoting transparency and trust in digital media environments

## REFERENCES

[1] A. Chakraborty, B. Paranjape, S. Kakarla, and N. Ganguly, "Stop click- bait: Detecting and preventing clickbaits in online news media," in *2016 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM)*, pp. 9–16, IEEE, 2016.

[2] M. Churi and D. Patil, "Clickbait detection using machine learning algorithm," 2022.

[3] B. Ghanem, P. Rosso, and F. Rangel, "An emotional analysis of false information in social media and news articles," *ACM Transactions on Internet Technology (TOIT)*, vol. 20, no. 2, pp. 1–18, 2020.

[4] D. S. Sisodia, "Ensemble learning approach for clickbait detection using article headline features," *Informing Science*, vol. 22, p. 31, 2019.

[5] P. Mowar, M. Jain, R. Goel, and D. K. Vishwakarma, "Clickbait in youtube prevention, detection and analysis of the bait using ensemble learning," *arXiv preprint arXiv:2112.08611*, 2021.