International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Real Time Sentiment Analysis on Video Streams with Deep learning Models

S. Sai Nadh Reddy[1], S. Sai Nikhil[2], K. Sai Pavan Reddy[3], T. Sai Pradeep[4], K. Sai Pranay Reddy[5], C M. Preeti[6]

[1,2,3,4,5] B. Tech, Malla Reddy University Hyderabad, India
[6]Professor, Malla Reddy University Hyderabad, India
[1]2111cs020439@mallareddyuniversity.ac.in, [2]2111cs020440@mallareddyuniversity.ac.in, [3]2111cs020441@mallareddyuniversity.ac.in, [4]2111cs020442@mallareddyuniversity.ac.in, [5]2111cs020443@mallareddyuniversity.ac.in, [6]preeticm@mallareddyuniversity.ac.in

ABSTRACT

This abstract introduces a novel approach to video sentiment analysis, an underexplored research area where emotions and sentiments are extracted from speakers by processing video frames, audio, and text. The system concurrently analyzes visual features such as facial expressions and mouth motion, and audio features like speech, addressing challenges related to real-time sentiment analysis in systems with constrained software or hardware capabilities. The model, developed to outperform conventional classifiers, offers a unified platform for audio and text processing. By combining statistics from visual, audio, and text components, the proposed system achieves robust and portable emotion detection, providing accurate predictions and adaptability to modern systems with minimal configuration adjustments

1. INTRODUCTION

The surge in children's engagement with social media platforms has spurred concerns about their potential exposure to unsuitable and emotionally distressing content. This highlights the critical need for effective content filtration mechanisms. Our project, "Sentiment Analysis on Video Streams using Lightweight Deep Neural Networks," addresses this pressing issue by leveraging advanced technology to mitigate the risks associated with harmful online content. Through the development of lightweight deep neural networks, our primary objective is to design an efficient system capable of identifying and filtering inappropriate, negative, and upsetting content in real-time video streams.

To achieve this objective, our project encompasses several key goals. Firstly, we aim to develop and train lightweight deep neural network models optimized for real-time video stream analysis, prioritizing computational efficiency without compromising accuracy. Additionally, we seek to create robust algorithms and techniques for sentiment analysis capable of detecting a wide range of emotional expressions and sentiments expressed in video content. Moreover, we plan to implement a scalable system architecture capable of processing large volumes of video streams in parallel, ensuring timely and effective content filtration. Through rigorous testing and validation, we will evaluate the performance of the sentiment analysis system, benchmarking it against existing methods and assessing its efficacy in mitigating the risks associated with harmful content. Furthermore, we aim to facilitate seamless integration of the sentiment analysis system into existing platforms and applications, empowering stakeholders to proactively safeguard children's online experiences.

2. LITERATURE REVIEW

The burgeoning field of sentiment analysis within video content has become pivotal in understanding human emotions and behaviors depicted in multimedia data. This literature review elucidates key research domains that contribute significantly to the advancement of sentiment analysis techniques within video streams. Firstly, scholars emphasize the necessity of comprehensively understanding and harnessing the multi-modality of video data, encapsulating visual, auditory, and textual information, to glean richer insights and bolster the accuracy of sentiment analysis models. Research endeavors aim to integrate and analyze multiple modalities synergistically, thereby enriching the understanding of emotional content conveyed within videos.

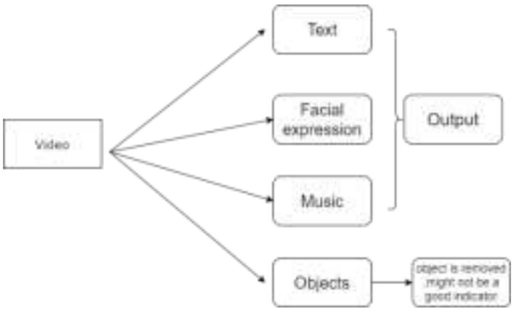
Moreover, facial expression recognition stands as a cornerstone in this discourse, as it serves as a primary indicator of emotions conveyed by individuals within video content. This exploration spans traditional computer vision methodologies to more advanced deep learning-based approaches, all aimed at accurately recognizing and interpreting facial expressions to enhance sentiment analysis outcomes. Additionally, object detection within video streams assumes significance, as researchers delve into methodologies to identify and localize objects within video frames, providing valuable contextual information that influences the overall sentiment conveyed. Text sentiment analysis and audio analysis also emerge as vital dimensions contributing to sentiment analysis within video content, with scholars exploring techniques for extracting sentiment from textual elements and features indicative of emotions from audio signals, respectively. These research endeavors collectively propel the development of advanced sentiment analysis techniques tailored to the multi-modal nature of video data, deepening the understanding of emotional content within multimedia streams and advancing the field of sentiment analysis in video content analysis.

Existing system:

Our research introduces a sophisticated multimedia analysis framework, amalgamating specialized models to comprehensively process and interpret diverse media content. Our system intertwines various components, each addressing distinct facets of multimedia data analysis. Firstly, utilizing the Wav2Vec library, we convert speech data into textual format, facilitating the extraction of textual information from audio inputs. Secondly, music analysis, inclusive of genre classification, mood detection, and lyric sentiment analysis, enriches our understanding of emotional and thematic elements conveyed through music within multimedia content. Thirdly, employing models like Logistic Regression and Multinomial Naive Bayes, we conduct sentiment analysis to extract subjective information accurately from textual data. Furthermore, facial recognition capabilities, powered by Convolutional Neural Networks (CNNs), enable precise identification and analysis of facial features, thereby recognizing individuals and discerning their emotional expressions within multimedia content. Lastly, object detection functionality, implemented through models such as the YOLO (You Only Look Once) model, enhances our system's ability to localize and classify objects within images or video frames. This integration of specialized models within our framework facilitates nuanced analysis across textual, auditory, and visual domains, promising significant advancements in multimedia data interpretation and understanding.

Proposed system:

In our proposed sentiment analysis system, we present a comprehensive framework integrating multiple models to analyze different aspects of textual and visual content. Text analysis will leverage advanced Natural Language Processing (NLP) techniques, employing a Sentiment Analysis model to extract sentiments and opinions expressed within textual content. Facial expression analysis will utilize a Convolutional Neural Network (CNN) Emotion Detection model to accurately interpret facial expressions captured in images or video frames, providing insights into the emotional states of depicted individuals. Music mood analysis will be enhanced through the incorporation of a specialized Music Mood Sentiment Model, enabling the extraction of mood-related features from audio data and classification of emotional tones conveyed through music. Object detection functionality will be facilitated by integrating the YOLO (You Only Look Once) Object Model, enabling efficient localization and classification of objects within visual content. This integration of specialized models within our framework promises significant advancements in sentiment analysis, facilitating comprehensive interpretation across textual and visual domains



3. PROBLEM STATEMENT

The increasing involvement of children in social media platforms has raised significant concerns about their exposure to unsuitable and emotionally distressing content. This escalation underscores the urgent need for effective content filtration mechanisms to safeguard children's digital experiences. Our project aims to address this pressing issue by leveraging lightweight deep neural networks for real-time sentiment analysis on video streams, offering a proactive solution to mitigate the risks associated with harmful online content. Through collaborative efforts and ongoing research, we seek to contribute to the development of robust strategies aimed at protecting children from potentially harmful content encountered online.

The primary objective of our project, titled "Sentiment Analysis on Video Streams using Lightweight Deep Neural Networks," is to develop and implement an efficient system for real-time sentiment analysis on video content, particularly focusing on identifying and filtering inappropriate, negative, and upsetting content. This encompasses designing and training lightweight deep neural network models optimized for real-time analysis of video streams, prioritizing computational efficiency without compromising accuracy. Additionally, our goals include the development of robust algorithms and techniques for sentiment analysis capable of detecting a wide range of emotional expressions and sentiments expressed in video content. We aim to implement a scalable and adaptable system architecture capable of processing large volumes of video streams in parallel, ensuring timely and effective content filtration. Moreover, we will evaluate the performance of the sentiment analysis system through rigorous testing and validation, benchmarking against existing methods, and assessing its efficacy in mitigating the risks associated with harmful content. Our project also seeks to facilitate seamless integration of the sentiment analysis system into existing platforms and applications, enabling stakeholders to proactively safeguard children's online experiences. Additionally, we will provide comprehensive documentation and resources to support the deployment and maintenance of the sentiment analysis system, empowering stakeholders to utilize and customize the solution according to their specific needs**.**

4. METHODOLOGY

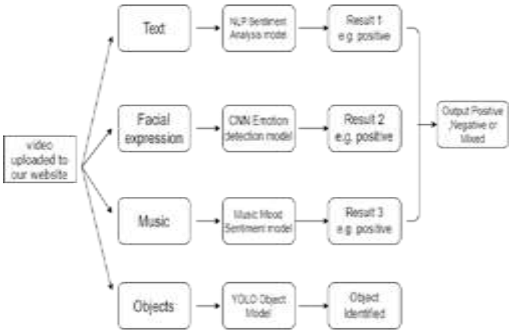
Our methodology encompasses a multifaceted approach, integrating various algorithms and techniques tailored to different modalities for comprehensive sentiment analysis in multimedia data.

Text Analysis: We begin by employing Natural Language Processing (NLP) techniques such as tokenization, stemming, and lemmatization to preprocess textual data. Subsequently, sentiment analysis is conducted utilizing lexicon-based methods, machine learning models such as Logistic Regression or Support Vector Machines, and deep learning architectures like Recurrent Neural Networks (RNNs) or Transformers.

Facial Expression Analysis: Facial expression analysis involves the utilization of Convolutional Neural Networks (CNNs), specifically trained to recognize facial expressions from images or video frames. These CNN models, including architectures like VGG networks or Residual Networks (ResNet), extract features from facial images and classify them into different emotion categories.

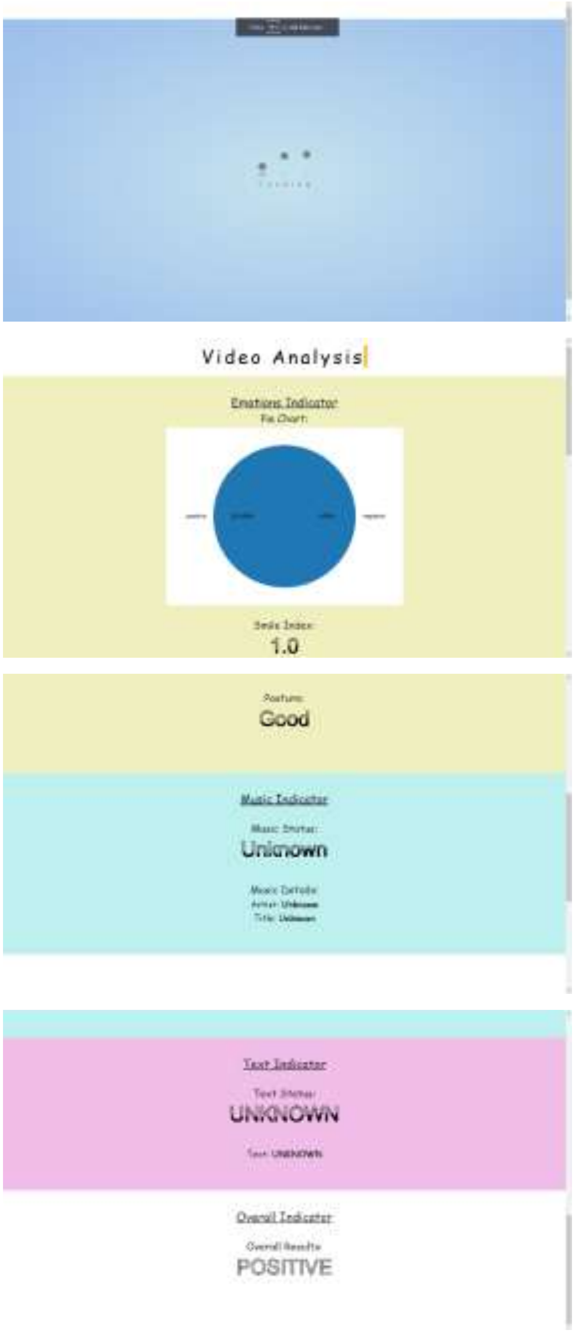
Music Analysis: For music analysis, we employ audio feature extraction techniques such as Mel-Frequency Cepstral Coefficients (MFCCs) to capture relevant characteristics of music. Supervised learning algorithms like Random Forests or Support Vector Machines are trained on annotated music datasets to predict mood labels. Additionally, deep learning architectures like Recurrent Neural Networks (RNNs) or Convolutional Neural Networks (CNNs) are employed to learn complex patterns in music data for mood analysis tasks.

Object Detection: In object detection, deep learning-based techniques are utilized to identify and localize objects within images or video frames. Popular choices include the YOLO (You Only Look Once) Object Detection Model known for its efficiency and accuracy in real-time object detection. Other algorithms and architectures such as Single Shot Multibox Detector (SSD), Faster R-CNN, and RetinaNet are also considered, each offering its own strengths in terms of accuracy and speed.


video uploaded to our website
Text
NLP Sentiment Analysis model
Result 1 e.g. positive
Facial expression
CNN Emotion detection model
Result 2 e.g. positive
Music
Music Mood Sentiment model
Result 3 e.g. positive
Objects
YOLO Object Model
Object Identified
Output Positive ,Negative or Mixed

Overall, our methodology encompasses a blend of lexicon-based methods, deep learning models, feature extraction algorithms, and object detection techniques to effectively analyze text, facial expressions, music, and objects within multimedia data, facilitating comprehensive sentiment analysis across diverse modalities.

5. EXPERIMENTAL RESULTS



Video Analysis

Emotions Indicator

Pie Chart:

Smile Index:

1.0

Posture:

Good

Music Indicator

Music Status:

Unknown

Music Details:

Artist: Unknown

Title: Unknown

Text Indicator

Text Status:

UNKNOWN

Text: UNKNOWN

Overall Indicator

Overall Results

POSITIVE

6. CONCLUSION

In this study, we have achieved significant milestones by completing and demonstrating four pivotal models essential for conducting comprehensive sentiment analysis across multimedia data. Our focus areas include Natural Language Processing (NLP) Sentiment Analysis, Convolutional Neural Network (CNN) Emotion Detection for facial expressions, Music Mood Sentiment Modeling, and YOLO Object Detection. Each model underwent meticulous development and training to effectively analyze different modalities of multimedia content, thereby contributing to a holistic approach towards understanding sentiment expressions.

Moreover, we successfully integrated these models into a unified pipeline, showcased through an additional website interface. This integration not only streamlines the collaboration between individual components but also facilitates the seamless processing of multimedia data. Through rigorous experimentation and validation, we have effectively tackled the identified problem of implementing robust content filtration mechanisms to safeguard digital experiences, especially among vulnerable demographics like children. Our research underscores a commitment to leveraging advanced technologies to mitigate the risks associated with harmful online content and to foster a safer and more enriching digital environment for all users.

7. FUTURE ENHANCEMENT

Enhanced Sarcasm Detection in Text: Building upon the existing Natural Language Processing (NLP) techniques, future enhancements could focus on refining the algorithms to better detect and interpret sarcasm in textual data. This may involve leveraging advanced machine learning models and deep learning architectures trained on larger and more diverse datasets specifically annotated for sarcasm.

Aspect-Based Sentiment Analysis: To further enhance sentiment analysis capabilities, future efforts could delve into aspect-based sentiment analysis. This entails identifying specific aspects or attributes within text and analyzing sentiment associated with each aspect individually. Advanced NLP techniques and machine learning algorithms could be employed to extract and analyze sentiment at a more granular level, providing deeper insights into user opinions and preferences.

Improved Speech-to-Text Accuracy: Addressing the challenge of accurate speech-to-text conversion, future enhancements could focus on developing more robust algorithms capable of handling loud or unclear audio inputs. This may involve incorporating advanced signal processing techniques and machine learning models trained on diverse audio datasets to improve transcription accuracy, especially in challenging audio environments.

Background Music Sentiment Analysis: Expanding the scope of sentiment analysis to include background music, future enhancements could involve developing algorithms capable of classifying music as positive or negative, irrespective of whether it is a structured song or ambient background music. This could be achieved through the integration of audio feature extraction techniques and machine learning models trained on annotated music datasets.

Object-Based Sentiment Analysis: To incorporate object sentiment analysis, future enhancements may require context-aware algorithms capable of analyzing sentiment associated with multiple objects within a scene. This could involve developing sophisticated computer vision techniques and deep learning architectures capable of understanding contextual relationships between objects and inferring sentiment accordingly. Additionally, leveraging multimodal approaches combining visual and textual cues could further enhance the accuracy of object sentiment analysis in complex scenarios.

8. REFERENCES

[1] Abbas, M., Ali, K., Jamali, A., Ali Memon, K., & Aleem Jamali, A. (2019). Multinomial Naive Bayes Classification Model for Sentiment Analysis Overview of China View project Classification for Sentiment Analysis View project Multinomial Naive Bayes Classification Model for Sentiment Analysis.IJCSNS International Journal of Computer Science and Network Security, 19(3), 62. https://doi.org/10.13140/RG.2.2.30021.40169

[2] Abdu, Sarah & Hassan Yousef, Ahmed & Salem, Ashraf. (2021). Multimodal Video Sentiment Analysis Using Deep Learning Approaches, a Survey. Information Fusion. https://doi.org/10.1016/j.inffus.2021.06.003

[3] Alotaibi, F. M. (2019). Classifying Text-Based Emotions Using Logistic Regression. VAWKUM Transactions on Computer Sciences, 7(1), 31–37. https://doi.org/10.21015/vtcs.v16i2.551a

[4] Baevski, A., Zhou, H., Mohamed, A., & Auli, M. (2020). wav2vec 2.0: A framework for self-supervised learning of speech representations. Advances in Neural Information Processing Systems, 2020-December(Figure 1), 1–12.

[5] Gunawan, T. S., Ashraf, A., Riza, B. S., Haryanto, E. V., Rosnelly, R., Kartiwi, M., & Janin, Z. (2020). Development of video-based emotion recognition using deep learning with Google Colab. Telkomnika (Telecommunication Computing Electronics and Control), 18(5), 2463–2471. https://doi.org/10.12928/TELKOMNIKA.v18i5.16717

[6] Jiang, P., Ergu, D., Liu, F., Cai, Y., & Ma, B.(2022). A review of Yolo algorithm developments. Procedia Computer Science, 199, 1066–1073. https://doi.org/10.1016/j.procs.2022.01.135

[7] Liu, T., Han, L., Ma, L., & Guo, D. (2018). Audio-based deep music emotion recognition. AIP Conference Proceedings, 1967(May 2018). https://doi.org/10.1063/1.5039095

[8] L. Stappen, A. Baird, E. Cambria and B. W. Schuller, "Sentiment Analysis and Topic Recognition in Video Transcriptions," in IEEE Intelligent Systems, vol. 36, no. 2, pp. 88-95, 1 March-April 2021, https://doi/10.1109/MIS.2021.3062200

[9] Mellouk, Wafa & Wahida, Handouzi. (2020). Facial emotion recognition using deep learning: review and insights. Procedia Computer Science. 175. 689-694. https://doi.org/10.1016/j.procs.2020.07.101

[10] Scott, M. (2022, May 16). How music affects the emotions of viewers. Royalty Free Music. Retrieved September 19, 2022, from https://www.soundstripe.com/blogs/how-music-affects-the-emotions-of-view