

# Multimodal Sentiment Analysis Using Deep Learning And Attentive Mechanism

Dr. R. S. Gaikwad, Ms. Harshada P. Patil, Ms. Bharati S. Phapale,  
Ms. Swejal A. Phapale, Ms. Monika S. Satpute  
Amrutvahini College of Engineering, Sangamner, India

**Abstract:** In today's digital era, interpreting human emotions from online content has become vital in sectors like marketing, customer support, healthcare, and social media analytics. This paper presents a robust multimodal sentiment analysis framework that combines textual and visual data to gain deeper emotional understanding. The approach employs Long Short-Term Memory (LSTM) networks to grasp contextual and sequential patterns in text, while EfficientNet, a cutting-edge convolutional neural network, is used to extract high-level features from images, including facial expressions and relevant visual cues.

By merging the outputs from both modalities, the model accurately classifies emotions into five categories: very positive, positive, neutral, negative, and very negative. This dual-modality setup addresses the shortcomings of single-source sentiment analysis—such as the vagueness in text or the lack of depth in standalone images. Experimental results reveal that this integrated model delivers significantly higher accuracy and precision compared to traditional unimodal systems.

Designed with real-time applications in mind, the system is well-suited for scenarios like monitoring social media trends, analyzing customer opinions, and improving virtual assistant interactions. The model's performance has been thoroughly validated using standard multimodal datasets, ensuring its adaptability and reliability. Ultimately, this research highlights the effectiveness of deep learning-based multimodal systems in decoding complex human emotions across a wide range of real-world applications.

**Keywords:** Multimodal sentiment analysis, deep learning, image-text integration, machine learning

