

Framework for Detecting Cyberbullying through Social Media Text Images

M. Varalakshmi¹, B. Raj Kamal², S. Upender Rao³

Assistant Professor, Department of IT¹

UG Students, Department of IT^{2,3}

Mahatma Gandhi Institute of Technology Hyderabad, Telangana, India

Abstract: Cyberbullying has become a critical issue with the rise of social media, as harmful behavior increasingly appears in both text and visual forms such as memes and altered images. Traditional detection methods primarily focus on text, often overlooking implicit or symbolic messages embedded in visuals. To address this gap, a comprehensive framework is proposed that integrates Optical Character Recognition (OCR), Natural Language Processing (NLP), and advanced deep learning techniques. Using Tesseract OCR, the system extracts text from images and refines it through NLP processes like slang correction, stemming, and lemmatization. Contextual understanding is achieved through BERT, while a BiLSTM network with attention mechanisms analyzes textual patterns. Concurrently, Convolutional Neural Networks (CNNs) process visual elements to identify symbols and cues missed by text-only approaches. By combining both text and image analysis, the framework significantly enhances detection accuracy across diverse media. Evaluations on labeled datasets demonstrate its effectiveness in identifying subtle and nuanced forms of cyberbullying, surpassing traditional models. This integrated approach sets the foundation for real-time monitoring to promote safer online environments, with future developments aimed at multilingual support, cross-platform detection, and expanded datasets to better capture visual bullying content.

Keywords: Cyberbullying, Social Media, OCR, NLP, BERT, BiLSTM, CNN, Deep Learning, Multimodal Analysis, Online Safety.

