

Abstract

Social media sites, most notably Twitter, have become active centers of community participation, knowledge sharing, and communication in the modern digital era. However, this initial expansion has been followed by an alarming rise in cyberbullying, a ubiquitous problem. Cyberbullying is a broad term that refers to various harmful actions that take place online, such as aggressive bullying, harassment, and intimidation. Those who experience cyberbullying frequently struggle with extreme stress, prolonged sadness, and intense emotional upheaval, which can dangerously deteriorate their mental health. The unsettling link between cyberbullying and its harmful impacts on mental health highlights the urgent need for efficient detection and support methods, especially on intricate and expansive platforms like Twitter. This study assesses four techniques: CNN, GRU, CNN – GRU hybrid, and GRU – CNN hybrid, for detecting cyberbullying on Twitter. GloVe and TF-IDF were employed for feature expansion and extraction. Evaluating each strategy's effectiveness is crucial for identifying instances of cyberbullying. Test results reveal accuracy levels: GRU (80.58%), CNN – GRU hybrid (80.41%), GRU – CNN hybrid (80.37%), and CNN (80.30%). This research contributes to cyberbullying detection on intricate platforms like Twitter, emphasizing the urgency of mitigating its impact.

Keywords: Twitter, cyberbullying, CNN, GRU, hybrid models, GloVe, TF-IDF
