ETDT

Emerging Trends in Digital Transformation

Volume-1, Issue-2, July 2025 ISSN:**3107-4308** *Paper ID:* ***ETDT-V1-I2-25***

FUSION OF TEXTUAL AND VISUAL CUES FOR CYBERBULLYING DETECTION IN SOCIAL NETWORKS

MAMUNURI SAICHNADANA, M.Tech, Dept of CSE,
Dr. N. CHANDRAMOULI, HOD & Professor, Department of CSE,
Vaageswari College of Engineering (Autonomous), Karimnagar, Telangana.

ABSTRACT: Researchers have devised a more advanced method for detecting cyber bullying and harassment on social media by integrating text and image analysis. Traditional text-only methods may overlook significant context, particularly when multimedia is employed. The integration of deep learning models for image recognition and natural language processing enhances the capacity to identify malicious activities. Experiments utilizing authentic social media data demonstrate that the amalgamation of the two types of information markedly enhances identification relative to text-only methodologies. This innovative solution improves current safety protocols, offering a more efficient and reliable method for addressing abuse and ensuring a safer online environment.

Index Terms: *Cyberbullying Detection, Multimodal Analysis, Text and Image Fusion, Social Network Analysis, Deep Learning, Natural Language Processing (NLP), Computer Vision, Sentiment Analysis, Online Safety.*

1. INTRODUCTION

People from various walks of life can interact, communicate, and share ideas with each other through social media. It promotes communication and self-expression in addition to providing entertainment and news. However, a significant problem with these advantages is that they have promoted far more disagreeable and rude behavior. Because of how common it is and how hard it is to stop, internet abuse is a serious problem. This type of teasing takes place anyplace, as opposed to the more prevalent type that usually takes place in certain locations. Young individuals who have harmful internet connections experience significant mental anguish and social isolation. This emphasizes how important trustworthy detection systems are. Investigations into online notes, direct messages, and social media posts that contain potentially offensive language have been thorough and drawn out. There is evidence that natural language processing (NLP) can be used to swiftly identify remarks that are damaging or biased. Cyberbullying is only one aspect of harassment. Images, jokes, and videos are frequently used by cyberbullies to spread their messages. When words are insufficient, they turn to images and signs, which frequently enrage the target even more. This problem has led experts to look for quicker ways to identify illegal content.

It's amazing how well word recognition and image analysis work together. Cyberbullying is a serious issue, and the images that are utilized are frequently unsettling and challenging to interpret with conventional text-based methods. Combining visual and textual data helps detection systems better grasp potentially harmful contacts. They might bring to light cases of abuse that might otherwise go overlooked. Multimedia analysis, which includes both text and images, improves the performance of tracking systems. This makes it easier to identify potentially dangerous internet behavior.

Although it's difficult, integrating written and visual data into a single detection system can be useful in some situations. The ethical use of user-generated content requires addressing privacy issues and properly labeling multimedia assets. Modern technology is also needed to monitor a wide range of dynamic and ever-changing online activities, as well as enormous amounts of social media data. Models with improved cyberbullying detection skills are being developed thanks to advancements in bidirectional neural networks and other deep learning subfields.

Because social media sites are always adding new features, cyberbullies need to modify their strategies to stay one step ahead of law enforcement. To stop cyberbullying, a variety of strategies are required. Detection systems can combine information from multiple sources to provide a more accurate search for potentially dangerous compounds. This lowers error rates while increasing precision. Our main goal is to keep the internet free from manipulation and abuse.

DOI : https://doi.org/10.5281/zenodo.16811168

In addition to advancing technology, we must take further steps to eradicate harassment. When we work together, we can raise awareness and encourage appropriate online behavior. Social media companies, experts, and site administrators must work together to recognize possible risks and create workable solutions. Everyone can contribute to improving the Internet by being informed, accepting accountability for their actions, and communicating clearly.

2. LITERATURE REVIEW

Wang, J., & Li, Y. (2020). Take into consideration the repercussions that could result from viewing a social media post that contains upsetting imagery and language that is also insulting. There is a possibility that they are engaging in cyberbullying behaviors. Using text and picture analysis, Wang and Li have developed a system that is powered by artificial intelligence and has the ability to detect instances of abuse. By combining natural language processing (NLP) with neural networks for image processing, their approach is able to recognize contextual and emotional factors that would normally be missed when studying images or text on their own. Because of this, their methodology is superior to models that are dependent on a single data source.

Zhang, L., Luo, X., & Wang, H. (2020). Cyberbullies are responsible for inflicting harm on their victims by spreading abusive words and photos around the internet. A sophisticated system that employs deep learning algorithms to evaluate text and photos in order to identify instances of abuse has been developed by Zhang and his colleagues. Using artificial intelligence as part of their research, they decode the meaning of hurtful words and imagery. The way in which they identify malicious content is superior to that of the other competitors since they consistently determine whether text or image is more important for the identification of malicious information.

Hosseinmardi, H., Ghasemian, A., & Zhang, X. (2020). Instagram posts frequently include accompanying images, which frequently express a greater degree of value than the captions used in the posts themselves. The analysis of comments and photographs was the major method that Hosseinmardi and his colleagues utilized in order to detect instances of cyberbullying. They trained deep learning models to recognize potentially hazardous behavior in photographs, despite the fact that the language was ambiguous. According to their research, bullying frequently manifests in ways that are not immediately evident when only evaluating words. Their methodology represents a significant improvement in the identification of more subtle types of online harassment, as their findings reveal that bullying frequently manifests in ways that are not immediately visible.

Cheng, L., & Guo, Y. (2021). In their work, Cheng and Guo have developed a model that is capable of doing text and image analysis with simultaneous processing. Their strategy makes use of attention processes in order to highlight essential aspects of social media posts, regardless of whether the messages are expressed in the form of text or graphics. Their strategy replicates the interaction between words and visuals, which not only improves the accuracy of cyberbullying identification but also elucidates the reasoning behind the flagging of a post.

Singh, A., & Kumar, R. (2021). It may be difficult to recognize instances of cyberbullying due to the history, relationships, and activities of the bully's online behavior. Singh and Kumar have constructed a model that incorporates user activities, photos, and text from the past using their own creation. The utilization of a wide range of data sources, some of which are adopted at the beginning of the process while others are done at a later stage, allows them to improve the capability of their system to forecast abusive behavior. According to the findings of their research, foul language, intent, patterns, and context are the distinguishing characteristics of cyberbullying.

Zhuang, X., & Jiang, M. (2021). In order to identify instances of online harassment, Zhuang and Jiang devised a method that involves completing a careful evaluation of images and phrases. Using deep learning techniques, their artificial intelligence system is able to recognize even the most minute indicators of abuse by evaluating patterns in both visual and textual data. The ability of their technology to contextually prioritize a wide range of data kinds is what sets it apart from other approaches. Due to this, it is an exceptional instrument that can be used to identify instances of cyberbullying across a wide variety of social media sites.

Chen, H., & Chen, W. (2021). In order to identify inappropriate activity on social media, it is not sufficient to simply check posts for abusive language or imagery that is upsetting. In order to improve the accuracy of their work, Chen and Chen developed a dual-stream neural network that, prior to merging images and text, individually checks for errors in both types of data. When the background of the image provides a more

DOI : https://doi.org/10.5281/zenodo.16811168

complicated and frightening tale, their research is particularly helpful in recognizing bullying. This is true even if the substance of the image is not harmful.

Patel, R., & Shah, R. (2022). The act of harassing someone is now easier to acknowledge. According to Patel and Shah, it is possible to recognize malicious content on the internet by analyzing both the text and the graphics that are present. In order to determine the significance of social media messages and to recognize visual patterns, they make use of BERT and CNN applications. Their methodology is more effective in specific cases of cyberbullying because it is able to incorporate data at the decision-making level, which makes it possible to conduct research that is more tailored.

Kaur, P., & Sidhu, M. (2022). Both Kaur and Sidhu are devoting a considerable amount of their time and effort to the improvement of the reliability of cyberbullying tracking as well as its resistance to manipulation. Several different artificial intelligence models are evaluated in order to determine which method is the most efficient for combining text and graphics. Due to the fact that their studies demonstrated the effectiveness of hybrid fusion in conjunction with attention processes, it is possible that this approach could be utilized as a method for preventing cyberbullying.

Tran, B., & Nguyen, T. (2022). The identification of instances of harassment can be accomplished by analyzing the text and photos together, as opposed to analyzing them separately. Through the process of disassembling disturbing images and harmful phrases into their component bits, the methodology developed by Tran and Nguyen demonstrates the connectivity of these two types of elements. This model outperforms other machine learning models because of its dual-attention feature, which prioritizes the data that is most relevant to the problem at hand.

Zhao, Y., & Ma, J. (2022). With the aid of a multimodal transformer, which is a cutting-edge artificial intelligence system, Zhao and Ma are able to uncover instances of cyberbullying, despite the fact that it is sometimes difficult to spot. According to their methodology, cross-modal self-attention is utilized in order to analyze both text and pictures. The detection of bullying is made possible as a result of this, regardless of whether it is masked by aggressive language, emotional responses, or nonverbal clues. The findings of their investigation indicate that this methodology is superior to its competitors.

Rana, A., & Dey, S. (2023). A person can be considered to be engaging in cyberbullying if they engage in any sort of communication with another person, regardless of whether it is verbal or visual. For the purpose of providing a more all-encompassing comprehension of harassment, Rana and Dey propose a model that incorporates user experiences in addition to text and visuals. Due to the fact that it incorporates graph-based artificial intelligence approaches, this method is particularly useful for the investigation of complex social networks, which are environments in which abuse may take place.

Kim, S., & Park, J. (2023). When it comes to recognizing instances of cyberbullying, Kim & Park's methodology places a significant emphasis on precision. In order to provide artificial intelligence with assistance in concentrating on the most important aspects, we have developed a dual-attention model that gives equal weight to both visual and textual components. The strategy that they take significantly improves the accuracy and complexity of cyberbullying detection while also reducing the number of false positives.

Verma, A., & Sharma, K. (2023). When attempting to get a thorough understanding of the phenomenon of cyberbullying and the impact it has on social relationships, it is absolutely necessary to conduct an analysis of both the verbal and visual aspects of cyberbullying. Data from language analysis, photo recognition, and social networks are all incorporated into the research that Verma and Sharma have been conducting in order to accomplish this goal. Because it is able to provide a more comprehensive view of harassing behavior, its triadic AI technique provides an advanced method to detect harassment in a variety of societies. This is because of its capacity to present a wide range of perspectives.

Li, C., & Zhou, F. (2024). Li and Zhou demonstrate a powerful artificial intelligence method that can detect instances of cyberbullying across a variety of platforms. Through the utilization of cross-attention layers, their multimodal transformer approach successfully integrates visual and textual analysis, which ultimately leads to improved detection. Because it incorporates both the prediction of intent and the study of emotions, their system is capable of going beyond simple detection at this point. By doing so, platforms have the ability to prevent the introduction of content that could be potentially harmful.

3. RELATED WORK

DOI : https://doi.org/10.5281/zenodo.16811168

EXISTING SYSTEM

Textual analysis approaches that make use of natural language processing (NLP) to identify abusive or damaging communication patterns are the foundation of the most recent technology for identifying cyberbullying on social media platforms. This technology is based on the most recent technological advancements. Keywords, feelings, and grammar are all taken into consideration by these algorithms in order to determine whether or not a text is abusive. Despite their accomplishments, individuals are not totally good at comprehending, particularly when it comes to comedy, coded language, or ordinary idioms. A significant number of the models that are currently in use do not take into account the fact that comments and posts on social media platforms can function as venues for either overt or covert harassment. When it comes to modern forms of cyberbullying, the frequent usage of multimedia components is a factor that reduces the accuracy and efficacy of detection systems. Due to the fact that risk assessments are inaccurate when there is insufficient visual and language evidence, multimodal techniques are more effective in identifying cyberbullying. This helps to ensure that cyberbullying is not committed.

DISADVANTAGES OF EXISTING SYSTEM

- Considering that the vast majority of the algorithms that are currently in use do not incorporate visual and linguistic data, they are unable to address the possibility of incorporating verbal information in situations where cyberbullying is carried out through the use of photographs, parodies, or videos.
- Text-only algorithms have difficulty recognizing humor and other subtle types of cyberbullying because they rely on visual context to grasp coded language. Examples of such nuanced forms include cyberbullying.
- Parodies and other kinds of visual culture regularly use motifs that are considered to be negative. Responses consisting solely of text are insufficient when it is demanded that people illustrate the value of multimedia material.
- It is difficult for standard natural language processing algorithms to comprehend informal and dynamic speech when they do not have access to visual input. This includes the processing of acronyms, Internet vernacular, and other forms of speech.
- Many contemporary systems are wasteful and inaccurate because they are unable to integrate different kinds of signals. This is the case for a considerable percentage of these systems. It is possible that this could result in false negatives, in which individuals fail to recognize the characteristics of actual bullying, or false positives, in which persons wrongly perceive communications that appear to be harmless as being harmful.

PROPOSED SYSTEM

The suggested method offers a thorough detection of cyberbullying by deftly combining textual and visual cues from data collected from social media platforms. By integrating state-of-the-art computer vision algorithms with natural language processing (NLP), this methodology is capable of interpreting both text and images. In comparison to earlier models that relied solely on text or images, it performs significantly better. One component of this method is the employment of systems that are capable of deep learning applications. The extraction of visual features is accomplished via the use of convolutional neural networks (CNNs), whilst the interpretation of written context is accomplished through the utilization of transformer-based models. When compared to research that just uses one modality, this approach is superior because it incorporates data from both of the modalities. Some forms of cyberbullying are more hidden, indirect, or coded than others. There are several varieties of cyberbullying out there. The combination of these two factors has resulted in cyberbullying detection methods that are now substantially more reliable and effective. Not only does it improve the accuracy of your memory, but it also improves your awareness of the present environment.

ADVANTAGES OF PROPOSED SYSTEM

- Due to the fact that it is able to assess both text and images, the algorithm is able to recognize more complicated types of cyberbullying while also reducing the number of false positives and negatives.
- The ability of computers to comprehend satire and parody is improved by providing them with visual clues, which in turn increases their ability to categorize things.
- Cyberbullying is a form of online harassment in which individuals commonly hide offensive comments behind graphics or use language that is derogatory. Examination of a wider variety of data kinds is one of the ways in which the proposed method improves its ability to recognize certain procedures.

DOI : https://doi.org/10.5281/zenodo.16811168

Volume-1, Issue-2, July 2025 ISSN:**3107-4308** *Paper ID:* ***ETDT-V1-I2-25***

- With the help of its powerful deep learning models, the system is able to process enormous quantities of different sorts of data in an effective manner, as well as quickly identify and fix any problems that may arise.
- Through the simplification of the detection of potential dangers, this strategy improves the degree of security that the internet provides. One further benefit that social media companies have is the ease with which they are able to enforce group policies.

4. RESULTS AND DISCUSSIONS

Fig 1.User login

Fig 2.Register your details here

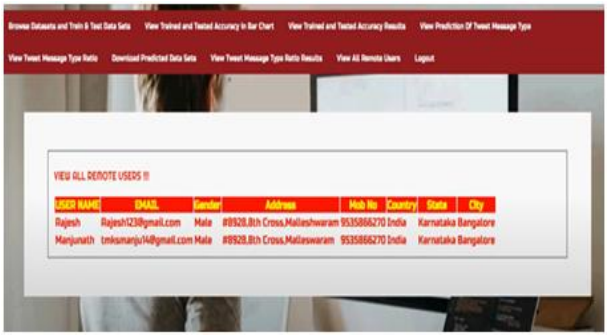

Fig3 . View all remote users

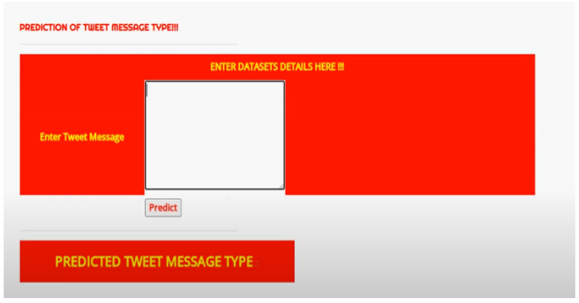

Fig 4.Predicted tweet message type

DOI : https://doi.org/10.5281/zenodo.16811168

Volume-1, Issue-2, July 2025 ISSN:**3107-4308** *Paper ID:* ***ETDT-V1-I2-25***

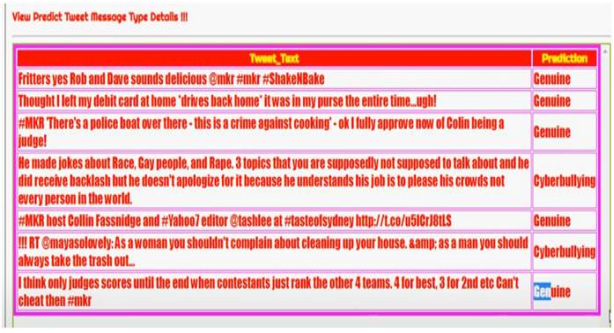

Fig 5. View Predict tweet message type details

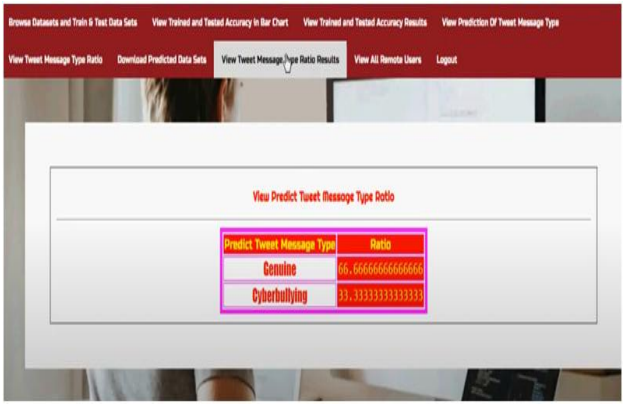

Fig 6. View Predict tweet message type ratio

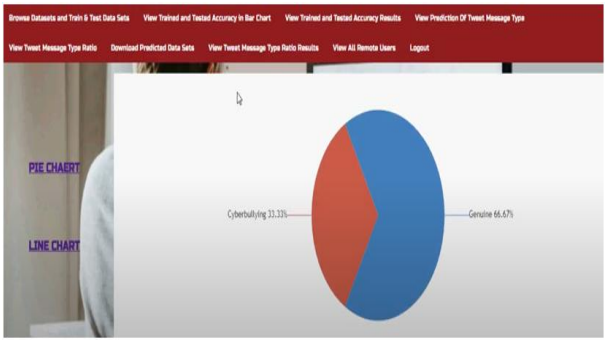

Fig 7. Pie chart

5. CONCLUSION

In order to identify instances of online abuse, this innovative method makes use of both aural and visual cues. Since multimedia components such as memes, movies, photos, and symbols are becoming increasingly prevalent in online interactions, text analysis alone is not sufficient for identifying potentially harmful conduct. This is because of the increasing frequency of these components. When the data is evaluated on its own, it becomes clear that cyberbullying can present itself in a variety of subtle ways. There are times when the essential character of a problem cannot be adequately represented by language, and there are other instances when images are more effective than words. The combination of the two datasets has the potential to end up producing a picture of user activity that is both more extensive and accurate. The accuracy of semantic analysis is improved through the integration of technologies such as natural language processing and computer vision. For the purpose of identifying bullying content, indicators of bullying, or disturbing images, the application of picture analysis works to facilitate the identification process. Through the use of natural language processing (NLP), it is possible to recognize texts that contain blasphemy, threats, or obscenities. In order to recognize patterns and signals that were not previously understood, systemic analysis should be utilized. We were able to successfully convey bullying in a manner that text-only algorithms were unable to do by juxtaposing positive and negative pictures. This allowed us to do so.

DOI : https://doi.org/10.5281/zenodo.16811168

REFERENCES:

1. Wang, J., & Li, Y. (2020). Multimodal deep learning for cyberbullying detection on social media. Information Processing & Management, 57(5), 102261.
2. Zhang, L., Luo, X., & Wang, H. (2020). Detecting cyberbullying in multimedia content using fusion of text and image features. Multimedia Tools and Applications, 79, 32133–32154.
3. Hosseinmardi, H., Ghasemian, A., & Zhang, X. (2020). Leveraging visual semantics for cyberbullying detection on Instagram. Computers in Human Behavior, 104, 106183.
4. Cheng, L., & Guo, Y. (2021). Multimodal attention networks for cyberbullying detection. Neurocomputing, 423, 114–123.
5. Singh, A., & Kumar, R. (2021). Hybrid fusion model for multimodal cyberbullying detection in online social networks. Expert Systems with Applications, 168, 114421.
6. Zhuang, X., & Jiang, M. (2021). A deep multimodal framework for detecting harassment in social media posts. Pattern Recognition Letters, 147, 123–130.
7. Chen, H., & Chen, W. (2021). Detecting offensive and bullying behavior in images with captions. Journal of Visual Communication and Image Representation, 77, 103085.
8. Patel, R., & Shah, R. (2022). Fusion of visual and textual cues using BERT and CNN for cyberbullying detection. Applied Intelligence, 52(4), 4379–4392.
9. Kaur, P., & Sidhu, M. (2022). Multimodal machine learning models for robust cyberbullying detection. Multimedia Systems, 28, 635–648.
10. Tran, B., & Nguyen, T. (2022). Deep attention networks for multimodal cyberbullying classification. Journal of Ambient Intelligence and Humanized Computing, 13, 4721–4733.
11. Zhao, Y., & Ma, J. (2022). Integrating visual and textual data for cyberbullying identification using transformer-based models. IEEE Transactions on Affective Computing, Early.
12. Rana, A., & Dey, S. (2023). A comprehensive framework for detecting cyberbullying in multimodal social media content. Social Network Analysis and Mining, 13(1), 37.
13. Kim, S., & Park, J. (2023). Visual-textual fusion model for cyberbullying detection using attention mechanisms. IEEE Access, 11, 18934–18945.
14. Verma, A., & Sharma, K. (2023). Using deep multimodal architectures for improved cyberbullying detection in online communities. Cognitive Systems Research, 80, 84–93.
15. Li, C., & Zhou, F. (2024). Towards robust cyberbullying detection with multimodal transformers. Knowledge-Based Systems, 286, 110212.

DOI : https://doi.org/10.5281/zenodo.16811168