Toward Multimodal Cyberbullying Detection

Vivek K. Singh
Rutgers University
New Brunswick, NJ 08904, USA
v.singh@rutgers.edu

Souvick Ghosh
Rutgers University
New Brunswick, NJ 08904, USA
souvick.ghosh@rutgers.edu

Christin Jose
Rutgers University
New Brunswick, NJ 08904, USA
christin.jose@rutgers.edu

Abstract

As human beings utilize computing technologies to mediate
multiple aspects of their lives, cyberbullying has grown as
an important societal challenge. Cyberbullying may lead to
deep psychiatric and emotional disorders for those affected.
Hence, there is an urgent need to devise automated meth-
ods for cyberbullying detection and prevention. While recent
cyberbullying detection efforts have defined sophisticated
text processing methods for cyberbullying detection, there
are as yet few efforts that leverage visual data processing to
automatically detect cyberbullying. Based on early analysis
of a public, labeled cyberbullying dataset, we report that vi-
sual features complement textual features in cyberbullying
detection and can help improve predictive results.

ACM Classification Keywords

J.4 [Computer Applications]: Social and Behavioral Sci-
ences

Author Keywords

Cyberbullying detection; Social Media Analysis

Introduction

According to National Crime Prevention Council report,
more than 40% of teenagers in the US have reported being
cyberbullied [2]. Multiple studies have highlighted the neg-
ative effects of cyberbullying [48, 6], which include deep

emotional trauma, psychological and psychosomatic disorder, and in extreme cases, even suicide [18].

Cyberbullying may be defined as "When the Internet, cellphones or other devices are used to send or post text or images intended to hurt or embarrass another person" [13]. While many researchers have worked on the effects of cyberbullying on teenagers [25, 50] and also tried to identify automated methods for cyberbullying detection [13, 12, 40], such approaches are yet to consider the dramatically changed social media landscape that the teenagers are dealing with now compared to even five or ten years ago. For example, recent CHI studies have reported that teenagers make extensive use of image and video sharing apps (e.g.,Instagram, Vine) and limited time messages (e.g., Snapchat) [39]. Specifically, visual (image, video) content now accounts for more than 70% of all web traffic [38, 33]. At the same time, there has been a significant growth in using image and video content for cyberbullying [42, 25, 44], and it has been argued that "cyberbullying grows bigger and meaner with photos, video" [35]. In fact, one of the important themes identified in our recent cyberbullying focus group study was the growing prevalence of image and multimodal content for cyberbullying [45].

While the importance of understanding multimodal content for cyberbullying detection has been widely acknowledged, the cyberbullying detection literature is still primarily focused on (sophisticated) text processing [13, 12, 40, 53], and their accuracy remains limited. There are as yet, few efforts that leverage the visual features for cyberbullying detection. This is in part attributable to the sophisticated expertise needed in processing images or videos for handling such issues. With the availability of recent computer vision Application Programming Interfaces (APIs) such as Microsoft's Project Oxford, Face++, Imagga, CMU's Open-

Face, image analysis has become accessible to larger pool of HCI and information scientists. These APIs can be used for studying diverse social, political, and information interaction issues and are being readily adopted by HCI researchers in different contexts (e.g.: [5, 20]). Hence, this paper focuses on using these APIs for deriving visual features and detecting cyberbullying.

Cyberbullying is an important societal problem and is actively researched in multiple disciplines (e.g. education, psychology, libraries, data mining, HCI). Different strands of research focus on different aspects including understanding the phenomenon, its psychological implications, automatic detection, and mitigation steps. While this extended abstract focuses on one aspect of this research - automatic detection - it is part of our larger research agenda to understand, detect, and reduce cyberbullying [45, 46]. With further enhancements we expect the obtained results to feed into novel ways to actively counter cyberbullying. This is in similar spirit as discussed in other recent efforts by the CHI community, such as [4, 17, 49].

Proposed Approach

Most of the previous research on cyberbullying detection focuses on text-based approaches. In this work, we concentrate on designing predictive models for automatic cyberbullying detection, which include the use of visual features (e.g. gender, race, nudity, portrayed emotions etc.) to complement textual features. Hence, we closely follow recent works using textual features for cyberbullying detection [41, 13, 15, 30, 47] as well as utilizing multiple channels of data (e.g. social and textual [31, 46] for cyberbullying detection.

Given the relatively underdeveloped literature on visual cyberbullying detection, in this late-breaking report we focus

on features that make sense based on the findings made using other methods (focus groups or text-based analysis), as reported in the literature (e.g. [40, 36, 24]).

Just as in the early days of natural language processing, the accuracy of these individual detectors varies greatly (typically ranging from 40% to 96%, depending on the criteria used) [22, 3, 23]. Similarly, reliable detectors for certain other relevant concepts (e.g. obesity, make-up) are not yet easily available for the visual content. While this may change in the near future, in this research we do not consider image-based features as a replacement for textual features but rather as a complementary approach. Consistent with the "gestalt" principle (the whole is greater than the sum of its parts) [34], we hypothesize that a combination of modestly accurate features coming from heterogeneous data modalities can outperform methods that employ a single modality [19, 32, 37].

Textual features used in cyberbullying detection
The value of textual features for cyberbullying detection is well established [40, 36, 11, 41]. In our work, we consider textual features indicative of emotion, gender specific terminology, sexual connotations, as well as the relative distribution of different parts of speech [40, 36, 11, 41, 13, 15]. Specifically, we use Linguistic Inquiry and Word Count (LIWC) to analyze the text for cyberbullying. LIWC provides more than 90 descriptive variables which includes word counts, summary variables, various word categories, personal concern categories, details and frequency of punctuations and the informal languages used. The list of features includes:

- **Psychological Processes** It includes various subcategories like anger, negative emotion, sadness, anxiety, etc. Words depicting anger (e.g. mad, kill), negative emotion (e.g. nasty), sadness (e.g. grief, gloomy), and

anxiety (e.g. agitated, tensed) often indicates cyberbullying.
- **Tone and Word count of the comments** Prior literature suggests that the tone in cyberbullying sessions are usually more aggressive and contains higher word counts [52].
- **Presence of informal language** Informal languages involve the use of swear words like damn, shit, etc. Swear or curse words are often used to direct abuses to victims.
- **Use of sexual words** Words like horny, incest, love which has a sexual connotation have reported higher instances of cyberbullying [16].
- **Words expressing personal concern** Posts involving words like bury, kill, death has a negative emotion attached to it. These words could act as indicators of cyberbullying.
- **Use of third person singular pronouns** Cyberbullying literature has frequently identified that cyberbullying involves more thirds person singular pronouns than normal text [7].

Visual features used in cyberbullying detection
As mentioned earlier, features extracted from images may also play an important role in cyberbullying detection. We used Microsoft's Project Oxford to extract visual features from the Instagram images [1]. It is a computer vision API which analyzes the images for the dominant colors, number of people present, category, adult content, etc. Based on existing literature [26, 51, 28, 43, 45], the following features were obtained using the API and used for cyberbullying detection:

- **Age and gender of people in the image** Multiple earlier efforts have suggested a connection between de-

mographic variables like age and gender and the occurence of cyberbullying [26, 51]

- **Image Category** The taxonomy-based categories of the image. This includes the presence of tattoos, graffiti, drugs as suggested by [27] as well as a generic description of the content of the image.

- **Presence of racy or adult content** Inappropriate or sexually suggestive content can be a target for cyberbullying. The computer vision API tags all images for racy content and assigns a score to them. Pornographic or sexual content is another category which is often closely associated with cyberbullying. Multiple recent efforts have connected cyberbullying with sexual and racy content [43, 45].

- **Image Type** Describes if the image is black and white, or color, or if the image is a clipart or line drawing.

Experiments and Results

Corpus
Our work uses the dataset made available by Hosseinmardi et al. [29]. This dataset contains about 2000 media sessions from Instagram, including the posted image, caption, and comments from other users. However, only 699 of these media sessions still have images accessible from the recorded Instagram URL. We focus on these 699 labeled media sessions for the analysis in this paper.

Hosseinmardi et al. [29] used a snowball sampling method to identify Instagram ids. Each image and the accompanying comments were considered a media session. Only those sessions were considered which contained more than 15 comments. The authors used a dictionary of profanity words to determine the percentage of negativity in the sessions. Each media session in the resulting dataset was hand-labelled by five crowd-sourced (CrowdFlower) annotators for presence of cyberbullying [29].

Classification
In this work, we attempt to build an automatic machine learning based classifier to detect media sessions containing cyberbullying. For the analysis of textual and visual features, we have used 461 (i.e. 66%) instances for training and 238 (i.e. 34%) instances for testing purposes. We applied SMOTE technique to deal with imbalances in the training data [10]. SMOTE over samples the minority class and under samples the majority class. It over samples the minority class by generating newer (synthetic) examples by operating in 'feature space' rather than 'data space', thus countering some of the problems caused by simple replication of data points [10]. Once we train our model using a balanced training dataset, we test the model using the realistic, unbalanced test data. After minority oversampling, we had 781 instances for training and 238 instances for test.

The performance of different automated algorithms were compared against the hand-labelled "ground truth". We used Weka 3.0 for classification of the instances. We used the Classifier Subset Selection [21] approach to undertake feature selection and the Bagging (Bootstrapping Aggregation) [8] algorithm for classification.

The features selected by the algorithm: for textual, visual, and combined models for cyberbullying detection are shown in Table 1. Textual features selected include those which were along the expected lines - positive and negative emotions (anger, sadness), focus on the present time, etc. - as well as a few unexpected features (e.g. clout and insights exhibited by the commenter). We leave a detailed analysis of these features outside the scope of this short report. Similarly, visual features include those with adult content, raciness, median age as well as some structural image de-

Textual Features
Word count, positive emotions, anger in comments, sadness, causation implied in comments, tentativeness, drives for reward, number of dashes (punctuation), clout exhibited, insights exhibited, focus on the present time, and focus on relative association with surroundings.
Visual Features
Adult score, raciness, median age, text signs present, people portraits, other categories (i.e. outside the frequent categories defined by the API), abstract-shape, object- sculpture, abstract-nonphoto, object-screen present, image darkness.
Multimodal Model - Textual and Visual Features
Visual: people in a performance, text signs present, abstract-rectangles, outdoor scenes, *Textual:* comparisons made, sadness, certainty, health related, sexual, informal non-fluencies (e.g. err, hmmm), all punctuations present, punctuations - 'others', word count, dictionary words, total functional words, third person pronouns, perceptual processes, feelings expressed

Table 1: Features selected by the algorithm for cyberbullying detection

Approach	Accuracy	ROC Area
Textual	77.6%	0.746
Visual	70.5%	0.452
Combined	81.4%	0.844

Table 2: Classification results for textual, visual and composite models for cyberbullying detection

scriptors. Lastly, the combined model used both textual and visual features. We plan to undertake more detailed analysis of these features in our future work.

One of the peculiarities expected in cyberbullying prediction is that of dealing with highly imbalanced classes [31]. Hence, besides the traditional accuracy mesure, we also consider the Receiver Operating Characteristic (ROC) metric that is often used for evaluating the performance of such classifiers [9, 31].

The results of classification using textual, visual and combined features are shown in Table 2. While the classification using visual features show lower accuracy when compared to that using textual features, the result is not surprising as text-based analysis have received a lot more attention in recent years and have taken considerable strides in data analytics. The combined multimodal model shows a higher Accuracy and ROC score, compared to both visual and textual models considered individually. This is evident from Table 2 (0.844 ROC score for the combined model compared to 0.746 for the text only model indicating a 13.1% improvement in classification results.)

To gain an understanding of the kind of scenarios where the multimodal approach performs better than individual modalities, we zoomed into one such example. The example of Figure 1 includes the image posted by an user as well as five of the comments made on it. Human annotators identified the image as an example of cyberbullying. However, both textual and visual feature based models, working in isolation, failed to detect cyberbullying in this media session. As can be observed, while the original intent of the post was to promote a show involving dolphins, multiple commenters posted sexually explicit comments in response. There were also instances of other commenters supporting the original poster. Given the mixed (positive + negative) nature of comments as well as the references from text to the image (tattoo) it is understandable how individually, both the visual and the textual models failed to detect this post as involving cyberbullying. However, the combination of weak clues available from textual and visual features was perhaps able to push the confidence past the threshold for appropriate detection in the combined model.

This shows an early but practical example of multimodal cyberbullying detection, where a combination of weak clues

coming from multiple modalities may be useful for more accurate detection.
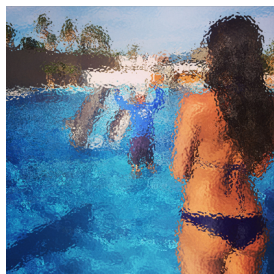
Original Post: Pool Day Anyone?! You'll see more of us swimming with rescued dolphins this Sunday on <name of a show>
Sample comments:
(1) DAT ass thoe
(2) What dolphins?
(3) Can i give u the other tattoo??
(4) Lemme <abusive term> that butt babe till it turns red and it hurts
(5) I get she is beautiful but you have to learn how to respect a beautiful woman

Figure 1: An example of a post which was labelled correctly for cyberbullying by a combination of textual and visual features. Note: Image blurred and some potentially identifying information removed.

Admittedly, the dataset employed in this work is modest sized ($N =\sim 700$) and considers a single social media site. Hence, we consider the results of this late-breaking effort as early and exploratory. However, the results obtained motivate and open up a new sub-area (multimodal cyberbullying detection using image analysis APIs) for further investigation and research contributions.

With the overwhelming amount of textual and visual content being generated in social media sites like Facebook, Instagram, Snapchat, it is almost impossible to detect cyberbullying manually. We posit that this work on automatic detection could provide an initial feedback to the relevant stakeholders (social network administrators, law enforcement authorities, parents, school authorities, peers) on possible cases of cyberbullying, thus, allowing them to manually validate the messages. This automatic detection will act as a "triage" step, scaling down "big data" to manageable scale for manual confirmation. Such an initial screening mechanism can also be used to trigger self-reflective interfaces [14] that

urge posters to re-think their potentially harmful messages on an online platform. This approach is applicable to all social media applications, which support image and textual content. In future, we can imagine similar approaches to be used to prevent cyberbullying incidents in different online networks.

Conclusion

This work proposes the use automated image and text analysis APIs for multimodal (visual and textual) cyberbullying detection. Increasing interactions with mobile devices have led to increase in visual content being exchanged online. Future research efforts could leverage the availability of vast quantity of visual content as well as visual processing APIs for better cyberbullying detection. While text-based features have outperformed visual features in the current dataset, this study highlights a promising new direction and reports that the performance of automatic cyberbullying detectors improves with the use of a multimodal (textual + visual) approach compared to a text only approach. With future refinements, the multimodal approach can play a significant role in reducing cyberbullying and improving the quality of life of thousands of individuals who are affected by cyberbullying each year.

References

[1] Computer Vision API. https://www.microsoft.com/cognitive-services/en-us/computer-vision-api. (????). Accessed: 2016-06-10.

[2] 2014. Stop bullying before it starts. http://www.ncpc.org/resources/files/pdf/bullying/cyberbullying.pdf. (2014).

Accessed: 2016-06-10.

[3] 2016. Gender/Age/Race Analysis. http://www.faceplusplus.com/tech_gender/. (2016). Accessed: 2016-06-10.

[4] Zahra Ashktorab and Jessica Vitak. 2016. Designing Cyberbullying Mitigation and Prevention Solutions through Participatory Design With Teenagers. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 3895–3905.

[5] Saeideh Bakhshi, David A Shamma, and Eric Gilbert. 2014. Faces engage us: Photos with faces attract more likes and comments on instagram. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. ACM, 965–974.

[6] Linda Beckman, Curt Hagquist, and Lisa Hellström. 2012. Does the association with psychosomatic health problems differ between cyberbullying and traditional bullying? *Emotional and behavioural difficulties* 17, 3-4 (2012), 421–434.

[7] Kate Blackburn, Nicholas Brody, and Leah LeFebvre. 2014. The IâĂŹs, WeâĂŹs, and She/HeâĂŹs of Breakups Public and Private Pronoun Usage in Relationship Dissolution Accounts. *Journal of Language and Social Psychology* 33, 2 (2014), 202–213.

[8] Leo Breiman. 1996. Bagging predictors. *Machine learning* 24, 2 (1996), 123–140.

[9] Nitesh V Chawla. 2005. Data mining for imbalanced datasets: An overview. In *Data mining and knowledge discovery handbook*. Springer, 853–867.

[10] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. 2002. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research* 16 (2002), 321–357.

[11] Ying Chen, Yilu Zhou, Sencun Zhu, and Heng Xu. 2012. Detecting offensive language in social media

to protect adolescent online safety. In *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)*. IEEE, 71–80.

[12] Maral Dadvar, Dolf Trieschnigg, Roeland Ordelman, and Franciska de Jong. 2013. Improving cyberbullying detection with user context. In *Advances in Information Retrieval*. Springer, 693–696.

[13] Karthik Dinakar, Birago Jones, Catherine Havasi, Henry Lieberman, and Rosalind Picard. 2012a. Common sense reasoning for detection, prevention, and mitigation of cyberbullying. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 2, 3 (2012), 18.

[14] Karthik Dinakar, Birago Jones, Catherine Havasi, Henry Lieberman, and Rosalind Picard. 2012b. Common sense reasoning for detection, prevention, and mitigation of cyberbullying. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 2, 3 (2012), 18.

[15] Karthik Dinakar, Roi Reichart, and Henry Lieberman. 2011a. Modeling the detection of Textual Cyberbullying. In *The Social Mobile Web*.

[16] Karthik Dinakar, Roi Reichart, and Henry Lieberman. 2011b. Modeling the detection of Textual Cyberbullying. In *The Social Mobile Web*.

[17] Mingyue Fan, Liyue Yu, and Leanne Bowler. 2016. Feelbook: A Social Media App for Teens Designed to Foster Positive Online Behavior and Prevent Cyberbullying. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 1187–1192.

[18] Tyler Clementi Foundation. Tyler's Story. http://tylerclementi.org/tylers-story/#. (????). Accessed: 2016-06-10.

[19] Gerald Friedland and Ramesh Jain. 2014. *Multimedia Computing*. Cambridge University Press.

[20] Venkata Rama Kiran Garimella, Abdulrahman Alfayad, and Ingmar Weber. 2016. Social Media Image Analysis for Public Health. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 5543–5547. DOI:http://dx.doi.org/10.1145/2858036.2858234

[21] Mark A Hall and Lloyd A Smith. 1998. Practical feature subset selection for machine learning. (1998).

[22] Hu Han, Charles Otto, Xiaoming Liu, and Anil K Jain. 2015. Demographic estimation from face images: Human vs. machine performance. *IEEE transactions on pattern analysis and machine intelligence* 37, 6 (2015), 1148–1161.

[23] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Deep residual learning for image recognition. *arXiv preprint arXiv:1512.03385* (2015).

[24] Sameer Hinduja and Justin W Patchin. 2010. Bullying, cyberbullying, and suicide. *Archives of suicide research* 14, 3 (2010), 206–221.

[25] Sameer Hinduja and Justin W Patchin. 2012a. *School climate 2.0: Preventing cyberbullying and sexting one classroom at a time*. Corwin Press.

[26] Sameer Hinduja and Justin W Patchin. 2012b. *School climate 2.0: Preventing cyberbullying and sexting one classroom at a time*. Corwin Press.

[27] Homa Hosseinmardi, Sabrina Arredondo Mattson, Rahat Ibn Rafiq, Richard Han, Qin Lv, and Shivakant Mishra. 2015a. Analyzing Labeled Cyberbullying Incidents on the Instagram Social Network. In *International Conference on Social Informatics*. Springer, 49–66.

[28] Homa Hosseinmardi, Sabrina Arredondo Mattson, Rahat Ibn Rafiq, Richard Han, Qin Lv, and Shivakant Mishra. 2015b. Detection of cyberbullying incidents on the instagram social network. *arXiv preprint arXiv:1503.03909* (2015).

[29] Homa Hosseinmardi, Sabrina Arredondo Mattson, Rahat Ibn Rafiq, Richard Han, Qin Lv, and Shivakant Mishra. 2015c. Detection of cyberbullying incidents on the instagram social network. *arXiv preprint arXiv:1503.03909* (2015).

[30] Qianjia Huang, Vivek Kumar Singh, and Pradeep Kumar Atrey. 2014a. Cyber bullying detection using social and textual analysis. In *Proceedings of the 3rd International Workshop on Socially-Aware Multimedia*. ACM, 3–6.

[31] Qianjia Huang, Vivek Kumar Singh, and Pradeep Kumar Atrey. 2014b. Cyber bullying detection using social and textual analysis. In *Proc. Int. Workshop on Socially-Aware Multimedia*. ACM, 3–6.

[32] Alejandro Jaimes and Nicu Sebe. 2007. Multimodal human–computer interaction: A survey. *Computer vision and image understanding* 108, 1 (2007), 116–134.

[33] Peter Kafka. Streaming Video Now Accounts for 70 Percent of Broadband Usage. http://www.recode.net/2015/12/7/11621218/streaming-video-now-accounts-for-70-percent-of-broadband-usage. (????). Accessed: 2016-06-10.

[34] Kurt Koffka. 2013. *Principles of Gestalt psychology*. Vol. 44. Routledge.

[35] Janet Kornblum. 2008. Cyberbullying grows bigger and meaner with photos, video. *USA Today* (2008).

[36] Franklin B Krohn. 2004. A generational approach to using emoticons as nonverbal communication. *Journal of technical writing and communication* 34, 4 (2004), 321–328.

[37] Michael S Lew, Nicu Sebe, Chabane Djeraba, and Ramesh Jain. 2006. Content-based multimedia information retrieval: State of the art and challenges. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 2, 1 (2006), 1–19.

[38] Carla Marshall. Cat Videos, Binge TV Watching Will Account for 84 Percent of Internet Traffic, Cisco Says. http://www.reelseo.com/2019-internet-video-traffic/. (????). Accessed: 2016-06-10.

[39] Jessica A Pater, Andrew D Miller, and Elizabeth D Mynatt. 2015. This Digital Life: A Neighborhood-Based Study of Adolescents' Lives Online. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2305–2314.

[40] Kelly Reynolds, April Kontostathis, and Lynne Edwards. 2011. Using machine learning to detect cyberbullying. In *Machine Learning and Applications and Workshops (ICMLA), 2011 10th International Conference on*, Vol. 2. IEEE, 241–244.

[41] Marlies Rybnicek, Rainer Poisel, and Simon Tjoa. 2013. Facebook Watchdog: A Research Agenda for Detecting Online Grooming and Bullying Activities. In *2013 IEEE International Conference on Systems, Man, and Cybernetics*. IEEE, 2854–2859.

[42] Steven J Seiler and Jordana N Navarro. 2014a. Bullying on the pixel playground: Investigating risk factors of cyberbullying at the intersection of childrenâĂŹs online-offline social lives. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace* 8, 4 (2014).

[43] Steven J Seiler and Jordana N Navarro. 2014b. Bullying on the pixel playground: Investigating risk factors of cyberbullying at the intersection of childrenâĂŹs online-offline social lives. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace* 8, 4 (2014).

[44] Shaheen Shariff. 2014. *Sexting and Cyberbullying*. Cambridge University Press.

[45] Vivek Singh, Marie Radford, Huang Qianjia, and Susan Furrer. 2017. "They basically like destroyed the school one day": On Newer App Affordances and Cyberbullying in Schools. In *(To appear:) Proceedings of the international conference on Computer Supported Collaborative Work and Social Computing*. ACM.

[46] Vivek Kumar Singh, Qianjia Huang, and Pradeep Kumar Atrey. 2016a. Cyberbullying Detection Using Probabilistic Socio-Textual Information Fusion. In *Proc. IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. ACM.

[47] Vivek Kumar Singh, Qianjia Huang, and Pradeep Kumar Atrey. 2016b. Cyberbullying Detection Using Probabilistic Socio-Textual Information Fusion. In *Proc. IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. ACM.

[48] Andre Sourander, Anat Brunstein Klomek, Maria Ikonen, Jarna Lindroos, Terhi Luntamo, Merja Koskelainen, Terja Ristkari, and Hans Helenius. 2010. Psychosocial risk factors associated with cyberbullying among adolescents: A population-based study. *Archives of general psychiatry* 67, 7 (2010), 720–728.

[49] Claire Sutherland, Lynne Coventry, and Elizabeth Sillence. 2014. Using animated scenarios to explore severity of cyberbullying and reporting readiness. In *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: the Future of Design*. ACM, 448–451.

[50] Robert S Tokunaga. 2010a. Following you home from school: A critical review and synthesis of research on cyberbullying victimization. *Computers in human behavior* 26, 3 (2010), 277–287.

[51] Robert S Tokunaga. 2010b. Following you home from school: A critical review and synthesis of research on cyberbullying victimization. *Computers in human behavior* 26, 3 (2010), 277–287.

[52] Elizabeth Whittaker and Robin M Kowalski. 2015. Cyberbullying via social media. *Journal of School Violence* 14, 1 (2015), 11–29.

[53] Jun-Ming Xu, Kwang-Sung Jun, Xiaojin Zhu, and Amy Bellmore. 2012. Learning from bullying traces in social media. In *Proceedings of the 2012 conference of the North American chapter of the association for computational linguistics: Human language technologies*. Association for Computational Linguistics, 656–666.