Multimedia Tools and Applications (2019) 78:23973–24010
https://doi.org/10.1007/s11042-019-7234-z

Cyberbullying detection on social multimedia using soft computing techniques: a meta-analysis

Akshi Kumar [1] · Nitin Sachdeva [1]

Received: 26 September 2018 / Revised: 4 January 2019 / Accepted: 15 January 2019 /
Published online: 23 January 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

Cyberbullying is to bully someone in the digital realm. It has become extremely detrimental as the social media and the internet have become more popular and omnipresent. People use the internet services to viciously attack others from behind a screen. The substantial growth in the dimensionality, heterogeneity, subjectivity and multimodality of social media and the pressing need to timely curtail the damage instigated through cyberbullying, has fostered the need to devise automated mechanisms which detect such unfavorable activities. The use of soft computing techniques to handle such pernicious issue has been studied invariably and widely in literature. This study is to understand the viability, scope and significance of this alliance of using soft computing techniques for cyberbullying detection on social multimedia. This work is a systematic literature review to gather, explore, comprehend and analyze the research trends, gaps and prospects of this pairing in a well-organized way. The contribution of this study is noteworthy as it focuses on the use and application of soft computing techniques for cyberbullying detection on social multimedia utilizing a meta-analytic approach in order to integrate, interpret and critically analyze the findings in the original studies for expounding novel approaches to achieve comparable and effectual results pertaining to the defined research domain. Published studies starting April 2003, accessed from six digital portals (ACM, IEEE, Elsevier, Wiley, Springer and Taylor and Francis) have been reviewed to expound the state-of-art within the domain to give insightsand finally identify the directions of future research.

Keywords Cyberbullying · Social multimedia · Soft computing · Machine learning · Meta-analysis

✉ Akshi Kumar
akshikumar@dce.ac.in

Nitin Sachdeva
nits.usit@gmail.com

[1] Department of Computer Science & Engineering, Delhi Technological University, Delhi, India

1 Introduction

Information is power, but without a means to distribute information, people cannot harness this power. Social media has emerged as a key player which provides a platform for expression and distribution of content in today's world.The primary intent of social networking sites is to create, professional, interest, relationship-based virtual communities enabling stronger connections with everyone around the world. With the proliferation of these sites (Twitter, Tumblr, Google+, Facebook, Instagram, Snapchat, YouTube etc.), the user can post and share all kinds of multimedia content (text, image, audio, video) in the social setting using Internet without much knowledge about the Web's client-server architecture and network topology. The social networking sites have given 'everyone a voice' but at the same time, we're drowning in abundance, complexity of choices and unfortunately, the misappropriation or misdirection of influence. Moreover, when lots of individuals come together and that too from different countries, communities, races, ethnicities, gender, and varied age-groups, there are bound to be conflicts, controversies, and intimidation vulnerabilities. That is, although social networking sites proffer numerous benefits as these facilitate participation and collaboration but on the flip side hate speech, social distrust, cyberbullying, identity theft, cyber-stalking and cascading of rumors and fake stories are some antithetical concerns associated with it. The pervasive reach of these sites has irrefutably triggered, contributed and exacerbated bullying.

As per the National Bullying Prevention Center, *'Every child on Facebook likely has a bullying story, whether as the victim, bully or as a witness'* [35]. Cyberbullying is defined as bullying an individual or a group of individuals using Internet, mobiles or any other electronic device by sending inappropriate textual or non-textual multimedia message in order to hurt or cause embarrassment [52]. The one who bullies is called as 'bully' and the other is said to be 'victim'. The term 'Cyberbullying' was coined by Canadian educator and anti-bullying activist Bill Belseyin the year 2003 [10]. It is the repeated exposure of the negative actions on the part of one or more individuals in order to inflict humiliation, harassment, discomfort or injury upon another through the use of electronic medium [17] like emails, chat rooms, instant messaging, cell phones or by posting videos, audios, images etc. Bullying has been a part of the human civilization history which involves hurting someone either by humiliating or harassing in any form, involving mental, verbal or physical damage. When this assault takes place on cyberspace, it is referred to as cyberbullying/ cyberharassment/ cybervictimization [30].

According to a recent study, nearly 43% of the teenagers in the United States are victims of cyberbullying [53]. It is more persistent way of bullying an individual in front of the entire online community especially within the social setting which can eventually lead to psychological, mental and emotional breakdown for the victim inculcating the sense of low self-esteem, low self-confidence, anger, depression, stress, loneliness, sadness, health degradation etc. [49]. Many such intense cases have tragically ended in self-injury or suicides, underlining the grave nature of this critical issue [30]. The following Fig. 1 presents an example of cyberbullying from twitter.

With the technological advancement, the social freedoms that the networking sites give and larger audience cyberbullying has spread manifolds affecting the individual not only limited to their workplace but has also children and young adults in their daily lives. Anonymity further allows bullies to be more aggressive and offensive due to the reduced chance of being detected and punished, making it critical to efficiently detect cyberbullying behavior in a real- time setting. This poses significant threat to the physical and mental health of the victims making it

Fig. 1 Example of cyberbullying

a public health concern. Various studies have reported that victims of cyberbullying have lower self-esteem, higher levels of depression, suffer from behavioral issues and addict to substance abuse. Bullying victimization may trigger a sequence of events that results in suicidal behavior. The first reported case of cyberbullying was of an American middle school student, Ryan Halligan of Vermont in 2003 [26]. Ryan was constantly bullied in person and online by his classmates and this bullying was attributed as his reason to commit suicide.

Pertinent studies indicate that social media is one of the mostly favored medium by bullies and various factors such as socio-demography, physiological distress and time frames are related to cyberbullying. The massive volumes of human-centric, real-time, multimodal, heterogeneous and unstructured social media data makes manual detection intractable. Moreover, the social web applications/services are not restricted to the text-based data but extend to the partially unknown complex structures of image, audio and video. This fosters the need to develop intelligent tools and techniques for identifying, detecting and assessing cyberbullying from the available social multimedia data to lower down its hazardous impact. Design and development of contemporary tools which tap and analyze online detrimental behavior automatically from the high-dimensional social multimedia are imperative. Automated cyberbullying detection is typically a classification problem where the intent is to classify each abusive or offensive comment/ post/ message/ image as either a bullying or a non-bullying.

Studies have been conducted to explore new paradigms which handle uncertainty, imprecision, approximation, and fuzziness generated by thesocial multimedia content making automated cyberbullying detection techniques computationally proficient. In recent years, soft computing based multimedia analytics as a technology-based solution for automated detection of cyberbullying has attracted a lot of attention by both researchers and practitioners. Soft computing is a blanket term which comprises of variety of techniques such as machine learning, fuzzy logic, swarm and evolutionary computing, deep learning, amongst others. This work is a study to understand the feasibility, scope and relevance of this alliance of using soft computing techniques for cyberbullying detection on social multimedia portals. A systematic literature review on the use of soft computing techniques to detect cyberbullying automatically from the social multimedia is presented. Systematic literature review (SLR) is a detailed,structured, criterion-based and formal approach of gathering and reviewing literature studies pertaining to well-defined research questions on the selected domain. Meta-analysis summarizes the results of these studies to generate a meta-summary of the qualitative filtered studies. It is a statistical pooling to quantify the systematic review and eventually build an evidence pyramid. A systematic review does not necessitate a meta-analysis, but a meta-analysis (if used to answer a quantitative question) must be based on a systematic review. The following Fig. 2 illustrates the relationship of SLR and meta-analysis.

Thus, a SLR may contribute to a quantitative evidence synthesis known as meta-analysis. Meta-analysis identifies heterogeneity among multiple studies, identifies data gap in the

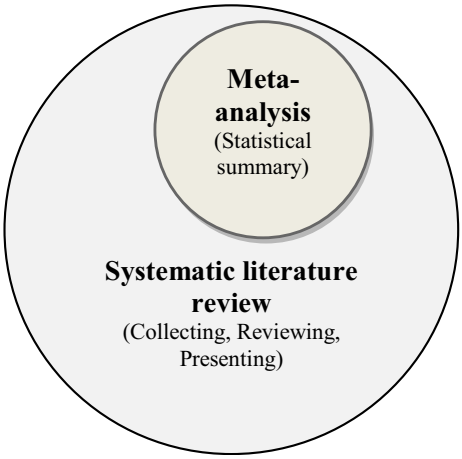

Fig. 2 Relationship of SLR and meta-analysis

knowledge base and at the same time suggest directions for future research. It is also referred to as research synthesis, pooled analysis, or quantitative review. In this work, the observational studies are thus meta-analyzed to provide quantitative answers to the specific, pre-defined questions. The goal is to gather empirical evidences and analyze results from existing studies to give a critically evaluated discussion on the existing trends in available research, identify gaps in current search and provide future prospects in the area by means of answering the established research questions.

A previously published systematic survey on cyberbullying [44] focused on the approaches to automated detection, namely, supervised learning, lexicon based, rule based and mixed-initiative approaches. The survey conducted in this work maps the state-of-the-art in cyberbullying detection research focused essentially on automated approaches using soft computing. Published studies within past 15 years, starting April 2003, accessed from six main digital portals (ACM, IEEE, Elsevier, Wiley, Taylor and Francis and Springer) have been reviewed. A quantitative, scientific synthesis of research results known as meta-analysis is done to finally summarize the results. The rest of the paper is organized as follows: Section 2 explicates the two primary concepts discussed in this review, which are cyberbullying and soft computing. Section 3 elaborates the review methodology enlisting the research questions identified to conduct this study review. Section 4 overviews the literature survey of the selected studies concisely. Finally section 5 provides the results and discussion followed by the conclusion in section 6 which enlists the gaps for potential research efforts.

2 Automated cyberbullying detection and soft computing

Netiquette refers to good manners on the Internet and treating other people on the Internet as you would like to be treated yourself. Unfortunately, some people use the Internet and/or mobile phones to offend or harass others. This is referred to as cyberbullying. Automated cyberbullying detection is a pro-active strategic technology based tool. It is a generic classification task where the multimodal, multimedia content is categorized as bullying or non-bullying. It is characterized as a predictive learning model in the social setting which detects the presence of cyberbullying in an online post (textual/non-textual) so that it does not inflict

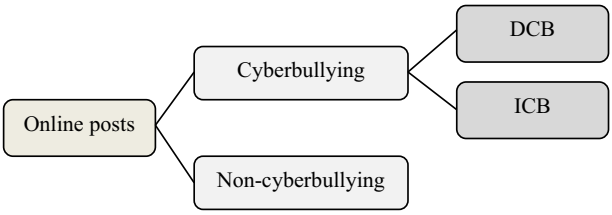

Fig. 3 Classification of social media posts

seriously or damage the victim's emotional, psychological and social state. The posts classified as bullying can further divided into two categories, namely, direct cyberbullying (DCB) and indirect cyberbullying (ICB) [45], as shown in Fig. 3. DCB involves direct sending of harmful content to a person either via email or SMS etc. ICB comprises of posting harmful contents about any person on social media or sharing it with others, for example posting an improper photograph of someone on Facebook is an example of ICB.

The direct and indirect cyberbullying is categorized into the different types as depicted in Fig. 4.

ICB messages are further categorized into six types [45]. These include Indirect Harassment, Denigration, Impersonation, Outing, Trickery and Exclusion, whereas DCB includes Flaming and Direct Harassment. The following Table 1 briefly explains these types of cyberbullying.

The elusive nature of cyberbullying undermines the self-esteem of the cyber victim, affecting him or her mentally, socially and psychologically. Automated detection model consists of multiple tasks which identify and classify posts as bullying or not. Considered as a generic classification problem, a typical cyberbullying detection process extracts the features from the pre-processed data and classifies the posts accordingly as shown in the Fig. 5 below.

The pre-processing phase includes cleaning the acquired data by removing unwanted URL's or strings etc., handling missing values, correcting words etc. and then transforming it in a representation suitable for feature extraction. After pre-processing, features such as keywords depicting bad/nasty/rude/abusive/hateful/attacking words, N-grams, pronouns, skip-grams are extracted. Next phase uses supervised learning techniques to classify the messages as either containing bullying content or not. This study comprehends the application of soft computing techniques to classify online bullying content. These techniques use approximate calculations to provide imprecise but usable solutions to complex computational problems. Also, referred to as computational intelligence techniques, soft computing techniques are generally divided into

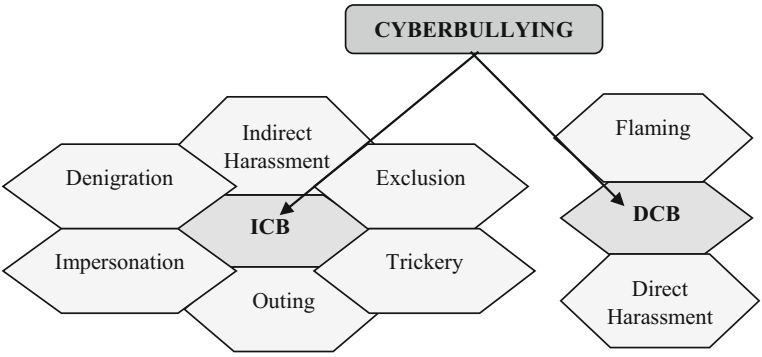

Fig. 4 Types of cyberbullying

Table 1 Types of cyberbullying

Types of cyberbullying	Details
Flaming	It is online fight between individuals. They usually exchange vulgar electronic messages [30, 45]. People fighting on online forums exchanging obscene message is an example of Flaming.
Direct harassment	It is directly harassing a person either by insulting or threatening him or her via messages [30, 45]. It includes only two parties- the one who bullies and the other who is bullied. Threatening or harassing a person either by sending email or SMS directly is an example of direct harassment.
Indirect harassment	It is indirectly harassing a person either by insulting or threatening him or her via messages posted online [30, 45]. It includes many parties. Posting embarrassing photos on social media in order to harass the other person indirectly is an example of indirect cyberbullying
Denigration	It is spreading hearsay or rumors about others in order to ruin their reputation [30, 45]. It puts the status of the cyber-victim on stake. Posting skewed contents on forums or blogs etc. in order to turn down cyber-victim's reputation is an example of denigration.
Impersonation	It is acting or pretending as other person and then doing anti-social activities in order to embarrass or damage his or her reputation [30, 45]. Imitating cyber-victim either by creating any fake profile or through hacking and sending messages that may instigate other users to attack the victim is an example of impersonation.
Outing	It is sharing private information of a cyber-victim without his or her consent in order to hurt the victim [30, 45]. Posting a humiliating picture of someone in order to hurt the cyber victim is an example of outing.
Trickery	It is obtaining sensitive information about a user by faking the trust of cyber victim and then eventually violating that trust [30, 45]. Obtaining a personal video by faking as close friend and then posting it online is an example of trickery.
Exclusion	It involves the exclusion of the cyber victim from online communities or groups etc. [30, 45]. Excluding a person knowingly from WhatsApp group is an example of exclusion.

following categories *(Machine learning (ML), Neural networks (NN), evolutionary computation, fuzzy logic and probabilistic reasoning)* [1] as depicted in the following Fig. 6.

The machine learning techniques are further categorized as shown in the Fig. 7 below.

From Figs. 6 and 7, it is apparent that soft computing is a *'blanket term'* comprising of several methodologies which are themselves inter-related to one other. Deep learning (DL) is considered as a part-of a broader family of ML [1] based on learning data representations, in contrast to the task-specific algorithms and where learning can be supervised, semi-supervised or unsupervised. DL consists of various techniques such as deep NN (DNN), recursive NN, recurrent NN, convolutional NN and deep belief networks, whereas neural networks is also an established sub-category of soft computing (SC) techniques [1] which includes feed forward; multi-layer perceptron; radial basis; Kohonen self-organizing; modular; shallow and deep NN (DNN). Thus, it can be inferred that SC, ML and DL are inter-connected to each other as shown in the following Fig. 8:

Social media is inherently an informal way of communication with all kinds of multimedia content. The following Fig. 9 depicts the multimedia types supported by popular social networking sites.

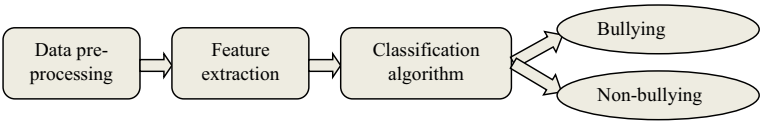

Fig. 5 Generic cyberbullying detection process

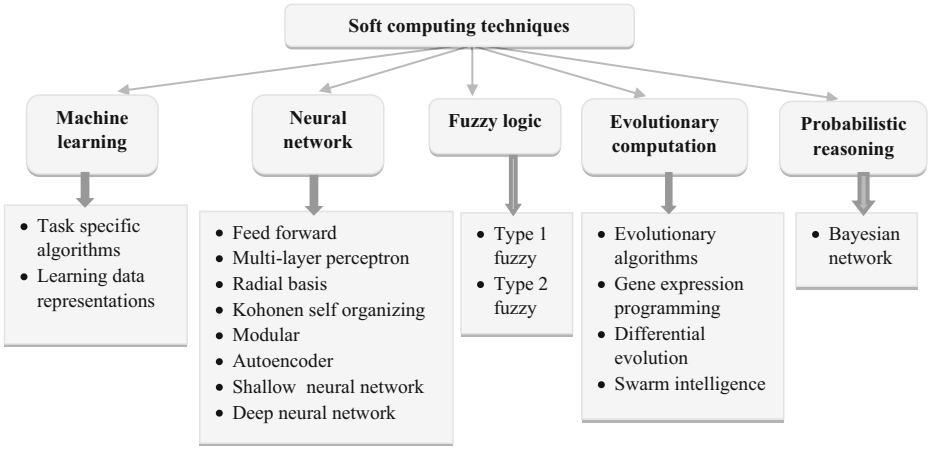

Fig. 6 Soft computing techniques

Social media dynamics keep changing with respect to increasing user base and user-activity which makes it a high -dimensional, complex and fuzzy data space for analytical processing. Soft computing has a solution to many real world problems. The guiding principle of soft computing is to exploit the tolerance for imprecision, uncertainty and partial truth to achieve robustness, low cost solutions [1]. The alliance of these two domains conceives the necessary balance and compels investigation regarding the feasibility, trends and scope of using soft computing techniques for cyberbullying detection on the social multimedia data. Recent trends demonstrating the use of soft computing techniques to automate cyberbullying detection has been observed. This makes it necessary to document the preliminary work and review the ongoing work within this domain. The next section describes the process of review adopted for this systematic study.

3 Review process

A systematic literature review (SLR) intends to identify, critically assess and integrate the findings of all pertinent, high quality primary studies addressing specific research questions pertaining to the research domain. This SLR was planned and conducted based on the format put forward by Ketchenham and Charters in 2007 [28]. The overall review process was categorized into six phases as per the guidelines given. The following Fig. 10 depicts the process of SLR.

The first phase ascertains and formulates the research questions (RQ's) within the domain identified for the survey. The next phase, that is, the '*Search strategy'* is designed to recognize and locate the relevant research studies addressing the defined research questions. In the next phase, *'Study selection',* the scope of the study is constricted by using a selection criterion known as '*Inclusion-Exclusion'* criteria. This phase yields the number of the relevant studies that can be included within the specified domain. Next phase is called as '*Quality assessment*' that evaluates the worthiness of the selected studies. The goal is to ensure the quality and similarity of included studies, and clearly define the boundaries of the review. Post this phase is the *'Data extraction'* phase which extracts the relevant and required data in order to answer the specified research questions. It produces a summarized critique to evaluate, extend and/or identify gaps and incosistencies, if any and provide directions for future research. The last *'Data synthesis'* phase summarizes the results of these qualitative filtered studies using

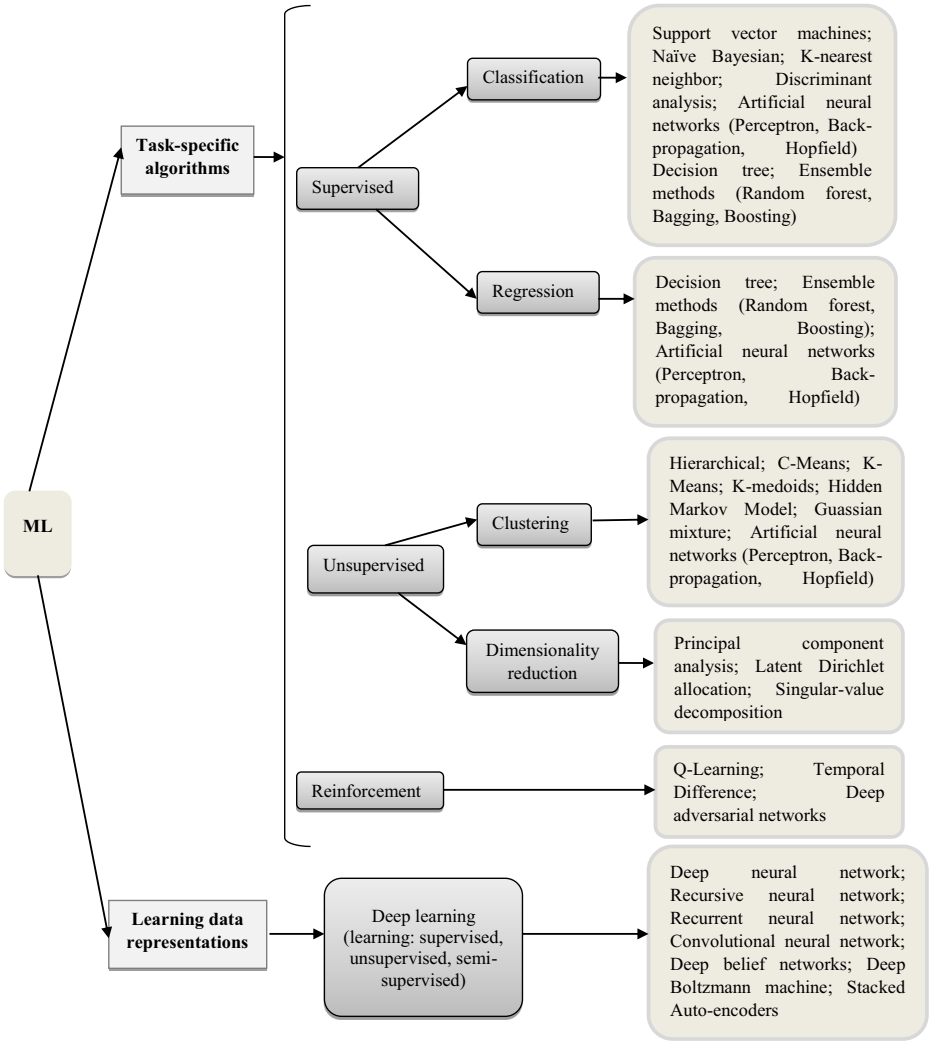

Fig. 7 Machine learning techniques

visualization tools such as graphs, charts and meta-summary tablesThe following sub-sections identify the relevant RQ's which this SLR intends to answer followed by the details of selection and examination of the selected relevant studies.

3.1 Research questions

The following RQs were identified:

- RQ1: Which are the most distinguished and relevant conferences and journals with published studies?
- RQ2: On which datasets and domains the studies using soft computing techniques for cyberbullying detection on social multimedia have been conducted?

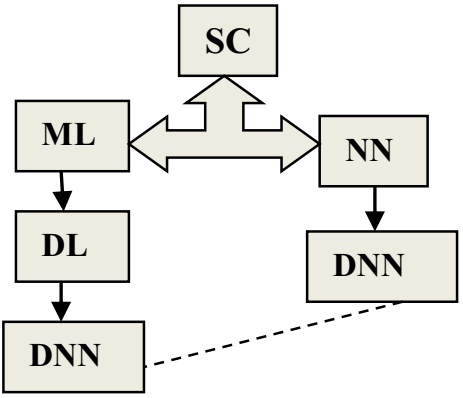

Fig. 8 Relation between SC, ML and DL

- RQ3: Which is the most frequently used soft computing techniques used for cyberbullying detection on social multimedia?
- RQ4: What are the widely applied key performance indicators that are used for evaluating the applied techniques?

3.2 Search strategy

An adept strategy of extensive search (starting the first reported study in April 2003-till date), was set up of in order to extract as many as potential relevant research studies which expound the use of soft computing techniques in cyberbullying detection. This phase fragmented the selected RQ's into individual concepts in order to create the '*search terms*' which could further be searched in the identified databases/e-portals/digital libraries/digital portals. The search terms were identified for this SLR were: cyberbullying, soft computing techniques, machine learning, supervised, unsupervised and social multimedia and were explored in titles, keywords and abstracts of studies to extract all related primary research studies from journals of high repute and highest relevance to the topic of study, available within six prominent digital

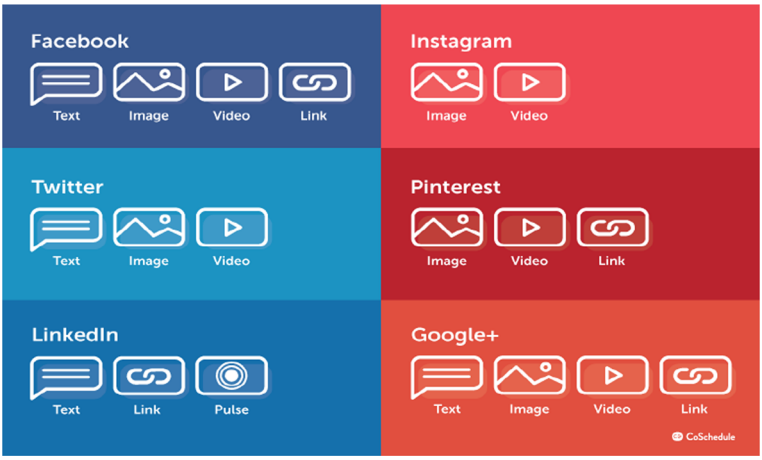

Fig. 9 Multimedia support by popular social networking sites [27]

Fig. 10 Process of SLR

libraries (publishers), namely, ACM, IEEE, Elsevier, Wiley, Springer and Taylor and Francis. The grammatical variations of these terms such as synonyms etc. were also used to search exhaustively. Boolean expressions (or/and) were further used for expanding or narrowing the sweep of the search to filter potentially relevant papers. The reference section of the relevant studies was also examined to extract cross-citations. Thus, the purpose of this step was to identify, select and extract the relevant set of research papers for conducting review.

3.3 Study selection

This phase involved isolating the irrelevant, redundant and non-pertinent studies based on the 'Exclusion-Inclusion' selection criteria. It performed the filtering procedure by either selecting or rejecting the studies which facilitate or directly answer at least one RQ within the selected problem domain. The following Inclusion-Exclusion Criteria was adopted:

Inclusion criteria:

- Studies published in the last 15 years i.e. from April 2003-till date (Sept'2018).
- Studies representative of cyberbullying detection on social multimedia.
- Studies focusing on the application of unsupervised and supervised learning algorithms.
- Studies focusing on the application of supervised machine learning (ML) algorithms like Decision Tree (DT), Support Vector Machine (SVM), k Nearest Neighbor (kNN), Random Forests (RF), Linear Regression (LR), Logistic Regression (LogR), Boosting (Bos), Bagging (Bgg), Adaboost (Adb), Multiple Regression (MR), Maximum Entropy (MaxE) etc. for detecting cyberbullying on social multimedia.
- Studies with unsupervised machine learning algorithms in soft computing such as K-Means Clustering (KMC), C-Means Clustering (CMC), Hierarchical Agglomerative Clustering (HAC) etc.
- Studies including soft computing techniques such as Probabilistic Reasoning which includes Naïve Bayesian (NB) or Bayes Network (BN), Neural Networks (NN), Fuzzy logic (FL), Evolutionary Computing (EC) for cyberbullying detection on social multimedia.
- Studies representing the application of deep learning (DL) techniques like Convolutional Neural Network (CNN) etc. for cyberbullying detection on social multimedia.
- Studies with hybrids of soft computing techniques for detecting cyberbullying on social multimedia.
- Studies involving the comparison of above mentioned techniques.
- Studies involving detecting cyberbullying on social media in English language only.

- Studies involving cyberbullying detection in multimedia like images, texts, videos etc.

Exclusion criteria:

- Studies which are without proper empirical analysis or benchmark comparisons.
- Studies that are purely reviews or surveys or theoretical concepts on cyberbullying detection without any implementations.
- Studies on languages other than English (for example Dutch, Portuguese, Latin, Chinese, Arab, Spanish etc.) and multilingual cyberbullying detection (such as mash-up languages i.e. mixed usage of different languages, for example, Hinglish is a mixture of English and Indian Hindi language).

3.4 Quality assessment

In order to maintain the quality standard of the selected studies a careful consideration had been affirmed by taking the novelty of technique proposed and the technical content (data set and evaluation methods used). The quality check had been imposed in order to evaluate the worthiness, significance and strength of the selected studies based on various w eighing parameters, as discussed next:

- ***Novelty***: to judge whether the proposed technique is a novel one or just an enhancement or improvement over an existing one
- ***Technically content***: to discover the real and clear motivation behind the proposed technique. Also to find whether the scope and limitation of the proposed technique is evident and unambiguous.
- ***Result and analysis***: to assess whether the proposed technique is tested on a standard benchmark data set or a random data set,with proper evaluation of efficacy measures and compared with existing techniques.
- ***Publication***: to identify whether the selected study belongs to a conference or a high impact journal and the number of the citations that the study has. Although not much weightage has been given to this parameter as a recent study may not have many citations.

Thus, each selected study was evaluated out of 10 and scored on the following basis: 2 for novelty, 1.5 for publisher, 5 for results and analysis in which 2 was for data set, 2 for evaluation criteria used and 1 for the comparison with any of the existing technique and the rest 1.5 for technical writing. This qualitative assessment is given in the following Table 2:

3.5 Data extraction

In this phase, finally the key information was extracted from the selected studies and was summarized based on the mapping of the selected study to one or more RQ's. The information acquired from the extracted research studies was: details like authors, year of publication, datasets used, techniques applied, domains targeted, social media which was used for analysis, textual or multimedia content on which the techniques were implemented, type of cross validation that was used, key performance indicators that were employed for evaluation of the techniques, followed by remarks. All this information was then stored in a table for data synthesis.

Table 2 Quality assessment

Quality level	Number of studies	Percentage
Outstanding (9.5 < score < =1)	4	7.69
Excellent (8 < score < = 9.5)	9	17.31
Good (7 < score < =8)	10	19.23
Average (5.5 < score < =7)	18	34.61
BelowAverage(4 < score < =5.5)	6	11.54
Poor (score < =4)	5	9.62

3.6 Data synthesis

The goal of the '*Data Synthesis*' phase is to summarize and interpret the extracted information in order to finally output the result of the SLR as direct answers to the identified RQs using critical analysis, discussions and different visual representations such as tables, graphs, charts, etc. To filter the most significant and qualitative work, search and study selection procedures were carried twice. Meta-analysis then quantitatively combined results of the studies in the SLR. It collated data to generate statistically significant resultsand summaries from the pooled set of relevant studies. Identified search terms were input as search query inthe six selected digital libraries which resulted in 320 papers. After removing redundant studies, 255 studies were obtained on which the inclusion and exclusion criteria was applied. Fifty-two potentially relevant studies were then filtered for further qualtitative analysis out of which 47 high quality studies eventually formed the basis of this review. Figure 11 depicts the overall search process applied in order to fetch the most relevant studies.

4 Literature survey

The review of the final set of studies identified for this SLR which demonstrate the use of soft computing techniques for cyberbullying detection on social multimedia is given in Table 3. As

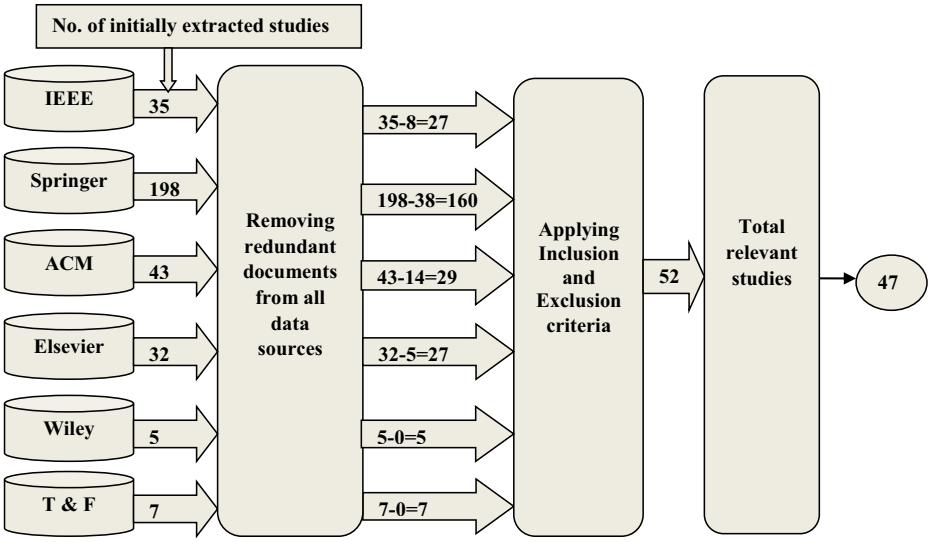

Fig. 11 Systematic literature review procedure

Table 3 Summary details of the studies selected for this review

S.No.	Author	Publication	Year	Techniques	Social Media	Textual or multimedia	Data set
1.	Reynolds et al. [43]	International Conference on Machine Learning and Applications, IEEE	2011	DT, kNN, SVM	Formspring.me	Texts	Author collected data from the Formspring.me and extracted the information from the sites of 18,554 users, containing 2696 posts for training and 1219 for testing purposes.
2.	Nahar et al. [32]	Asia-Pacific Web Conference. Springer, Berlin, Heidelberg	2012	BoW, PLSA, Bayes method, SVM	Kongregate, Slashdot and MySpace	Texts	Data taken from the workshop on Content Analysis (http://caw2.barcelonamedia.org/).
3.	Xu et al. [52]	Conference of the North American chapter of the association for computational linguistics: Human language technologies, ACM	2012	NB, L-SVM (linear), R-SVM (RBF) and LogR, CRF	Twitter	Texts	Data taken from uniformly sampled 990tweets for manual inspection by five experienced annotators.
4.	Kontostathis et al. [30]	Proceedings of the 5th annual acm web science conference. ACM.	2013	BoW, EDLS, tf-idf	Formspring.me	Texts	Author collected data from the Formspring.me and contained 13,652 posts. The data has been labelled using AMT.
5.	Dadvar et al. [17]	European Conference on Information Retrieval, Springer, Berlin, Heidelberg	2013	SVM	YouTube	Texts	Author collected 4626 comments from 3858 distinct users
6.	Sheeba et al. [48]	International Conference on Computational Intelligence and Computing Research, IEEE.	2013	FL, MaxE, CMC, Fuzzy C means, Fuzzy DT	Twitter, Facebook	Texts, audios	Random
7.	Nahar et al. [33]	In Australasian Database Conference Springer, Cham	2014	Naive Bayes multinomial, Stochastic Gradient Descent, RF, LogR, Fuzzy SVM (FSVM), Kernel-based Fuzzy C-Means (K-FCM) clustering	Myspace, Kongregate, and Slashdot.	Texts	Data provided by Fundacion Barcelona Media for the workshop on content analysis (http://caw2.barcelonamedia.org/).
8.	Parime&Suri [37]	International Conference on Circuit, Power and	2014	SVM	Myspace	Texts	Myspace

Table 3 (continued)

S.No.	Author	Publication	Year	Techniques	Social Media	Textual or multimedia	Data set
		Computing Technologies [ICCPCT], IEEE					
9.	Dadvar et al. [18]	Canadian Conference on Artificial Intelligence, Springer, Cham	2014	NB, DT, SVM, MCES (Expert System)	YouTube	Texts	Author collected 54,050 comments from 3825 distinct users
10.	Michalopoulos et al. [31]	Computers & security, Elsevier.	2014	FL, NB, kNN, MaxE, SVM	Perverted-justice website	Texts	Random
11.	Holt et al. [24]	Journal of Criminal Justice, Elsevier	2014	LogR	–	Texts	Author collected a self-administered questionnaire for 6th to 12th grade students in 14 middle and high schools in the Iredell-Statesville School System (ISS) in North Carolina. 1972 students had completed the survey.
12.	Byrne et al. [9]	Journal of Computer-Mediated Communication, Wiley	2014	LogR	General survey conducted	Texts	C + R Research, a professional research firm in Chicago collected data from parents and children. The data was related to the survey involving questions on cyberbullying.
13.	Rafiq et al. [39]	International Conference on Advances in Social Networks Analysis and Mining, ACM	2015	Snowball sampling method, NB, Adb, DT, RF	Vine	Videos	Author collected 652 K media sessions from Vine that contained information such as user id, profile information, videos posted by a user, post id's etc.
14.	Chavan and Shylaja [14]	Advances in computing, communications and informatics (ICACCI), International Conference, IEEE	2015	SVM, LogR	Kaggle	Texts	Authorcollected 2647 comments

Table 3 (continued)

S.No.	Author	Publication	Year	Techniques	Social Media	Textual or multimedia	Data set
15.	Balci and Salah [6]	Computers in Human Behavior, Elsevier	2015	DT, SVM, KMC, Bayes Point Machine (BPM)	Okey (CCSoftOkey Player Abuse (COPA) Database)	Texts	Author gathered 800,000 Okey games along with the player interactions in the chat area over a period of six months.
16.	Nandhini and Sheeba [34]	International Conference on Advanced Computing Technologies and Applications (special issue of Procedia Computer Science), Elsevier	2015	NB, FL, FuzGen (hybrid of Fuzzy Logic and GA)	Formspring.me and MySpace	Texts	–
17.	Balakrishnan V [5]	Computers in Human Behavior, Elsevier	2015	LogR	Facebook	Texts	Author prepared a questionnaire of 393 participants consisting of questions related to cyberbullying involving victims and perpetrators, sexting (i.e. sharing sexually suggestive photos or messages through mobile phones and other mobile media), and personalities (i.e. questions related to their overall self-esteem and family).
18.	Zhang Et al. [55]	International Conference on Machine Learning and Applications, IEEE	2016	PCNN, CNN	Twitter and Formspring.me	Texts	Author collected 1313 messages from Twitter and 13,000 messages were Collected from Formspring.me and labeled by a web service called Amazon Mechanical Turk
19.	Zhao et al. [57]	Proceedings of the 17th international conference on distributed computing and networking. ACM.	2016	Continuous Bag of Words, Semantic-enhanced BoW Model, Embeddings-enhanced Bag-of-Words (EBoW), Latent Dirchilet Allocation (LDA),	Twitter	Texts	Author collected 1762 random tweets as on 6 Aug 2011.

Table 3 (continued)

S.No.	Author	Publication	Year	Techniques	Social Media	Textual or multimedia	Data set
				tf-idf, Latent Semantic Analysis (LSA), word embeddings, SVM			
20.	Hosseinmardi et al. [25]	International Conference on Advances in Social Networks Analysis and Mining (ASONAM), IEEE	2016	Snowball sampling method, backward feature selection approach, LogR, ridge regression classifier	Instagram	Images	Author collected data from around 41 K Instagram user ids using a snowball sampling method via the Instagram API.
21.	Gordeev [20]	International Conference on Speech and Computer, Springer, Cham	2016	RF, CNN-non-static, CNN (POS)	Imageboards (4chan.org, 2ch.hk)	Texts	Author collected 1000 English messages
22.	Hammer [23]	International Conference on Industrial Networks and Intelligent SystemsSpringer, Cham.	2016	Logistic LASSO regression	YouTube	Texts	Author collected a total of 24,840 sentences where 1469 sentences were violent
23.	Gordeev [21]	Procedia-Social and Behavioral Sciences, Elsevier	2016	word2vec, NN, RF	Imageboards (4chan.org, 2ch.hk)	Texts	Author collected 1000 English messages
24.	Al-garadi et al. [3]	Computers in Human Behavior, Elsevier	2016	synthetic minority oversampling technique (SMOTE), NB, SVM RF, and kNN	Twitter	Texts	Author collected data between January 2015 and February 2015 and contain 2.5 million geo-tagged tweets. Author randomly selected 10,606 tweets from collected data.
25.	Potha et al. [38]	Knowledge-Based Systems, Elsevier	2016	HAC, Bayesian hierarchical clustering, SVM	Perverted-Justice (PJ)	Texts	–
26.	Rafiq et al. [40]	Social Network Analysis and Mining, Springer	2016	LDA, Adb, DT, RF, Extra tree classifier, SVM (SVM Linear, SVM Polynomial, SVM rbf (radial basis function), SVM Sigmoid), kNN, NB, MLP, LogR	Vine	Videos	Author collected Vine information from 59,560 users about 652 K media sessions.
27.	Papegnies et al. [36]	International Conference on Statistical Language and	2017	Bag of words, TF-IDF, Probability of n gram emission, Context	MMO (online game chat systems)	Texts	Author accessed database containing 4, 029, 343 messages

Table 3 (continued)

S.No.	Author	Publication	Year	Techniques	Social Media	Textual or multimedia	Data set
		Speech Processing. Springer, Cham		based (SVM) and graph based classifier			where 779 messages were flagged by one or more users as abusive and 1558 as non-abusive. Total 2000 random messages were fetched.
28.	Sedano et al. [46]	International Conference on Artificial Intelligence and Soft Computing. Springer, Cham	2017	SVM, FL	Twitter	Texts	Author collected 18,504 tweets from June to December 2016
29.	Thu and New [50]	International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), IEEE	2017	SVM	Twitter, Amazon	Texts	Author collected 674 real news-wire articles and 226 satire newswire articles. Also, they had gathered and pre-processed around 20 K-30 K tweets for each satire and non-satire corpus. Publicly available dataset was also used (https://rpubs.com/pm0kjp/satire_serious_reviews).
30.	Zhao and Mao [56]	IEEE transactions on affective computing	2017	Semantic-enhanced Marginalized Stacked DenoisingAutoencoder - smSDA (deep learning method), mSDA, SVM, LSA, LDA, BoW, sBo W, BWM (Bullying word matching), word embeddings	Twitter and MySpace	Texts	Author collected 7321 tweets from Twitter from 6 Aug 2011 to 31 Aug 2011 and 1539 data samples from MySpace.
31.	Raisi and Huang [41]	International Conference on Advances in Social Networks Analysis and Mining, ACM	2017	Participant-Vocabulary Consistency (PVC) using Alternating Least Squares, snowball sampling	Ask.fm, Instagram, and Twitter	Texts	Author collected 296,308 tweets from Twitter from 1 Nov 2015 to 14 Dec 2015. Author had also gathered 2,863,801 question-answer pairs from Ask.fm and 9,828,760 messages

Table 3 (continued)

S.No.	Author	Publication	Year	Techniques	Social Media	Textual or multimedia	Data set
							from Instagram. For labelling the data, a web service called Amazon Mechanical Turk (AMT) was used.
32.	Chatzakou et al. [11]	Conference on Hypertext and Social Media, ACM	2017	K means, Expectation-maximization algorithm, RF	Twitter	Texts	Author collected 650 K tweets about the Gamergate (GG) controversy. Author had also gathered a baseline dataset with 1 M random tweets.
33.	Chatzakou et al. [12]	International Conference on World Wide Web Companion, ACM	2017	RF	Twitter	Texts	Author collected 650 K tweets on hate related topics and around 1 M baseline tweets random from Jun to Aug 2016.
34.	chatzakou et al. [13]	Proceedings of the 2017 ACM on Web Science Conference. ACM	2017	NB, DT, RF, NN, CBoW	Twitter	Texts	Author collected 650 K tweets on hate related topics and around 1 M baseline tweets random from Jun to Aug 2016.
35.	Garc'ıa--Recuero [19]	International Conference on Advances in Social Networks Analysis and Mining, ACM	2017	MinHashes, DT, RF, Extra Trees, Gradient Boosting, Adb and SVM, Voting	Twitter	Text	Author collected data from 163 trusted humans that provided 14,193 annotations.
36.	Ashktorab et al. [4]	Web Science Conference, ACM	2017	Snowball sampling, Latent Dirichlet Allocation topic modelling, NB	Ask.fm	Texts	Author searched for the key terms that are generally associated with cyberbullying. From Ask.fm and had then fetched random user profile information from it.
37.	Bourgonje et al. [7]	International Conference of the German Society for Computational Linguistics and Language Technology, Springer, Cham	2017	Bayes, Bayes expectation maximization, DT, Multivariate LogR, MaxE, Winnow2, BoW	Twitter, Wikipedia Talk pages	Texts	Author gathered data from Twitter containing 15,979 tweets and from Wikipedia Talk pages containing 11,304 annotated Comments.
38.	Haidar et al. [22]	Cyber Security in Networking Conference IEEE	2017	NB, SVM	Twitter	Texts	Author collected 91,431 tweets

Table 3 (continued)

S.No.	Author	Publication	Year	Techniques	Social Media	Textual or multimedia	Data set
39.	Wint et al. [51]	Digital Information Management (ICDIM), Twelfth International Conference, IEEE	2017	CNN, NB	Twitter, Formspring.me	Texts	Author collected 1,578,627 tweets from Twitter and 18,554 users from Formspring.me
40.	Sarna& Bhatia [45]	International Journal of Machine Learning and Cybernetics,Springer	2017	NB, kNN, DT, SVM	Twitter	Texts	Author collected random tweets via customized crawler written in Python.
41.	Rakib and Soon [42]	Asian Conference on Intelligent Information and Database Systems, Springer, Cham	2018	RF	Reddit, kaggle	Texts	Author collected 6594 raw comments.
42.	Agrawal and Awekar [2]	European Conference on Information Retrieval,ECIR, Springer, Cham	2018	CNN, LSTM, BLSTM, BLSTM with attention, logistic regression, SVM, RF, NB	Formspring, Twitter, Wikipedia	Texts	Author collected 12 k posts for FormSpring, 16 k for Twitter and 100 k for Wikipedia
43.	Chen et al. [15]	Neural Computing and Applications, Springer	2018	SVM, Logistic regression, LSTM+2D TF-IDF, CNN +2D TF-IDF, LSTM+ EMBEDDING, CNN + EMBEDDING	Twitter	Texts	Author collected random aggressive comments from Twitter.
44.	Koban et al. [29]	Computers in Human Behavior, Elsevier	2018	MR	Facebook	Texts	Author conducted an online survey of 256 participants concerning personality dispositions, participants' Internet and Facebook usage, as well as single-item measures of their interest and their level of expertise in four different news subjects (i.e., politics, sport, social issues, and terrorism).
45.	Coletto et al. [16]	World Wide Web Conference 2018,ACM	2018	LogR	Twitter	Texts	Author used twitter datasets. The first one contained 977 k English tweets from Dec 2008 to Jan 2009 and the other contained

Table 3 (continued)

S.No.	Author	Publication	Year	Techniques	Social Media	Textual or multimedia	Data set
							1 M English tweets collected in Dec 2015.
46.	Sharma et al. [47]	International Conference on Advances in Computing and Communication Engineering (ICACCE) 2018, IEEE	2018	LogR, SVM, RF, Gradient Boost	Kaggle	Texts	Author collected data that contained2235 samples.
47.	Bu et al. [8]	International Conference on Hybrid Artificial Intelligence Systems, Springer	2018	CNN, LSTM	Kaggle	Texts	Author collected data that contained 8815 comments. Amongst all, 2818 comments were labeled as cyberbullying.

S.No.	Domain	Tools	CV	KPI	Remarks
1.	Random	Weka	10	A	DT outperformed the other classifier with A of around 79%.
2.	Games	LibSVM	10	A	Feature selection had shown improved accuracy. SVM performed best with 500 features.
3.	Random	WEKA	5	A, P, R, F	Key problems have been identified in using social multimedia data and formulate them as NLP tasks, including text classification, role labelling, sentiment analysis, and topic modelling.
4.	Random	R	2	P, R, TP	Author proposed a model that resulted in better results for CD.
5.	Movies	–	10	P, R, F	Author observed that incorporation of context in the form of User's activity history improves CD accuracy.
6.	Meeting transcripts	–	–	A	Author obtained improved CD results using FL techniques.
7.	Random	Weka	10	P, R, F	Author proposed a semi-supervised method that had shown improved results as compared to traditional methods for CD.

Table 3 (continued)

S.No.	Domain	Tools	CV	KPI	Remarks
8.	Random	RapidMiner,	10	Cn	Author had taken into account the psychological factors related to cyberbullying for identifying the absence and presence of abusive content.
9.	Random	WEKA 3	10	AUC	Author found that Naive Bayes outperformed the other two algorithms.
10.	Romantic movies, chats	Rainbow (front-to-end document classification tool that performs instant classification of a given text)	10	A, FN, FP	Author developed a 'Grooming Attack Recognition System' for real-time identification, assessment and control of cyber grooming attacks in favor of child protection.
11.	Adolescent problem behavior	–	–	M, SD, Rg, Cr	Author discussed the implications of various demographic factors for policy responses to bullying victimization.
12.	Adolescents problems, pornography or sexual imagery	–	–	A	Author obtained accuracy of more than 70% for the cases where kids are cyberbullied by others.
13.	Well known celebrities	CrowdFlower, NLTK toolkit	10	A, P, R	Amongst all, Adb had obtained the highest accuracy of around 76%.
14.	Random	–	–	P, R, AUC,ACC	Author proposed that the suggested hypothesis increase the accuracy by 4%.
15.	Player demographics, statistics, game records, interactions and complaints	Infer.NET library	10	TP, TN, FP, FN, P, Sn, Sp.	Author proposed a model for assessing different types of features for detecting abuse automatically.
16.	–	–	–	P, A, R, F	Author proposed a hybrid approach for CD where GA is used for optimizing the parameters and to obtain precise output and FL has been used to retrieve relevant data for classification from the input.
17.	Sexting, self-esteem, family.	–	–	MW, KW	Author claimed that the proposed model for cyberbullying was significant where age and gender were found to be insignificant predictors for Cyber-victims and cyberbullies.
18.	Random	eSpeak, Theano package, Python	5 and 10	A, R, P, F, TP, FP, TN, FN	Author proposed a novel PCNN model for CD and the results show that the novel approach had outperformed the existing methods in terms of accuracy.

Table 3 (continued)

S.No.	Domain	Tools	CV	KPI	Remarks
19.	Random	Word2vec	5	P, R, F	Author proposed a novel learning method called as EBoW for CD that yielded enhanced results.
20.	Elite users (famous personalities, like actors, singers etc	–	5	P, R, F, FP, ROC, AUC	Author achieved high performance in predicting cyberbullying using the proposed approach.
21.	Random	–	10	F	Author observed that Random Forest classifier surpassed CNN for the task of detecting aggression for the English language
22.	Religious and political	R (Glmnet package)	–	MSE	Author proposed a method which can automatically detect threats of violence using machine learning.
23.	Random	NLTK-toolkit	10	A	Author proposed a method that detected automatic aggression with 88% accuracy.
24.	Random	WEKA	10	P, R, AUC, F	The results exhibited that the Random forest using SMOTE alone showed the best AUC (0.943) and f-measure (0.936).
25.	Sexual conversations	Matlab	–	R	Author proposed clustering based method for extracting patterns in sexual cyberbullying data and had shown improved results.
26.	Public or user profiles	CrowdFlower	10	A, P, R	RF yielded best A, P and R of more than 85%
27.	Games	Python-iGraph, Sklearn (SVC (C-Support Vector Classification) toolkit	10	P, R, F	Author presented an approach based on graph features for automatically detecting online abuse.
28.	School students and staff members	Java Swing, Twitter streaming API, Tweepy, Mysql database, PHP	–	A	Author presented a model where the output of SVM is fed as input to Fuzzy Logic for identifying the bullying severity.
29.	News-articles and Amazon Products Review s	Tweepy (Python library), SÉANCE (Sentiment analysis and social cognition engine) text analysis tool, Bag of words, term frequency-inverse document frequency (TFIDF), term	10	P, R, F, A	Author proposed an approach for detecting satirical languages in both short text (tweets) and long text (newswire articles and product reviews). Author had shown that the model with supervised weighting TFRF worked better in long text whereas the model with unsupervised weighting TFIDF worked better for short text.

Table 3 (continued)

S.No.	Domain	Tools	CV	KPI	Remarks
		frequency relevance frequency (TFRF), SenticNet			
30.	Random	word2vec	10	A, F	Author proposed an approach called as smSDA for text based CD that showed improved results.
31.	Random	scipy.sparse	3	P	Author proposed a weakly supervised PVC model for CD that had shown enhanced results.
32.	Gamergate Controversy [hate speech] and random	SentiStrength tool	–	P, R, ROC	Author proposed an unsupervised machine learning analysis for better understanding the behaviors of abusive users on social multimedia alike Twitter.
33.	Random and hate related (Gamergate controversy) [hate speech]	–	10	P, R, CK, ROC	The study depicted that the author's approach had produced promising results with high accuracy for detecting aggressive and bully users on Twitter.
34.	Random and hate related (Gamergate controversy) [hate speech]	CrowdFlower, SentiStrength Tool, WEKA	10	P, R, ROC, CK, RMSE	Author proposed a robust approach for understanding the properties of bullies and aggressors on Twitter and improved results were achieved.
35.	Elite users	Crowdflower, Trollslayer	5	P, R, F	MinHashes obtained better abuse detection rates with supervised learning and also minimized the amount of computation as well.
36.	Random	–	–	CK, P, R, F	The proposed approach encouraged the performance of the classifier reasonably for accurate automatic detection of different discourse categories.
37.	Random	McCallum 2002	10	A, P, R, F	Author observed that logistic regression implementation, using word unigrams, outperformed the best scoring feature set in Twitter dataset
38.	Random	WEKA, PHP, python, mongoDB, SentiStrength	–	P, R, F, TP, FP	Author presented a solution for detecting and stopping cyberbullying using SVM and Naive Bayes
39.	Random	–	–	A	Author measured accuracy on collected datasets using CNN and Naïve Bayes.
40.	Random	MATLAB, Python, Part-Of Speech (POS)	–	P, R, F	Author had shown that less users are involved in indirect cyberbullying than direct cyberbullying.
41.	Random	MongoDB, word2vec,Python, GloVe	–	AUC, P	Author depicted that the presented model had2% improvement of precision over the next best score.
42.	Random	GloVe, SSWE	5	P, R, F	

Table 3 (continued)

S.No.	Domain	Tools	CV	KPI	Remarks
					This study analyzedcyberbullying detection on various topics across multiple SMPs using deep learning based models and transfer learning.
43.	Random	Word2Vec, Glove	–	A, Micro-AUC, Macro-AUC	Author achieved improvement in convolutional neural networks (CNN) using 2-dimensional tf-idffeatures.
44.	Personality, politics, sport, social issues, and Terrorism.	–	–	M, SD, Cr	Author presented a study that examined participants' intention to comment in an uncivil manner that typically hinders a productive public discussion.
45.	Random	Face++	–	P, R	The results showed that aggressive users smile less. Also, they appeared not happy in their profile pictures,
46.	Random	Python	–	A, P, R AUC	Results depicted that Logistic Regression and Random Forest performed better than SVM andGradient Boosting.
47.	Random	–	10	ROC, AUC,	Author proposed hybrid architecture of character-level CNN and word-level LSTMthat outperformed other machine learning methods.

discussed in the data extraction phase, the information extracted from the selected studies included details about the authors, publication, year of publication, techniques applied, social media which was selected, type of multimedia on which SC techniques were applied, domains targeted, tools which were used, type of cross validation (CV) used, key performance indicator (KPI) [Accuracy (A), Precision (P), Recall (R), F score/ F1 score/ F1 measure/ F measure (F), Confidence (Cf), Sensitivity (Sn), Specificity (Sp), True Positives (TP), True Negatives (TN), False Positives (FP), False Negatives (FN), ROC, AUC, Cohen's Kappa measure (CK), Root Mean Squared Error (RMSE), MSE, ACC, Mean (M), Standard Deviation (SD), Range (Re), Correlation (Cr), Mann–Whitney U-test (MW), Kruskal–Wallis test (KW)] and the remarks if any.

From these studies, it was observed that instances of cyberbullying can be easily seen and are more flared-up with the widespread use of social media. This digital abuse risks the mental health and social well-being of the netizens, especially victims. Effective method to analyze online interactions between people to detect cyberbullying activities is an important area of research among social media researchers. Various tools and techniques have been identified and assessed for cyberbullying detection and prediction with linguistic, visual, profile, demographic, and activity features. One such consortium of techniques is soft computing (SC) that envisages group of data-driven approaches that have been used in the past decade for cyberbullying detection (CD) on social multimedia (SM).

These studies show that various SC techniques have been used to detect and predict cyberbullying online. Few initial efforts of using supervised machine learning (ML) for CD in Web 2.0 textual content were reported in 2009 [54] and 2011 [43]. Post that, a range of supervised ML techniques such as SVM, NB, DT, RF, Essential Dimensions of Latent Semantic Indexing, Stochastic Gradient Descent, Bayes Point Machine, Ridge Regression, Logistic LASSO Regression, Extra Tree, Gradient Boosting have been used on a variety of SM namely Twitter, Facebook, MySpace, Formspring.me, Kaggle, Kongregate, Slashdot for CD particularly for textual and audio messages. These proposed methods yield enhanced results as compared to traditional methods (lexicon, rule-based methods) for CD. Likewise, an increasing use of SC techniques such as FL, GA, MaxE, NN etc. and Ensemble methods including Adb, Bos, Bgg have been demonstrated for CD on SM for all sorts of multimedia including images, videos, audios and texts. The implementation was done on the data collected from varied social media such as Vine, Instagram, Amazon and YouTube. The work of the authors depicted improved performance in terms of accuracy for these approaches over statistical methods. Few of the studies also presented the use of distance based algorithms such as kNN and regression based methods namely LogR, LR and MR for CD on SM. Some of the studies also focused on the use of unsupervised and semi-supervised SC techniques for CD on SM. These techniques included C means clustering, Probabilistic Latent Semantic Analysis, Conditional Random Fields, Expectation-Maximization, MinHashes, Graph based classifier and Hierarchical agglomerative clustering. Many reported studies also applied the hybrid techniques namely Fuzzy C means (FCM), Fuzzy Decision Tree (FDT), Fuzzy CD detection on SM in order to boost the performance of the system in terms of accuracy etc. More recently, studies on CD on SM utilizing deep learning (DL) techniques such as Convolution Neural Network (CNN) and pronunciation based convolutionneural network (PCNN), marginalized Stacked Denoising Autoencoders (mSDA), Semantic-Enhanced Marginalized Denoising Auto-Encoder (smSDA), Long short-term memory (LSTM) and Bidirectional Long Short Term Memory (BLSTM) have been reported for multimedia messages collected from social media like Twitter and Formspring.me.

5 Results and discussion

In this section, the results obtained as answers to the RQs defined in the SLR are discussed. Table 4 depicts the mapping of the research articles to the respective RQs.

- *Most distinguished and relevant journals/conferences with published studies in the identified research domain (RQ1)*

Reviewing the literature exhaustively helped us identify the journals/conferences within the selected digital libraries and the ones prominently publishing research in the defined problem domain. The following Table 5 demonstrates the distribution of the research articles in the identified journals/conferences, their proportions and corresponding cumulative proportions of the studies included in this SLR.

The following Table 6 suggests the most relevant and distinguished journals/conferences in this domain of study. Amongst the final selected articles for this SLR, majority of them belonged to Elsevier (Computers in Human Behavior, Elsevier), followed by ACM (Web science conference, International Conference on Advances in Social Networks Analysis and Mining), followed by Springer (European Conference on Information Retrieval), IEEE (International Conference on Machine Learning and Applications) and ACM (International Conference on World Wide Web, ACM).

- *Widely used datasets and domains in which the studies for soft computing techniques in cyberbullying detection on social multimedia have been conducted (RQ2)*

Reviewing the pertinent literature it was observed that the research studies have considered a gamut of data sources including publically available datasets. The probed datasets were either publically available, a set of random data fetched in real-time from various SM like Twitter, Vine, Facebook, MySpace etc., or a random set of collected questionnaires/surveys within a selective topic/ subject/ domain. Publically available datasets were also used in the reported studies. One of the data was provided by 'Fundacion Barcelona Media (FBM)' (http://caw2.barcelonamedia.org/) for the workshop on content analysis. FBM supplied five datasets. Amongst all, the author in [32, 33] had used only three of them, primarily belonging to Kongregate, Slashdot and MySpace. The latter was the data available at (https://rpubs.com/pm0kjp/satire_serious_reviews) that comprised of serious and satire reviews from Amazon and Twitter about the news articles and Amazon's products reviews [50], gathered by Payton and Weigandt for their project 'Satire and Data Science: An Exploration into one of the Current Final Frontiers'.

Table 4 Mapping of RQ's with the relevant research articles

#RQ addressed	Research reference
RQ1	[3, 10, 14, 19, 20, 27, 32, 33, 38, 40]
RQ2	[2–9, 11–25, 29–34, 36–43, 45–48, 50–52, 55–57] (http://caw2.barcelonamedia.org/, https://rpubs.com/pm0kjp/satire_serious_reviews)
RQ3	[2, 3, 6, 14, 15, 17–19, 22, 31, 32, 36–38, 40, 45–47, 50, 52, 56, 57] (https://rpubs.com/pm0kjp/satire_serious_reviews)
RQ4	[2–4, 6, 7, 11–14, 16, 17, 19, 22, 25, 30, 33, 34, 38, 39, 41, 42, 45, 47, 50, 52, 55, 57]

Table 5 Distribution of the research articles in regard to the respective journals/conferences

S. No.	Journal/Conference name	Description	#papers	Proportion (%)	Cumulative Proportion (%)
1.	Elsevier Journals	• Computers & security, Elsevier.	1	17	17
		• Journal of Criminal Justice, Elsevier	1		
		• Computers in Human Behavior, Elsevier	4		
		• Procedia-Social and Behavioral Sciences, Elsevier	1		
		• Knowledge-Based Systems, Elsevier	1		
2.	Elsevier Conferences	• International Conference on Advanced Computing Technologies and Applications (special issue of Procedia Computer Science), Elsevier	1	2	19
3.	Springer Journals	• Social Network Analysis and Mining, Springer	1	6	25
		• International Journal of Machine Learning and Cybernetics, Springer	1		
		• Neural Computing and Applications, Springer	1		
4.	Springer Conferences	• Asia-Pacific Web Conference. Springer, Berlin, Heidelberg	1	26	51
		• European Conference on Information Retrieval, Springer, Berlin, Heidelberg	2		
		• In Australasian Database Conference Springer, Cham	1		
		• Canadian Conference on Artificial Intelligence, Springer, Cham	1		
		• International Conference on Speech and Computer, Springer, Cham	1		
		• International Conference on Industrial Networks and Intelligent Systems, Springer, Cham.	1		
		• International Conference on Statistical Language and Speech Processing. Springer, Cham	1		
		• International Conference on Artificial Intelligence and Soft Computing. Springer, Cham	1		
		• International Conference of the German Society for Computational Linguistics and Language Technology, Springer, Cham	1		
		• Asian Conference on Intelligent Information and Database Systems, Springer, Cham	1		
		• International Conference on Hybrid Artificial Intelligence Systems, Springer	1		
5.	IEEE Journals	• IEEE Transactions on Affective Computing	1	2	53
6.	IEEE Conferences	• International Conference on Computational Intelligence and Computing Research, IEEE	1	21	74
		• International Conference on Circuit, Power and Computing Technologies [ICCPCT], IEEE	1		
		• Advances in computing, communications and informatics (ICACCI), International Conference, IEEE	1		
		• International Conference on Machine Learning and Applications, IEEE	2		

Table 5 (continued)

S. No.	Journal/Conference name	Description	#papers	Proportion (%)	Cumulative Proportion (%)
		• International Conference on Advances in Social Networks Analysis and Mining (ASONAM), IEEE	1		
		• International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), IEEE	1		
		• Cyber Security in Networking Conference IEEE	1		
		• Digital Information Management (ICDIM), Twelfth International Conference, IEEE	1		
		• International Conference on Advances in Computing and Communication Engineering (ICACCE) 2018, IEEE	1		
7.	ACM Conferences	• Conference of the North American chapter of the association for computational linguistics: Human language technologies, ACM	1	24	98
		• Web science conference. ACM.	3		
		• International Conference on Advances in Social Networks Analysis and Mining, ACM	3		
		• Proceedings of the 17th international conference on distributed computing and networking. ACM.	1		
		• Conference on Hypertext and Social Media, ACM	1		
		• International Conference on World Wide Web, ACM	2		
8.	Wiley Journals	•Journal of Computer-Mediated Communication, Wiley	1	2	100

Many reported researches were carried on the data fetched directly from various social media using their respective API's. The fetched corpuses were from a variety of domains, topics and time period (referred as topic specific/topic oriented). These prominently belonged to domains including games and sports [6, 29, 32, 36], entertainment [17, 31], meeting transcripts [48], chats [31], adolescent behaviors and their problems [9, 24], pornography [9], sexual imagery [9], elite personalities and well known celebrities including singers, actors

Table 6 Relevant and distinguished journals/conferences

S. No.	Journal/Conference Name	# papers
1.	• Computers in Human Behavior, Elsevier (Journal)	4
2.	• Web Science Conference, ACM	3
	• International Conference on Advances in Social Networks Analysis and Mining, ACM	3
3.	• European Conference on Information Retrieval, Springer	2
	• International Conference on Machine Learning and Applications, IEEE	2
	•International Conference on World Wide Web, ACM	2

etc. s [19, 25, 29, 39], player demographics [6], statistics [6], complaints [6], sexting or sexual conversations [5, 38], family issues [5], self-esteem behaviors [5], religion [23], politics [23, 29], public profiles [40], user profiles [40], education [46], news [50], consumer products [50], hate speech controversy [11–13], social issues [29] and terrorism [29], etc. The following Fig. 12 depicts the year-wisetrend of published work from various domains taken as dataset for empirical evaluation.

Various other reported studies use random datasets [2–4, 7, 8, 14–16, 18, 20–22, 30, 31, 33, 37, 41, 42, 45, 47, 51, 52, 55–57]. Few studies reported self-administered questionnaires/ surveys conducted by the author involving questions related to cyberbullying comprising of victims and perpetrators [5, 9, 24, 29]. The following Fig. 13 shows the year-wise trend of published work with 'topic specified/ topic oriented' datasets from various domains.

The use of random data sets on general topics from various domains has been considered for research evaluations, especially since the year 2011. The following Fig. 14 specifies the wide array of topics from which the randomized datasets have been taken.

Further, the Table 7 identifies the wide list of SM covered for analyzing studies involving the application of SC techniques for CD.

The following Table 8 enumerates the type of multimedia upon which CD was done using SC techniques.

Likewise, Table 9 provides details about the types of tools and software libraries used for carrying out the analyses and implementation of CD on SM using SC techniques.

- *Most frequently used soft computing techniques for achieving efficient results for cyberbullying detection on social multimedia on SM (RQ3)*

Table 10 illustrates the year wise usage of various SC techniques for CD on SM over a wide range of domains.

Thus, the most frequently used soft computing technique for achieving efficient results for cyberbullying detection on social multimedia is machine learning (support vector machine) followed by deep learning techniques. The graph shown in Fig. 15 depicts the quantitative extent of the use of various soft computing techniques for cyberbullying detection on social media.

Apart from these, other subset of SC techniques that were used for analyzing CD on SM are categorized under the SC consortium as follows: *supervised learning, un-supervised learning and semi-supervised learning*. Techniques like Essential Dimensions of Latent Semantic Indexing (EDLSI) [30], Stochastic Gradient Descent (SGDes) [33], Bayes Point Machine

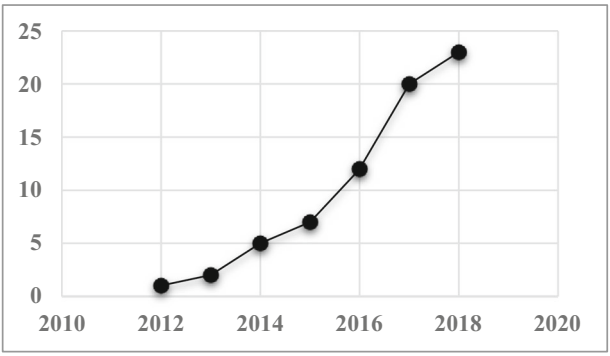

Fig. 12 Year wise cumulative assessment of topic oriented fetched datasets

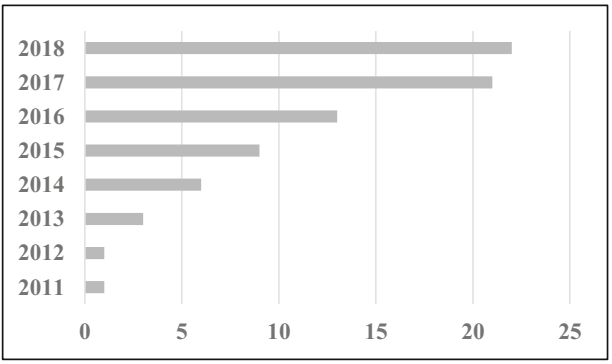

Fig. 13 Year-wise cumulative assessment of random (general) datasets fetched from various SM

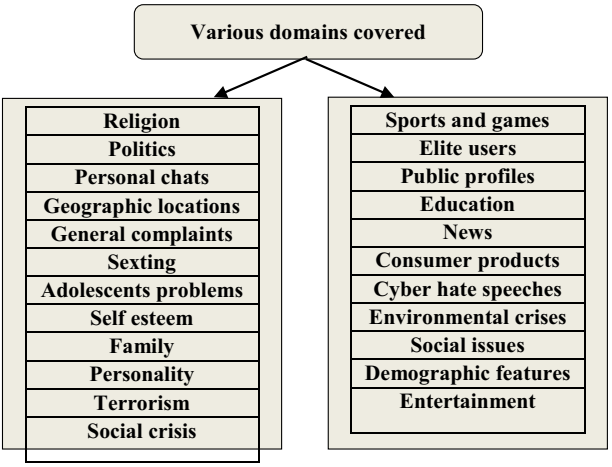

Fig. 14 Various domains covered for research for CD on SM using SC techniques

Table 7 List of social multimedia covered

Social Multimedia	# Papers	Referenced study
Twitter	20	[2, 3, 7, 11–13, 15, 16, 19, 22, 41, 42, 45, 46, 48, 50–52, 55–57] (https://rpubs.com/pm0kjp/satire_serious_reviews)
Forums	9	[2, 6, 7, 20, 21, 32, 33, 36, 50]
Formspring.com	6	[2, 30, 34, 43, 51, 55]
MySpace	5	[32–34, 37, 56]
Kaggle	4	[8, 14, 42, 47]
Facebook	3	[5, 29, 48]
YouTube	3	[17, 18, 23]
Perverted Justice Website	2	[31, 40]
Vine	2	[39, 40]
Ask.fm	2	[4, 52]
Instagram	2	[25, 52]
Slashdot	2	[32, 33]
Reddit	1	[42]

Table 8 List of multimedia covered

Multimedia	# Papers	Referenced Study
Texts	44	[2–9, 11–24, 29–34, 36–38, 41–43, 45–48, 50–52, 55–57] (http://caw2.barcelonamedia.org/, https://rpubs.com/pm0kjp/satire_serious_reviews)
Videos	2	[39, 40]
Audios	1	[48]
Images	1	[25]
Emoticons	0	–
GIF's	0	–

(BPM) [6], Ridge Regression (RR) [25], Logistic LASSO Regression (LLReg) [23], Gradient Boosting (GBos) [19], Multilevel Regression (MR) [29], Extra Tree classifier (ET) [19, 40] and Multilayer Perceptron (MLP) [40] comes under the supervised learning whereas C means clustering (CMC) [48], Probabilistic Latent Semantic Analysis (PLSA) [32], Conditional Random Fields (CRF) [52], Expectation-Maximization (ExpM) [11], MinHashes (MinH) [19], Graph based classifier [36] and Hierarchical agglomerative clustering (HAC) [38] lies under the category of un-supervised learning approaches. Many reported studies also applied the hybrid techniques namely Fuzzy C means (FCM) [33, 48], Fuzzy Decision Tree (FDT) [48], Fuzzy SVM (FSVM) [33], FuzGen (hybrid of Fuzzy Logic and GA) [34] and Bayesian hierarchical clustering (BHC) [38]. Furthermore, few of the studies had also reported the applicability of deep learning (DL) techniques for CD on SM such as pronunciation based convolution neural network (PCNN) [55], marginalized Stacked DenoisingAutoencoders (mSDA) [56], Semantic-Enhanced Marginalized Denoising Auto-Encoder (smSDA) [56],

Table 9 List of tools and software libraries used

Tools	# Papers	Referenced study
Python	8	[22, 36, 42, 45–47, 50, 55]
WEKA	7	[3, 13, 18, 22, 33, 43, 52]
Word2vec	4	[15, 42, 56, 57]
CrowdFlower	4	[13, 19, 39, 40]
SentiStrength	3	[11, 13, 22]
GloVe	2	[2, 15, 42]
MATLAB	2	[38, 45]
NLTK	2	[21, 39]
R	2	[23, 30]
RapidMiner	1	[37]
Rainbow	1	[31]
.NET	1	[6]
LibSVM	1	[32]
ESpeakTheano	1	[55]
IGraph	1	[36]
SKLearn	1	[36]
Java	1	[46]
Séance	1	[50]
Scipy.sparse	1	[41]
Trollslayer	1	[19]
McCallum	1	[7]
SSWE	1	[2]
Face++	1	[16]

Table 10 Year wise distribution of applicability of the SC techniques for CD on SM

Technique	#papers	Year	Referenced study
SVM	27	2012-2018	[2, 3, 6, 14, 15, 17–19, 22, 31, 32, 36–38, 40, 43, 45–47, 50, 52, 56, 57]
DL	15	2016-2018	[2, 8, 15, 20, 51, 55, 56]
NB	14	2012, 2014-2018	[2–4, 13, 18, 22, 31, 33, 34, 39, 40, 45, 51, 52]
RF	13	2014-2018	[2, 3, 12, 13, 19–21, 33, 39, 40, 42, 47]
LogR	13	2012, 2014-2018	[2, 5, 7, 9, 14–16, 24, 25, 33, 40, 47, 52]
DT	9	2014-2017	[6, 7, 13, 18, 19, 39, 40, 43, 45]
kNN	5	2016-2017	[3, 31, 40, 43, 45]
FL	4	2013-2015, 2017	[31, 34, 46, 48]
LDA	4	2016-2017	[4, 40, 56, 57]
MaxE	3	2013-2014, 2017	[7, 31, 48]
Adb	3	2015-2017	[19, 39, 40]
NN	2	2016-2017	[13, 21]
LSA	2	2016-2017	[56, 57]
KMC	2	2015, 2017	[6, 11]
Bys	2	2012, 2017	[7, 32]

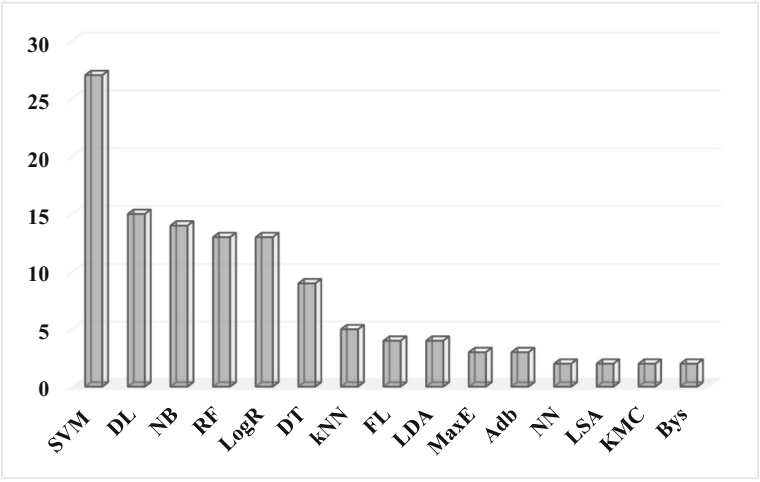

Fig. 15 Quantitative extent of the use of various SC techniques

Long short-term memory (LSTM) [15] andBidirectional Long Short Term Memory (BLSTM) [15]. These hybrid techniques utilizing the combination of supervised and un-supervised methods comes under semi-supervised techniques, such as Bayes Expectation Maximization

Table 11 Year wise distribution of the use of hybrid techniques

S. no.	Technique	#papers	Year	Referenced study
1.	FuzGen (Fuzzy Logic + GA)	1	2015	[34]
2.	FCM (Fuzzy Logic + C Means Clustering)	1	2013-2014	[33, 48]
3.	FDT (Fuzzy Logic + Decision Tree)	1	2013	[48]
4.	FSVM (Fuzzy Logic + Support Vector Machines)	1	2014	[33]
5.	BHC (Bayesian Classification + Hierarchical Clustering)	1	2016	[38]

Table 12 Key performance indicators used

S. no.	KPI	Description	Referenced study
1.	Precision (P)	It defines the exactness of any classifier. A higher precision value indicates fewer 'false positives' (FP) and vice versa. It is given as the ratio of true positives (TP) to all the predicted positives.	[2–4, 6, 7, 11–14, 16, 17, 19, 22, 25, 30, 33, 34, 36, 39–42, 45, 47, 50, 52, 55, 57]
2.	Recall (R)/Sensitivity (Sn)	It defines the sensitivity or the completeness of any classifier. A higher recall value indicates less 'false positives' and vice versa. Recall and precision are bounded by inverse relation with each other. It is given as the ratio of TP to all the actual positives (TP + FN)	[2–4, 7, 11–14, 16, 17, 19, 22, 25, 30, 33, 36, 38–40, 45, 47, 50, 52, 55, 57]
3.	F score/F1 score/F measure/F1 measure/F1 value(F)	It is defined as 'weighted' harmonic mean of Recall and Precision. It is the combination of the precision and recall measures.	[2–4, 7, 17, 19, 20, 22, 25, 33, 34, 36, 45, 50, 52, 55–57]
4.	Accuracy (A)	It is defined as proximity of a measurement to its true value. It is calculated as a proportion of TP and true negatives (TN) among total inspected cases.	[7, 9, 15, 21, 31, 32, 34, 39, 40, 43, 46–48, 50–52, 55, 56]
5.	AUC (Area under the curve)	It is interpreted as AUROC (area under the receiver operating characteristics curve) and is also used for measuring the efficacy of the classifiers based on two metrics i.e. true positive rate and false positive rate.	[3, 8, 14, 18, 25, 42, 47]
6.	ROC	ROC curve plots sensitivity as a function of Specificity for different cut-off points.	[8, 11–13, 25]
7.	Cohen's Kappa (CK)	This metric performs a comparison between the observed and the expected accuracy. It is used for evaluating performance of a single classifier and also among different classifiers.	[4, 12, 13]
8.	Mean (M)	Average of numbers is defined as mean. It is calculated by finding sum of all numbers and then divided the sum by count of numbers.	[24, 29]
9.	Standard Deviation (SD)	It measures the dispersion of data values from mean. Square root of variance gives SD.	[24, 29]
10.	Co-relation (Cr)	It is a measure that shows the degree to which two or more variables vary together.	[24, 29]
11.	Specificity (Sp)	Sp is referred as true negative rate and is denoted as ratio of TN to all the actual negatives (TN + FP).	[6]
12.	Confidence (Cn)	Cnis the probability of prediction of test sample to be in each class.	[37]
13.	Range (Rg)	It is defined as difference between highest and lowest values.	[24]
14.	ACC	It ensures that the label set predicted for a sample must tally with corresponding label set in ground truth labels.	[14]
15.	Mann-Whitney U-Test (MW)	It is used to compare whether two sample means are equal or not.	[5]
16.	Kruskal-Wallis Test (KW)	It is a non-parametric test to determine whether different samples emerge from same distribution.	[5]
17.	MSE	Mean Squared Error is the average of squared differences between actual and predicted values.	[23]
18.	Root Mean Squared Error (RMSE)	It is defined as square root of Mean Squared Error.	[13]
19.	Micro AUC	It is the area under ROC curve.	[15]
20.	Macro AUC	It is the average AUC of the distinct ROC curves for each class	[15]

(BExpM) [7], Fuzzy C means (FCM) [33, 48], Participant-Vocabulary Consistency (PVC) using Alternating Least Squares [41] etc. The following Table 11 presents the studies with hybrid techniques.

- *Widely used performance measures (RQ4)*

The following performance measures were observed in all selected studies to measure the performances of the applied SC techniques. They are referred to as the 'key performance indicators' (KPIs) or 'performance parameters' or 'efficacy measures'. Table 12 illustrates the KPI's used so far.

6 Conclusion

Cyberbullying can take many forms; however, it typically refers to repeated and hostile behavior online to intentionally and repeatedly harass or harm individuals. With the pervasive use of social media, cyberbullying is becoming rampant. Conventional methods to combat cyberbullying included guidelines on cyber-ethics, human moderators, and blacklisting based on the use of profane words. These methods are incapable to deal with the mounting velocity, volume and variety of data generated by the social media, thus necessitating the design and development of novel learning-based computational models. A systematic literature review was conducted on the use of soft computing techniques to detect cyberbullying activities across various social media domains to understand the theory, research and practice trends within the domain. Combining the results of individual studies, as a meta-summary, it was observed that the use of soft computing techniques for cyberbullying detection provided the intelligent analytic paradigm essential for predicting bullying behaviors and activities on both textual and non-textual social media. It is a promising direction of research with practical domain which primarily relies on exploring and understanding the extensibility of human expressions via both textual and non-textual unstructured web-data available. Based on the review, the following research gaps have been identified:

- Detecting and predicting cyberbullying behavior is a non-trivial task with anonymous cyberbullying as an added concern.
- Although researchers are keen in applying soft computing techniques, only few approaches have been explored. Techniques such as deep learning, neural network, ensemble methods, evolutionary computing and hybrids including neuro-fuzzy models have been least explored to substantiate their influence on CD.
- The existing models using SC techniques for CD on SM have majorly considered Twitter, MySpace, Formspring.me, Facebook as the database, making other SM technologies such as Reddit, Vine, Instagram, Flickr, Tumblr, Ask.fm, YouTube etc. open to further application and testing.
- Extracting, selecting and modeling computational features such as linguistic, visual, socio-demographic features (like person's economic status, age, gender, etc.), socio-ecological features (like parental monitoring, hours spent on Internet, racial/community differences etc.) and activity features (behavioral factors) needs further concurrence of soft computing techniques with natural language models and network analysis techniques.
- Most of the reported work done is to detect cyberbullying activity is using textual content, whereas other media types such as audio, video, images are open to research initiatives.

Also, the use of animated GIF's, memes has recently been reported to embarrass or target people on social media, making it an open area of research.

- The informal, short, noisy and unstructured social media further add to the challenges. The use of slangs, mal-formed or colloquial words,mash-up languages (mixed usage of different languages, for example, Hinglish is a mixture of English and Indian Hindi language) make detection of online bullying activities tricky and computationally hard.

Thus, the need to exploit new computational models to detect and predict cyberbullying on social multimedia is abundant making this domain of study a potentially active and dynamic for social media researchers. It compels to look for models that combine the cognition, intelligence and self-tuning behavior of soft computing techniques with disciplines like natural language processing, psychology and artificial intelligence.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Aggarwal CC (2018) Neural networks and deep learning: a textbook. Springer, Berlin
2. Agrawal S, Awekar A (2018) Deep learning for detecting cyberbullying across multiple social media platforms. In: European conference on information retrieval. Springer, Cham, pp 141–153
3. Al-garadi MA, Varathan KD, Ravana SD (2016) Cybercrime detection in online communications: the experimental case of cyberbullying detection in the twitter network. Comput Hum Behav 63:433–443
4. Ashktorab Z, Haber E, Golbeck J, Vitak J (2017) Beyond cyberbullying: self-disclosure, harm and social support on ASKfm. In: Proceedings of the 2017 ACM on Web Science Conference, p 3-12
5. Balakrishnan V (2015) Cyberbullying among young adults in Malaysia: the roles of gender, age and internet frequency. Comput Hum Behav 46:149–157
6. Balci K, Salah AA (2015) Automatic analysis and identification of verbal aggression and abusive behaviors for online social games. Comput Hum Behav 53:517–526
7. Bourgonje P, Moreno-Schneider J, Srivastava A, Rehm G (2017) Automatic classification of abusive language and personal attacks in various forms of online communication. In: International conference of the German Society for Computational Linguistics and Language Technology. Springer, Cham, pp 180–191
8. Bu SJ, Cho SB (2018) A hybrid deep learning system of CNN and LRCN to detect cyberbullying from SNS comments. International Conference on Hybrid Artificial Intelligence Systems 2018, Springer, p 561–572
9. Byrne S, Katz SJ, Lee T, Linz D, McIlrath M (2014) Peers, predators, and porn: predicting parental underestimation of children's risky online experiences. J Comput-Mediat Commun 19(2):215–231
10. Campbell MA (2005) Cyber bullying: an old problem in a new guise? J Psychol Couns Sch 15(1):68–76
11. Chatzakou D, Kourtellis N, Blackburn J, De Cristofaro E, Stringhini G, Vakali A (2017a) Hate is not binary: studying abusive behavior of# gamergate on twitter. In Proceedings of the 28th ACM conference on hypertext and social media, p 65-74
12. Chatzakou D, Kourtellis N, Blackburn J, De Cristofaro E, Stringhini G, Vakali A (2017b) Detecting aggressors and bullies on Twitter. In: Proceedings of the 26th International Conference on World Wide Web Companion, p 767-768
13. Chatzakou D, Kourtellis N, Blackburn J, De Cristofaro E, Stringhini G, Vakali A (2017c) Mean birds: Detecting aggression and bullying on twitter. In: Proceedings of the 2017 ACM on web science conference, p 13-22
14. Chavan VS, Shylaja SS (2015) Machine learning approach for detection of cyber-aggressive comments by peers on social media network. In: Advances in computing, communications and informatics (ICACCI), 2015 International Conference on IEEE, p 2354-2358
15. Chen J, Yan S, Wong KC (2018) Verbal aggression detection on twitter comments: convolutional neural network for short-text sentiment analysis. Neural Comput & Applic:1–10
16. Coletto M, Lucchese C, Orlando S (2018) Do violent people smile: social media analysis of their profile pictures. In: Companion of the Web Conference 2018. International World Wide Web Conferences Steering Committee, ACM, p 1465–1468

17. Dadvar M, Trieschnigg D, Ordelman R, de Jong F (2013) Improving cyberbullying detection with user context. In: European conference on information retrieval. Springer, Berlin, Heidelberg, pp 693–696
18. Dadvar M, Trieschnigg D, de Jong F (2014) Experts and machines against bullies: a hybrid approach to detect cyberbullies. In: Canadian conference on artificial intelligence. Springer, Cham, pp 275–281
19. García-Recuero Á (2017) Efficient privacy-preserving adversarial learning in decentralized online social networks. In: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ACM, p 1132–1135
20. Gordeev D (2016a) Detecting state of aggression in sentences using CNN. In: International Conference on Speech and Computer, Springer, Cham, p 240-245
21. Gordeev D (2016b) Automatic detection of verbal aggression for Russian and American image boards. Procedia Soc Behav Sci 236:71–75
22. Haidar B, Chamoun M, Serhrouchni A (2017) Multilingual cyberbullying detection system: Detecting cyberbullying in Arabic content. In: Cyber Security in Networking Conference (CSNet), IEEE, p 1-8
23. Hammer HL (2016) Automatic detection of hateful comments in online discussion. In: International Conference on Industrial Networks and Intelligent Systems, Springer, Cham, p 164-173
24. Holt TJ, Turner MG, Exum ML (2014) The impact of self control and neighborhood disorder on bullying victimization. J Crim Just 42(4):347–355
25. Hosseinmardi H, Rafiq RI, Han R, Lv Q, Mishra S (2016) Prediction of cyberbullying incidents in a media-based social network. In: Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, p 186–192
26. http://www.ryanpatrickhalligan.org/ Accessed 14 July 2018
27. https://coschedule.com/ Accessed 14 July 2018
28. Kitchenham B, Charters S (2007) Guidelines for performing systematic literature reviews in software engineering. Tech Rep EBSE 1:1–57
29. Koban K, Stein JP, Eckhardt V, Ohler P (2018) Quid pro quo in web 2.0. Connecting personality traits and Facebook usage intensity to uncivil commenting intentions in public online discussions. Comput Hum Behav 79:9–18
30. Kontostathis A, Reynolds K, Garron A, Edwards L (2013) Detecting cyberbullying: query terms and techniques. In: Proceedings of the 5th annual ACM web science conference, p 195-204
31. Michalopoulos D, Mavridis I, Jankovic M (2014) GARS: real-time system for identification, assessment and control of cyber grooming attacks. Computers & Security 42:177–190
32. Nahar V, Unankard S, Li X, Pang C (2012) Sentiment analysis for effective detection of cyber bullying. Asia-Pacific Web Conference, Springer, Berlin, Heidelberg, p 767-774
33. Nahar V, Al-Maskari S, Li X, Pang C (2014) Semi-supervised learning for cyberbullying detection in social networks. In: Australasian database conference. Springer, Cham, pp 160–171
34. Nandhini BS, Sheeba JI (2015) Online social network bullying detection using intelligence techniques. Procedia Computer Science 45:485–492. International Conference on Advanced Computing Technologies and Applications (ICACTA)
35. National Bullying Prevention Center https://www.pacer.org/bullying/. Accessed 26 July 2018
36. Papegnies E, Labatut V, Dufour R, Linares G (2017) Graph-based features for automatic online abuse detection. In: International conference on statistical language and speech processing. Springer, Cham, pp 70–81
37. Parime S, Suri V (2014) Cyberbullying detection and prevention: data mining and psychological perspective. In: Circuit, Power and Computing Technologies (ICCPCT), 2014 International Conference IEEE, p 1541-1547
38. Potha N, Maragoudakis M, Lyras D (2016) A biology-inspired, data mining framework for extracting patterns in sexual cyberbullying data. Knowl-Based Syst 96:134–155
39. Rafiq RI, Hosseinmardi H, Han R, Lv Q, Mishra S, Mattson SA (2015) Careful what you share in six seconds: detecting cyberbullying instances in Vine. In: Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ACM, p 617–622
40. Rafiq RI, Hosseinmardi H, Mattson SA, Han R, Lv Q, Mishra S (2016) Analysis and detection of labeled cyberbullying instances in vine, a video-based social network. Soc Netw Anal Min 6(1):88
41. Raisi E, Huang B (2017) Cyberbullying detection with weakly supervised machine learning. In: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ACM, p 409–416
42. Rakib TB, Soon LK (2018) Using the Reddit Corpus for cyberbully detection. In: Asian conference on intelligent information and database systems. Springer, Cham, pp 180–189
43. Reynolds K, Kontostathis A, Edwards L (2011) Using machine learning to detect cyberbullying. In: Machine learning and applications and workshops (ICMLA), 2011 10th International Conference, IEEE 2:241-244

44. Salawu S, He Y, Lumsden J (2017) Approaches to automated detection of cyberbullying: a survey. IEEE Trans Affect Comput 1:1–20
45. Sarna G, Bhatia MP (2017) Content based approach to find the credibility of user in social networks: an application of cyberbullying. Int J Mach Learn Cybern 8(2):677–689
46. Sedano CR, Ursini EL, Martins PS (2017) A bullying-severity identifier framework based on machine learning and fuzzy logic. In: International conference on artificial intelligence and soft computing. Springer, Cham, pp 315–324
47. Sharma HK, Kshitiz K, Shailendra (2018) NLP and machine learning techniques for detecting insulting comments on social networking platforms. International Conference on Advances in Computing and Communication Engineering (ICACCE) 2018, IEEE, p 265-272
48. Sheeba JI, Vivekanandan K (2013) Low frequency keyword extraction with sentiment classification and cyberbully detection using fuzzy logic technique. In: IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), p 1–5
49. The National Crime Prevention Cyberbullying. http://www.ncpc.org/cyberbullying. Accessed 10 March 2018
50. Thu PP, New N (2017) Implementation of emotional features on satire detection. In Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 2017 18th IEEE/ACIS International Conference, IEEE, p 149-154
51. Wint ZZ, Ducros T, Aritsugi M (2017) Spell corrector to social media datasets in message filtering systems. In: Digital Information Management (ICDIM), 2017 Twelfth International Conference, IEEE, p 209–215
52. Xu JM, Jun KS, Zhu X, Bellmore A (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, p 656-666
53. Ybarra M (2010) Trends in technology-based sexual and non-sexual aggression over time and linkages to nontechnology aggression. National Summit on Interpersonal Violence and Abuse Across the Lifespan: Forging a Shared Agenda
54. Yin D, Xue Z, Hong L, Davison BD, Kontostathis A, Edwards L (2009) Detection of harassment on web 2.0. Proceedings of the Content Analysis in the WEB 2:1-7
55. Zhang X, Tong J, Vishwamitra N, Whittaker E, Mazer JP, Kowalski R, Hu H, Luo F, Macbeth J, Dillon E (2016) Cyberbullying detection with a pronunciation based convolutional neural network. In: 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), p 740–745
56. Zhao R, Mao K (2017) Cyberbullying detection based on semantic-enhanced marginalized denoising auto-encoder. IEEE Trans Affect Comput 8(3):328–339
57. Zhao R, Zhou A, Mao K (2016) Automatic detection of cyberbullying on social networks based on bullying features. In: Proceedings of the 17th international conference on distributed computing and networking, p 43-48

Dr. Akshi Kumar is an Assistant Professor in the Department of Computer Science & Engineering at Delhi Technological University (formerly Delhi College of Engineering). She has been with the university for the past 10 years. She has received her doctorate degree in Computer Engineering in the area of Web Mining, from Faculty of Technology, University of Delhi in 2011. She completed her M. Tech (Master of Technology) with honors in Computer Science & Engineering from Guru Gobind Singh Indraprastha University, Delhi in 2005. She received her BE (Bachelor of Engineering) degree with distinction in Computer Science & Engineering from Maharishi Dayanand University, Rohtak in 2003. Dr. Kumar's research interests are in the area of Intelligent Systems, User-generated Big-data, Social Media Analytics and Soft Computing. She has many publications to her credit in various International Journals with high impact factor and International Conferences with best paper awards. She has authored books & book-chapters within her domain of interest. She is actively involved in research activities as the editorial review board member for international journals, session chair & technical program committee member for international conferences and providing research supervision at under-graduate, post-graduate and doctorate levels. She is a currently a member of IEEE, IEEE (WIE) and ACM and a life member of IACSIT, IAENG, ISTE and CSI.

Mr. Nitin Sachdeva is a Research Scholar in the Department of Computer Science & Engineering at Delhi Technological University, Delhi, India. He completed his Masters in Information Security [MTech (IS), 2014-2016] from the Department of Computer Science & Engineering at Dr. BR Ambedkar National Institute of Technology, Jalandhar, India. He pursued his Bachelors in Information Technology from USIT, Kashmere Gate, Main Campus, Delhi, India. His research interests include Web Technologies, Soft Computing, Intelligent Systems, Text Mining and Social Web Analytics.