2022-12-22

Investigation, Detection and Prevention of Online Child Sexual Abuse Material: A Comprehensive Survey

Vuong Ngo
Technological University Dublin, vuong.ngo@tudublin.ie

Christina Thorpe
Technological University Dublin, christina.thorpe@tudublin.ie

Cach N. Dang
Ho Chi Minh City University of Transport, Cach@ut.edu.vn

See next page for additional authors

Recommended Citation

V. M. Ngo, C. Thorpe, C. N. Dang and S. Mckeever, "Investigation, Detection and Prevention of Online Child Sexual Abuse Materials: A Comprehensive Survey," 2022 RIVF International Conference on Computing and Communication Technologies (RIVF), Ho Chi Minh City, Vietnam, 2022, pp. 707-713, doi: 10.1109/RIVF55975.2022.10013853.

Funder: The Safe Online Initiative of End Violence and the Tech Coalition

Investigation, Detection and Prevention of Online Child Sexual Abuse Material: A Comprehensive Survey

Vuong M. Ngo ✉
School of Computer Science
Technological University Dublin, Ireland
Email: Vuong.Ngo@tudublin.ie or Vuong.cs@gmail.com

Christina Thorpe
School of Informatics and Cybersecurity
Technological University Dublin, Ireland
Email: Christina.Thorpe@tudublin.ie

Cach N. Dang
Science and Technology Application for Sustainable Development (STASD) Research Group
Ho Chi Minh City University of Transport, Vietnam
Email: Cach@ut.edu.vn

Susan Mckeever
School of Computer Science
Technological University Dublin, Ireland
Email: Susan.Mckeever@tudublin.ie

***Abstract*—Child sexual abuse inflicts lifelong devastating consequences for victims and is a growing social concern. In most countries, child sexual abuse material (CSAM) distribution is illegal. As a result, there are many research papers in the literature which proposed technologies to detect and investigate CSAM. In this survey, a comprehensive search of the peer-reviewed journal and conference paper databases (including preprints) is conducted to identify high-quality literature. We use the PRISMA methodology to refine our search space to 2,761 papers published by Springer, Elsevier, IEEE and ACM. After iterative reviews of title, abstract and full text for relevance to our topics, 43 papers are included for full review. Our paper provides a comprehensive synthesis about the tasks of the current research and how the papers use techniques and dataset to solve their tasks and evaluate their models. To the best of our knowledge, we are the first to focus exclusively on online CSAM detection and prevention with no geographic boundaries, and the first survey to review papers published after 2018. It can be used by researchers to identify gaps in knowledge and relevant publicly available datasets that may be useful for their research.**

I. Introduction

With the rise of the digital age in the 21st century, the use of online platforms (e.g. social media) has grown exponentially over time with the growth of the Internet. Online platforms have allowed people to share information and free expression. However, these networks are also used for supporting sexual abuse behaviours, such as grooming, sex exploitation, indecent exposure, forced intercourse, sexual torture and coordinating sex trafficking ([1]). Online sexual abuse is a global issue but techniques for detecting and categorising are still limited in their availability and application ([2]). The production and consumption of child sexual abusive materials (CSAMs) have many negative lasting effects on the victims, and impacts society as a whole. Victims of child sexual abuse (CSA) can live with both short and long term physical and psychological effects, such as anxiety, depression, self-harm, suicide, sexually transmitted infections and/or pregnancy.

Abusers can create, consume or distribute CSAM in the form of text, image or video for their purposes. The number of unreported sexual abuse instances is typically far higher than the number actually reported, as indicated by U.S figures of (69%) unreported versus (31%) reported[1], because the victims are afraid to tell anyone, and the legal procedure for validating an episode is complex and traumatising for the victim. Sexual predator and CSAM detection is critical to not only protect child victims from being abused online but also prevent the cycle of abuse that is repeated each time as CSAMs are shared or viewed. Our aim in this paper is to provide an update to date analysis of research works in the domain of CSAM investigation and detection online.

Related work: Similar to our paper, there are 8 survey papers in the domain of CSAM; they are Ali et al. [3], Christensen and Pollard [4], Lee et al. [1], Russell et al. [5], Sanchez et al. [2], Slavin et al. [6], Steel et al. [7] and Steel et al. [8]. In [3], the authors reviewed 42 papers to study the Internet's role in developing CSA for commercial and non-commercial purposes. In [4], 6 law enforcement strategies combating CSAM were selected. The paper explained how the strategies can work and success. In [1], 21 CSAM research papers were reviewed about policy and legal frameworks, distribution channels and applied technologies. In that, distribution channels include P2P networks, darknet, web search engine and website, mobile devices and social media. The applied technologies were detailed as follows: an image hash database, web-crawler, visual detection algorithm and machine learning. In [5], 8 research papers were reviewed to identify the nature of CSA prevention strategies and interventions in

[1] https://www.rainn.org/statistics/criminal-justice-system

developing countries. In addition, the paper also analysed the typical settings and population groups which were used by the intervention strategies. In [2], some papers were studied to analyse the functions, accuracy, importance and effectiveness of the tools used by the child exploitation investigators. In [6], 21 papers were reviewed to measure potential risk factors and clinical correlates of compulsive child behaviour that may help to prevent abuse and treat victims more effectively. In [7], 20 papers were reviewed to learn about the cognitive distortions associated with sexual touching of CSAM offenders. In [8], 33 papers were reviewed to identify the technology used by CSAM offenders.

Our contributions: the review papers mentioned focus on papers published almost exclusively before 2019. Our review covers 43 papers, the vast majority published from 2018 to 2022. Moreover, our contribution is completely different from the previous review papers. This paper provides a comprehensive survey in the domain of online CSAM investigation and detection based on both information technology and social science. It analyses the data sets used, the tasks addressed and the AI techniques applied in the state-of-the-art CSAM literature. To the best of our knowledge, we are the first to focus exclusively on online CSA detection and prevention with no geographic boundaries, and the first survey to review papers published after 2018. This survey is a valuable resource for the research community working in the CSAM domain. It is a scoping review that identifies literature in the online CSAM, and the first assessment of the size of the literature since 2018. It can be used by researchers to identify gaps in knowledge and relevant publicly available datasets that may be useful for their research.

The rest of this paper is organised as follows: in Section II, we detail the methodology employed in selecting the papers form the literature. In Section III, the current research in CSAM is presented and analysed under the following headings: the datasets used, the tasks addressed and the analysis/modelling techniques applied. Section IV contains a discussion of the main findings from the survey. Finally, Section V presents conclusion and future work.

II. Method

We focused on reviewing research papers in the area of online CSAM. Our overall methodology which is used to identify the relevant papers is based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)[2], as used in previous review papers such as [1], [6], [7] and [8]. The following criteria were used:

- Published by Springer, Elsevier, IEEE or ACM which respectively have libraries being Springer Nature[3], Science Direct[4], IEEE Xplore[5] or ACM DL[6].
- Published from 2018 to 2022.
- Written in English, not discriminating by geographical area and dataset language.
- Title or keywords or abstract of each paper has keywords: (Child OR Children) AND (Sex OR Sexual) AND (Abuse OR Exploitation OR Material OR CSAM OR CSEM). The keywords are used in the Boolean search query based on the form requirements of each library.

[2]https://www.prisma-statement.org/Extensions/Protocols
[3]https://link.springer.com/
[4]https://www.sciencedirect.com
[5]https://ieeexplore.ieee.org
[6]https://dl.acm.org

TABLE I
THE NUMBER OF RELATED LITERATURES OF EACH PUBLISHER

Publisher with search criteria	**The number of literatures**		
	After downloading	After reviewing title and abstract	After reviewing full text
Springer	1,255	50	**13** (Res[a] 11, Sur[b] 2)
Elsevier	1,216	46	**12** (Res: 7, Sur: 5)
IEEE	274	14	**7** (Res: 7, Sur: 0)
ACM	16	9	**4** (Res: 4, Sur: 0)
Total	2,761	119	**36** (Res: 29, Sur: 7)
Others with unlimited publisher and pub. year			**7** (Res: 6, Sur: 1)
Final Review			**43** (Res: 35, Sur: 8)

[a] Res: the number of research papers.
[b] Sur: the number of survey papers.

As shown in Table I, with search criteria in above, there are 2,761 downloaded papers through the advanced search functions of the publisher libraries. After reviewing the title and abstract of the papers, we identified 119 relevant papers. For each of these, we continued to review the full text carefully, and selected 36 papers which are most relevant to the purpose of our paper. To avoid missing other relevant papers, we searched on Google Scholar[7] and did not limit publisher or publication year of papers. This search led us to find an extra 7 relevant papers, resulting in a total of 43 relevant papers. Finally, there are 35 research papers and 8 review papers in the domain of CSAM investigation and detection. We presented and compared the 8 review papers as related work in Section I. The 35 research papers will be analysed, classified and discussed in detail in the next sections.

III. Classification and Analysis

Table II presents classification of the current research in terms of datasets, the CSAM tasks addressed and the techniques applied. Each paper is listed under the associated publisher's name and the year of publication is specified. The datasets are categorised based on their accessibility and origin. The CSAM task addressed by the research is based on the nature of the CSAM media involved, with categories of text, image and video. Finally, the applied technique categories are statistic, natural language processing (NLP) and machine learning (ML)/deep learning (DL). Some additional specific categorisation is also highlighted in the rare cases where papers do not fit the main categories.

A. CSAM Tasks addressed

CSAM related activities manifest as text based communications, images and videos.

[7]https://scholar.google.com/

TABLE II
DATASETS, TASKS ADDRESSED AND APPLIED TECHNIQUES OF CURRENT RESEARCH

No	Study	Year	Datasets			Tasks Addressed			Applied Techniques		
Springer			Open	Self-Build	Third-Party	Text	Image	Video	Statistic	NLP	ML/DL
1	Akhter et al. [9]	2021	x		x	x					x
2	Cecillon et al. [10]	2021		x		x					x
3	Christensen and Pollard [11]	2022		x		x			presentation		
4	Emery et al. [12]	2019		x		x			x		
5	Graham et al. [13]	2018	x		x	x			x		
6	Guastaferro et al. [14]	2019		x		x			presentation		
7	Jonsson et al. [15]	2019		x		x			x		
8	Maas et al. [16]	2021		x		x			x		
9	Quayle [17]	2020			x	x			x		
10	Rind [18]	2018	x		x	x			x		
11	Shaw et al. [19]	2021	x		x	x			x		
Elsevier			Open	Self-Build	Third-Party	Text	Image	Video	Statistic	NLP	ML/DL
1	Borg et al. [20]	2019		x			x		direct medical examination		
2	Gangwar et al. [21]	2021	x	x			x				x
3	Guerra and Westlake [22]	2021			x		x			x	
4	Kissos et al. [23]	2020		x			x				x
5	Kokolaki et al. [24]	2020		x		x			x		
6	Ngejane et al. [25]	2021	x		x	x					x
7	Vitorino et al. [26]	2018			x		x				x
IEEE			Open	Self-Build	Third-Party	Text	Image	Video	Statistic	NLP	ML/DL
1	Andrews et al. [27]	2020		x		x				x	
2	Borj et al. [28]	2021	x		x	x					x
3	Bours and Kulsrud [29]	2019	x		x	x				x	x
4	Fauzi and Bours [30]	2020	x		x	x					x
5	Islam et al. [31]	2020		x		x					x
6	Li et al. [32]	2018		x		x				x	x
7	Samra et a. [33]	2021		x		sensor data			x		
ACM			Open	Self-Build	Third-Party	Text	Image	Video	Statistic	NLP	ML/DL
1	Bursztein et al. [34]	2019		x			x	x	x		
2	Laranjeira et al. [35]	2022	x		x		x				x
3	Struppek et al. [36]	2022			x		x				x
4	Sultana et al. [37]	2022		x		x			x		
Others			Open	Self-Build	Third-Party	Text	Image	Video	Statistic	NLP	ML/DL
1	Al-Nabki et al. [38]: arXiv	2020	x		x	x					x
2	Hole et al. [39]: AIC[a]	2022		x				x			x
3	Owen and Savage [40]: CIGI-RIIA[b]	2015		x		x			x		
4	Pereira et al. [41]: arXiv	2020		x		x					x
5	Westlake and Bouchard [42]: Elsevier	2016		x			x	x	x		
6	Woodham et al. [43]: Frontiers	2021		x		x			x		

[a] AIC: published by Australian Institute of Criminology.
[b] CIGI-RIIA: published by the Centre for International Governance Innovation and the Royal Institute of International Affairs.

1) Working on Text: The current social networks are regularly misused to post or spread CSA messages, comments and chat conversation. Hence, the 24 research papers [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [24], [25], [27], [28], [29], [30], [31], [32], [37], [38], [40], [41] and [43] proposed methods to detect, prevent and/or process CSAMs in text format.

Research works in this category involve two expert domain areas: computer science and social sciences. Sociologists studied about: (1) CSAM offenders ([11], [17]); (2) The effect of perceived informal social control on physical CSA ([12]); (3) CSA prevention related to parents ([14], [19]); (4) Association of CSA with culture, race/ethnicity, psychological health and/or risk behaviours ([13], [15], [18]); (5) Consideration of 'slut pages' as a social form of CSA images ([16]); (6) Correlation between the Greek Hotline reports and dark webs forum logs ([24]); (7) How to support victims and identify places unsafe for children ([37]); and (8) Dark web operation ([43]).

To provide a safe environment for children in online networks, computer scientists applied artificial intelligent algorithms for: (1) Detecting abusive comments, chat conversations and/or messages ([9], [10], [38], [31], [41]); (2) Analysing responses and posts on Twitter and Facebook in India ([27]); (3) Processing chat logs to discover behaviours of potential predators ([25]), or identify predatory conversations and sexual predators ([28], [29], [30]); (4) Analysing the complex adult service websites to determine human trafficking organizations operating the websites ([32]); and (5) Investigating CSAM collected on Tor Dark Net[8] ([40]).

2) Working on Images: In 2021, 29.3 million were detected online and removed [9]. The sheer volume and distribution levels

[8] https://www.torproject.org/
[9] https://www.missingkids.org/ourwork/ncmecdata

of CSA images is such that human experts can no longer handle the manual inspection. Seven papers [20], [21], [22], [23], [26], [35] and [36] processed or worked on CSA images. In [20], the authors proposed how to detect physical signs in the body parts of victims that will help general practitioners or hospital workers to detect and investigate CSA. The work in [21], [22], [26], [35], [36] detected the CSA images with the goal of removing them from social media sites. In addition, a CSA image is often considered as evidence of a crime in progress. In [23], to support professionals to detect victims of CSA when there is no forensic evidence, the authors addressed CSAM from self-figure drawings.

3) Working on Videos or Sensor Data: Three papers [34], [39] and [42] focused on processing video. While a fourth paper [33] used information from sensors to detect child abuse activities. In [34] and [42], the authors proposed methods to detect and prevent the distribution of CSA images and videos on online sharing platforms. In [39], Hole et al. developed a software prototype that used both faces and voices to match victims and offenders across CSA videos. The sensor-based work [33] designed and implemented a system which used a smartwatch and two pressure sensors on a belt and on underclothes to detect sexual abuse behaviours. The smartwatch also contains some sensors to detect the body temperature, the heart rate and skin conductance.

B. Datasets Used

The datasets to support research work consist of data gathered directly by the research group, versus third party datasets.

1) Self-Build: Twenty-one research works [10], [11], [12], [14], [15], [16], [20], [21], [23], [24], [27], [31], [32], [33], [34], [37], [39], [40], [41], [42] and [43] created their own datasets to evaluate and discuss their systems. For example, Quayle [17] collected CSA text from social networks and media sources, namely Facebook, BBC news, Skype, WhatsApp and Instagram. Westlake and Bouchard [42] created a dataset containing over 4.8 million web pages which were crawled from the 10 networks with 300 websites in each network. In [43], the authors used data based on descriptions of 53 anonymous CSAM suspects in the United Kingdom who were active on the dark web and noticed by the police. The number of pages of available data per suspect varied and ranged from 1 to 12 transcripts, and 1 to 411 pages.

With the exception of [21], none of the self-build group have published their dataset. In [21], the authors made Pornography-2M and Juvenile-80k datasets[10] which contain 2 million pornography images and 80 thousand age-group images, respectively.

2) Third-Party: The 14 papers [9], [13], [17], [18], [19], [22], [25], [26], [28], [29], [30], [35], [36] and [38] used datasets of third-parties to evaluate and analyse the methods. Of these, the datasets of the 10 papers are open, namely [9], [13], [18], [19], [25], [28], [29], [30], [35] and [38].

In [9], the dataset was collected from YouTube. These collected raw comments were cleaned and manually labelled into abusive or non-abusive[11]. In [13], Graham et al. used the 2012 national child abuse and neglect data system child file data[12] for the research. In [18], the dataset is from the national health and social life survey[13] which included more than 3,000 men and women aged 18–59. In [19], Shaw et al. used the data from the national health and social life survey[14] with the sample size of over 1,500 U.S. participants.

In the papers [25], [28], [29] and [30], the authors used the PAN-2012 dataset ([44]) which contains a total of 222,055 conversations with nearly 3 million online messages from sexual predators and non-malicious user. The conversations have short text abbreviations, emoticons, slang, digits, symbols and character repetitions. They were collected from various sources, such as regular conversations without any sexual content and sexual conversations between consenting adults from Omegle[15]. In [35], the authors created a dataset based on 2,138 CSAM, adult pornography and sensitive images from a benchmark child pornography dataset built by [45]. Their data includes 836 CSA images and 285 adult pornography images which have annotations of body parts (e.g. head, breast and buttocks) and demographic attributes (e.g. age, gender, and ethnicity). In [38], authors used a dataset published by the National Software Reference Library[16] that contains more than 32 million file names.

C. Applied Methods

1) Statistical: Statistical analysis of large datasets is a common approach used to discern patterns and trends without bias, enabling us to discover meaningful information from large amounts of raw unstructured data. The 14 papers [12], [13], [15], [16], [17], [18], [19], [24], [33], [34], [37], [40], [42] and [43] used statistical methods to analyse their datasets.

In [12], each participant (in 100 fathers from Seoul and 102 parents from Novosibirsk) was supplied with a questionnaire form. Random effects regression models were then used to detect the relationship between child abuse and informal social control. In [13], the generalized linear mixed models were used to analyse the CSA reports from the child welfare system. In [15], the authors used bi-variate statistics and step-wise multiple logistic regression models.

Research works [16], [17], [18] and [19] used SPSS tool to analyse their datasets. Paper [16] exploited features about age, gender, pornography use, social media use and team sport participation. Paper [17] used features including scale of the online CSA problem; cybercrime and the avoidance of digital technology. In [18] and [19] the authors exploited features,

[10] https://gvis.unileon.es/datasets/

[11] https://github.com/shaheerakr/roman-urdu-abusive-comment-detector

[12] https://www.ndacan.acf.hhs.gov/datasets/datasets-list-ncands-child-file.cfm

[13] https://www.icpsr.umich.edu/web/HMCA/studies/6647

[14] https://www.cdc.gov/violenceprevention/childabuseandneglect/vacs/country-reports.html

[15] www.omegle.com

[16] https://www.nist.gov/itl/ssd/software-quality-group/national-software-reference-library-nsrl/about-nsrl/nsrl-introduction

namely unhealthy last year, unhappy last year, same-sex age groups, emotional problems interfered with sex, specific sexual problems and/or ever forced woman sexually.

In [24], the illegal online content reports of the Greek Hotline and the ATLAS dark web dataset of Web-IQ were analysed. The authors wanted to discover the relationship between the open web and the dark web. For example, they discovered that more than 50% of the Hotline reports were discussed in the dark web. In [33], the system applied a simple algorithm based on a threshold of returned values of sensors worn. The system could detect and prevent most acts of sexual abuse or assault against children by calling the parents, or taking a picture or producing an alarm. In [34], the authors studied, measured and analysed CSAM distribution and the exponential increment of CSAM reports. In [37], the data of an online survey and semi-structured interviews was analysed. Some information found included: (1) Location and types of abuse incidents in Bangladesh; (2) The relationship between the abuser and the victims; (3) Challenges in combating CSA.

Applying simple statistical methods, Owen and Savage [40] analysed the type and popularity of the content, and Woodhams et al. [43] analysed the characteristics and behaviours of anonymous users of dark web platforms. Westlake and Bouchard [42] studied the hyperlinks of the websites distributing CSAM.

2) Natural Language Processing: The NLP techniques can be used to extract features and contents of CSAM, e.g. abuse chats and grooming conversation. There are 2 papers [22] and [27] applying NLP to detect CSAM. In [22], the authors analysed patterns in the locations and folder/file naming practices of the CSA websites to detect non-hashed CSA images. The motivation for this is due to the increasing volume of CSA images being created and distributed an increasing percentage of images are not classified in hash value databases used for automatic detection. In [27], the list of keywords about emotions and sentiments posted in social media was built. Then, the authors analysed the Twitter tweets and Facebook pages of some potential people in India based on the keyword list.

3) Machine Learning (ML) & Deep Learning (DL): ML/DL techniques are often used to classify or cluster CSA text or images. The 14 papers [9], [10], [21], [23], [25], [26], [28], [30], [31], [35], [36], [38], [39] and [41] used and proposed ML/DL techniques in CSAM.

Some papers [21], [26], [23] and [35] applied deep Convolutional Neural Network (CNN) to detect CSA images. In [21], visual attention mechanism, age-group scoring and porn image classification were exploited. In [26], the data-driven concepts and characterisation aspects of images were used. In [23], the authors detected CSA from self-figure drawings. In [35], the model combined object categories, pornography detection, age estimation and image metrics (i.e. luminance and sharpness)

The work of [9] used DL models, namely CNN, Long Short Term Memory Network (LSTM), Bidirectional Long Short Term Memory (BLSTM) and Contextual LSTM. Ngejane et al. [25] used ML models, namely Logistic Regression (LR), XGBoost and Multilayer Perceptrons (MLP). In [28], [30], [31], [41] and [38], the authors exploited Bag of Words, TF-IDF, Word2Vec to Support Vector Machine (SVM), Naive Bayes, Decision Trees, LR, Random Forest and/or K-Nearest Neighbors.

Cecillon et al. [10] used the DeepWalk model and graph embedding representations. While in [36], the deep perceptual hashing algorithm based on the Apple NeuralHash was used. In [39], Hole et al. proposed that combining facial recognition with other biometric modalities, namely speaker recognition, can reduce both false positive and false negative matches, and could enhance analytical capabilities during investigations.

4) Combined methods: Both [29] and [32] combine ML/DL and NLP. In [29], Bours and Kulsrud applied CNN and Naıve Bayes based on features of message, author and conversation. While, Li et al. [32] used the Hierarchical Density based Spatial Clustering of Applications with Noise Algorithm (HDBSCAN) to cluster CSAM. The HDBSCAN applied the phrase detection algorithm to find the template signatures reuse between clusters.

IV. Results and Discussion

In this survey we synthesised the state-of-the-art in the online CSAM detection and prevention literature, with a focus on datasets utilised, tasks considered, and the analysis/modelling techniques applied. In this section we will discuss each of these categories in detail.

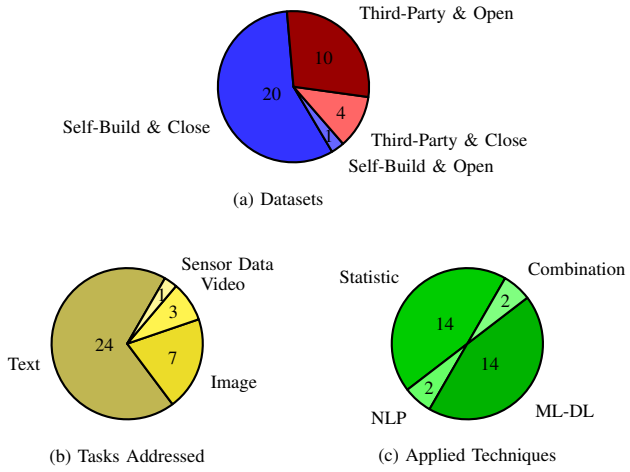

Fig. 1. The number of prior research are classified in terms of dataset, task addressing and applied technique

A. Datasets

Figure 1(a) presents the breakdown of the datasets used in the selected papers from the literature. The majority (approximately 57%) of the papers used self-generated and closed data, which is not publicly available to the research community. A slightly smaller percentage (approximately 29%) of the literature used datasets that were generated from a third-party such as social media and subsequently made publicly available

to the community. There are 7 open datasets created by third-parties. In that, 6 of them are CSA text and another is CSA images.

Typical pre-processing included the removal of personal data were applied before the information were extracted and exploited. Much smaller percentages of papers used datasets that were self-generated and then made available (approximately 3%) or which were generated by a third-party and kept private (approximately 11%).

B. Tasks Addressed

Figure 1(b) presents the breakdown of the type of task addressed in the selected papers from the literature. The tasks addressed focused on: (1) detecting sexual abuse comment or chat conversation; (2) understanding the abuser's behaviours; and (3) detecting CSA images based on porn content, face detection and age estimation.

The majority (approximately 69%) of the research analysed text-based CSAM. Of lower focus as image-based CSAM (approximately 20%), video (approximately 8%), with only 1 paper (approximately 3%) using sensor data. One explanation of this breakdown can be the very sensitive and illegal nature of image and video based CSAM, which requires special permission to access in many jurisdictions. Moreover, the disturbing nature of the content may be too traumatising for many of the research community (particularly true for technologists rather than specially trained sociologists) to willingly deal with graphic content of this nature. In terms of the low number of sensor data, this may be due to the relatively recent development of such devices and the novelty around the proposal to use sensor data to detect abusive activities.

C. Applied Analysis/Modelling Techniques

Figure 1(c) presents the breakdown of the analysis and modelling techniques proposed in the selected literature. In papers applying statistic methods, the authors often used information from questionnaires and national hotlines to exploit features about victims and/or abusers, such as age, gender, race, pornography usage and emotional problems interfered with sex. The ML/DL papers typically applied SVM, Naive Bayes and/or CNN. The list of key word was used in the NLP papers and combined with ML/DL models in the combination techniques.

The most popular techniques are statistic and ML/DL (same approximately 44%). NLP and combination approaches are much less common in the domain, with only approximately 6% of papers proposing these. One explanation of this breakdown can be that sociologists are familiar with statistical algorithms and tools more than NLP and DL/ML algorithms and tools. While, comparing between NLP and DL/ML, DL/ML processes CSA images and videos more effective than NLP. Moreover, NLP tools are more difficult to use than ML/DL tools (even with computer scientist) in applying on CSAM (even with CSA text). So, the researchers often use statistical and ML/DL on processing and detecting CSAM.

V. Conclusion

This paper presented a comprehensive survey about the investigation, detection and prevention of online CSAM. We focused on research works in domain of CSA published by Springer, Elsevier, IEEE, or ACM from 2018 to 2022 to identify the relevant literature. The techniques used in the research papers include statistics, NLP, and ML/DL, applied to addressing CSAM problems across text, image and video content. The types of research datasets and their availability was also addressed. This information gives us a broad perspective on applying statistic and AI methods for solving the problems in domain of CSA on social networks. In addition, the open datasets in CSAM were reviewed and presented.

The survey paper can be a valuable resource for the CSAM research community as it provides the scope and size of the literature in the domain since 2018. This work can help researchers identify the gaps in knowledge and potential datasets that can be utilised for analysis or validation of proposed solutions.

CSAM detection and reduction are a complex task. In the future, the multiple methods need to be used in creating a combined method which may be combinations of deep learning methods and NLP technologies. In addition, to evaluate the CSAM detection models, a benchmark dataset must be created. We will re-use open CSAM datasets and collect more CSAM in hotlines and dark webs.

Acknowledgement

The survey paper is a part of the N-Light project which is funded by the Safe Online Initiative of End Violence and the Tech Coalition through the Tech Coalition Safe Online Research Fund (Grant number: 21-EVAC-0008-Technological University Dublin).

References

[1] H. Lee, T. Ermakova, V. Ververis, and B. Fabian, "Detecting child sexual abuse material: A comprehensive survey," *Forensic Science International: Digital Investigation*, vol. 34, p. 301022, 2020.

[2] L. Sanchez, C. Grajeda, I. Baggili, and C. Hall, "A practitioner survey exploring the value of forensic tools, ai, filtering, & safer presentation for investigating child sexual abuse material (csam)," *Digital Investigation*, vol. 29, pp. S124–S142, 2019.

[3] S. Ali, H. A. Haykal, and E. Y. M. Youssef, "Child sexual abuse and the internet—a systematic review," *Human Arenas*, pp. 1–18, 2021.

[4] L. S. Christensen, S. Rayment-McHugh, T. Prenzler, Y.-N. Chiu, and J. Webster, "The theory and evidence behind law enforcement strategies that combat child sexual abuse material," *International Journal of Police Science & Management*, vol. 23, no. 4, pp. 392–405, 2021.

[5] D. Russell, D. Higgins, and A. Posso, "Preventing child sexual abuse: A systematic review of interventions and their efficacy in developing countries," *Child abuse & neglect*, vol. 102, p. 104395, 2020.

[6] M. N. Slavin, A. A. Scoglio, G. R. Blycker, M. N. Potenza, and S. W. Kraus, "Child sexual abuse and compulsive sexual behavior: A systematic literature review," *Current addiction reports*, vol. 7, no. 1, pp. 76–88, 2020.

[7] C. M. Steel, E. Newman, S. O'Rourke, and E. Quayle, "A systematic review of cognitive distortions in online child sexual exploitation material offenders," *Aggression and violent behavior*, vol. 51, p. 101375, 2020.

[8] C. M. Steel, E. Newman, S. O'Rourke, and E. Quayle, "An integrative review of historical technology and countermeasure usage trends in online child sexual exploitation material offenders," *Forensic Science International: Digital Investigation*, vol. 33, p. 300971, 2020.

[9] M. Akhter, Z. Jiangbin, I. Naqvi, M. AbdelMajeed, and T. Zia, "Abusive language detection from social media comments using conventional machine learning and deep learning approaches," *Multimedia Systems*, pp. 1–16, 2021.

[10] N. Cecillon, V. Labatut, R. Dufour, and G. Linares, "Graph embeddings for abusive language detection," *SN Computer Science*, vol. 2, no. 1, pp. 1–15, 2021.

[11] L. S. Christensen and K. Pollard, "Room for improvement: How does the media portray individuals who engage in material depicting child sexual abuse?" *Sexuality & Culture*, pp. 1–13, 2022.

[12] C. R. Emery, S. Wu, T. Eremina, Y. Yoon, S. Kim, and H. Yang, "Does informal social control deter child abuse? a comparative study of koreans and russians," *International journal on child maltreatment: research, policy and practice*, vol. 2, no. 1, pp. 37–54, 2019.

[13] L. M. Graham, P. Lanier, M. Finno-Velasquez, and M. Johnson-Motoyama, "Substantiated reports of sexual abuse among latinx children: Multilevel models of national data," *Journal of family violence*, vol. 33, no. 7, pp. 481–490, 2018.

[14] K. Guastaferro, K. M. Zadzora, J. M. Reader, J. Shanley, and J. G. Noll, "A parent-focused child sexual abuse prevention program: Development, acceptability, and feasibility," *Journal of child and family studies*, vol. 28, no. 7, pp. 1862–1877, 2019.

[15] L. S. Jonsson, C. Fredlund, G. Priebe, M. Wadsby, and C. G. Svedin, "Online sexual abuse of adolescents by a perpetrator met online: a cross-sectional study," *Child and adolescent psychiatry and mental health*, vol. 13, no. 1, pp. 1–10, 2019.

[16] M. K. Maas, K. M. Cary, E. M. Clancy, B. Klettke, H. L. McCauley, and J. R. Temple, "Slutpage use among us college students: the secret and social platforms of image-based sexual abuse," *Archives of sexual behavior*, vol. 50, no. 5, pp. 2203–2214, 2021.

[17] E. Quayle, "Prevention, disruption and deterrence of online child sexual exploitation and abuse," in *Era Forum*, vol. 21, no. 3. Springer, 2020, pp. 429–447.

[18] B. Rind, "First postpubertal male same-sex sexual experience in the national health and social life survey: Current functioning in relation to age at time of experience and partner age," *Archives of Sexual Behavior*, vol. 47, no. 6, pp. 1755–1768, 2018.

[19] S. Shaw, H. J. Cham, E. Galloway, K. Winskell, Z. Mupambireyi, C. Kasese, Z. Bangani, and K. Miller, "Engaging parents in zimbabwe to prevent and respond to child sexual abuse: A pilot evaluation," *Journal of Child and Family Studies*, vol. 30, no. 5, pp. 1314–1326, 2021.

[20] K. Borg, C. Snowdon, and D. Hodes, "A resilience-based approach to the recognition and response of child sexual abuse," *Paediatrics and child health*, vol. 29, no. 1, pp. 6–14, 2019.

[21] A. Gangwar, V. González-Castro, E. Alegre, and E. Fidalgo, "Attm-cnn: Attention and metric learning based cnn for pornography, age and child sexual abuse (csa) detection in images," *Neurocomputing*, vol. 445, pp. 81–104, 2021.

[22] E. Guerra and B. G. Westlake, "Detecting child sexual abuse images: traits of child sexual exploitation hosting and displaying websites," *Child Abuse & Neglect*, vol. 122, p. 105336, 2021.

[23] L. Kissos, L. Goldner, M. Butman, N. Eliyahu, and R. Lev-Wiesel, "Can artificial intelligence achieve human-level performance? a pilot study of childhood sexual abuse detection in self-figure drawings," *Child Abuse & Neglect*, vol. 109, p. 104755, 2020.

[24] E. Kokolaki, E. Daskalaki, K. Psaroudaki, M. Christodoulaki, and P. Fragopoulou, "Investigating the dynamics of illegal online activity: The power of reporting, dark web, and related legislation," *Computer Law & Security Review*, vol. 38, p. 105440, 2020.

[25] C. H. Ngejane, J. H. Eloff, T. J. Sefara, and V. N. Marivate, "Digital forensics supported by machine learning for the detection of online sexual predatory chats," *Forensic science international: Digital investigation*, vol. 36, p. 301109, 2021.

[26] P. Vitorino, S. Avila, M. Perez, and A. Rocha, "Leveraging deep neural networks to fight child pornography in the age of social media," *Journal of Visual Communication and Image Representation*, vol. 50, pp. 303–313, 2018.

[27] D. Andrews, S. Alathur, N. Chetty, and V. Kumar, "Child online safety in indian context," in *2020 5th International Conference on Computing, Communication and Security (ICCCS)*. IEEE, 2020, pp. 1–4.

[28] P. R. Borj, K. Raja, and P. Bours, "Detecting sexual predatory chats by perturbed data and balanced ensembles," in *2021 International Conference of the Biometrics Special Interest Group (BIOSIG)*. IEEE, 2021, pp. 1–5.

[29] P. Bours and H. Kulsrud, "Detection of cyber grooming in online conversation," in *2019 IEEE International Workshop on Information Forensics and Security (WIFS)*. IEEE, 2019, pp. 1–6.

[30] M. A. Fauzi and P. Bours, "Ensemble method for sexual predators identification in online chats," in *2020 8th international workshop on biometrics and forensics (IWBF)*. IEEE, 2020, pp. 1–6.

[31] M. M. Islam, M. A. Uddin, L. Islam, A. Akter, S. Sharmin, and U. K. Acharjee, "Cyberbullying detection on social networks using machine learning approaches," in *2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*. IEEE, 2020, pp. 1–6.

[32] L. Li, O. Simek, A. Lai, M. Daggett, C. K. Dagli, and C. Jones, "Detection and characterization of human trafficking networks using unsupervised scalable text template matching," in *2018 IEEE International Conference on Big Data (Big Data)*. IEEE, 2018, pp. 3111–3120.

[33] S. Samra, M. Alshouli, S. Alaryani, B. Aasfour, W. Shehieb, and M. Mir, "Shield: Smart detection system to protect children from sexual abuse," in *2021 13th Biomedical Engineering International Conference (BMEiCON)*. IEEE, 2021, pp. 1–6.

[34] E. Bursztein, E. Clarke, M. DeLaune, D. M. Elifff, N. Hsu, L. Olson, J. Shehan, M. Thakur, K. Thomas, and T. Bright, "Rethinking the detection of child sexual abuse imagery on the internet," in *The world wide web conference*, 2019, pp. 2601–2607.

[35] C. Laranjeira, J. Macedo, S. Avila, and J. Santos, "Seeing without looking: Analysis pipeline for child sexual abuse datasets," in *Proceedings of 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT'22)*. ACM, 2022, p. 2189–2205.

[36] L. Struppek, D. Hintersdorf, D. Neider, and K. Kersting, "Learning to break deep perceptual hashing: The use case neuralhash," in *2022 ACM Conference on Fairness, Accountability, and Transparency*, 2022, pp. 58–69.

[37] S. Sultana, S. T. Pritha, R. Tasnim, A. Das, R. Akter, S. Hasan, S. R. Alam, M. A. Kabir, and S. I. Ahmed, "'shishushurokkha': A transformative justice approach for combating child sexual abuse in bangladesh," in *CHI Conference on Human Factors in Computing Systems*, 2022, pp. 1–23.

[38] M. W. Al-Nabki, E. Fidalgo, E. Alegre, and R. Alaiz-Rodríguez, "Short text classification approach to identify child sexual exploitation material," *arXiv preprint*, 2020.

[39] M. Hole, R. Frank, K. Logos, B. Westlake, D. Michalski, D. Bright, E. Afana, R. Brewer, A. Ross, T. Swearingen *et al.*, "Developing automated methods to detect and match face and voice biometrics in child sexual abuse videos," *Trends and Issues in Crime and Criminal Justice [electronic resource]*, no. 648, pp. 1–15, 2022.

[40] G. Owen and N. Savage, "The tor dark net," *Chatham House*, 2015.

[41] M. Pereira, R. Dodhia, H. Anderson, and R. Brown, "Metadata-based detection of child sexual abuse material," *arXiv preprint arXiv:2010.02387*, 2020.

[42] B. G. Westlake and M. Bouchard, "Liking and hyperlinking: Community detection in online child sexual exploitation networks," *Social science research*, vol. 59, pp. 23–36, 2016.

[43] J. Woodhams, J. A. Kloess, B. Jose, and C. E. Hamilton-Giachritsis, "Characteristics and behaviors of anonymous users of dark web platforms suspected of child sexual offenses," *Frontiers in Psychology*, vol. 12, p. 623668, 2021.

[44] G. Inches and F. Crestani, "Overview of the international sexual predator identification competition at PAN-2012," in *CLEF 2012 Evaluation Labs and Workshop, Online Working Notes*, vol. 1178. CEUR-WS.org, 2012.

[45] J. Macedo, F. Costa, and J. A. dos Santos, "A benchmark methodology for child pornography detection," in *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, 2018, pp. 455–462.