Conceptual Model for Crowd-sourcing Digital Forensic Evidence

Stacey O. Baror,[†] H. S. Venter[*], Victor R. Kebande[z]

[†*]University of Pretoria, Pretoria, South Africa
[†*]stacey.baror@cs.up.ac.za,[*]hventer@cs.up.ac.za
[z]Blekinge Institute of Technology, karlskrona, Sweden
[z]victor.kebande@bth.se

Abstract. COVID-19 scourge has made it challenging to combat digital crimes due to the complexity of attributing potential security incidents to perpetrators. Existing literature does not accurately pinpoint relevant models/frameworks that can be leveraged for crowd-sourcing digital forensic evidence. This paper suggests using feature engineering approaches for crowd-sourcing digital evidence to profile potential security incidents, for example, in a COVID-19 scenario. The authors have proposed a conceptual Crowd-sourcing (CRWD) model with three main components: Forensic data collection, feature engineering and application of machine learning approaches, and also assessment with standardized reporting. This contribution is significantly poised to solve future investigative capabilities for forensic practitioners and computer security researchers.

Keywords: Crowd-sourcing, citizen-media · Digital forensics · Digital evidence · COVID-19

1 Introduction

Digitalization, technological advancements and the rise of cyber-related incidents has meant it is vital to address the need for attribution [1][2][3]. Currently, managing these kinds of tasks across heterogeneous environments, where there is a plethora of data with many features is seen to be a key challenge for research practitioners. That notwithstanding, the emergence of the COVID-19 scourge with associated conspiracies has been capitalized by adversaries, where it has become apparent that adversaries can camouflage to perpetuate digital crimes.

Despite the earlier advancements and the need for long-lasting attribution solutions, the global media industry has faced critical challenges. The main challenge has been the shortage of on-sight reporters due to several months of lock-downs or restricted movements. As a result, it may be essential to practice citizen or community journalism. “Citizen journalism, also known as collaborative media, or street journalism, is based upon public citizens ”playing an active role in the process of collecting, reporting, analyzing, and disseminating news and information" [1]. Tracking or monitoring incidents based on personal traits, characteristics, locations or collecting aspects of digital data that can be used in hypothesis creation is positioned to offer a practical approach towards combating cybersecurity incidents albeit from post-event response approach [4][5][6].

In this paper, the authors propose a Crowd-Sourcing (CRWD) model that can be used to identify crucial and relevant digital evidence based on heterogeneous sources. The objective of the CRWD model is to show that based on the existing features of digital information, it can be possible to use feature engineering approaches coupled with machine learning techniques to solve attribution challenges in the post-event response strategy. Furthermore, the study contributes to a contextual discussion based on how the perspective of the CRWD model has been presented. Specifically, this study intends to achieve the following:

- Propose a generic Crowdsourcing (CRWD) model and show the relevance across heterogeneous environments by relying on a COVID-19 scenario as a baseline
- Provide a contextual evaluation of the CRWD model and assess the degree of its influence in the digital forensic community

The remainder of the paper is structured as follows: Section 2 provides background study while Section 3 discuss Related Work. After this, Section 4 gives the scenario while Section 5 discuss the proposed digital forensic evidence crowdsourcing model. A discussion on the propositions is given in Section 6, while Section 7 concludes the paper with a mention of future work.

2 Background

2.1 Digital forensics

Digital forensic science is drawn from the traditional science of forensics developed in conjunction with the biological sciences and has undergone a continuous growth that encompasses all digital devices [7] [8][9][10][11]. It mainly emerged from forensic science, which is a larger body of knowledge in which science is used to solve crime [12] [13][14]. Furthermore, it has been defined as the use of *"scientifically derived and proven mathematical methods toward the preservation, collection, validation, identification, analysis, interpretation, documentation and proper expertise witness presentation of digital evidence, derived from digital sources to facilitate or further the reconstruction of events found to be criminal", or helping to anticipate unauthorized actions"* [15] [7]. Other definitions conclude that digital forensics [16] is concerned with the investigation of any suspected crime or misbehaviour that may be manifested by digital evidence. For example, Sibiya et al., [17] looks at digital forensics as a discipline that combines elements of law and computer science to collect and analyze data from computer systems, networks, wireless communications, and storage devices in a way that is admissible as evidence in a court of law. Studies have mainly tried to map digital forensics to contact tracing strategies to enhance electronic discoveries. For example, authors [18][19][20][6][21], have proposed different proactive techniques, which in the context of this study could also be mapped to how concurrent contact tracing strategies could be accomplished. On the other hand, digital evidence constitutes relevant facts that are being investigated, which in their entirety needs to be reliable and with the highest form of integrity as highlighted by ISO/IEC 27043; ISO/IEC 27037 for purposes of admissibility [22] [7]. This is owing to the fact that it is digital evidence that has to be presented to support or refute forensic hypotheses during litigation.

2.2 Crowd sourcing

Crowd-sourcing is a sourcing process of collecting and gathering information, opinions, media content from a group of people, usually sourced to achieve a certain goal [23][24]. Crowd-sourcing has been in existence since the stone age; however, it recently gained popularity in the advent of the internet and social media technology evolution. Crowdsourcing typically involves using online connectivity to attract participants and divide the task objective to achieve a cumulative result.

Organizations have employed the use of crowdsourcing to share and publish research, especially when the organization sourcing out the skills do not possess the relevant resources or skills required to solve a specific research problem [8][25][24] [23]. Furthermore, some crowdsource endeavours require a situation where the potential problem-solvers receive financial incentives. The challenge, however, with crowdsourcing is such that the owner of the intellectual property becomes contextual.

Another form of crowdsourcing involves the situation where the crowd's creativity which is usually based on the individual's domain knowledge, are drawn together to build the basis for a product, example of a crowdsourcing endeavour of this sort is the *wikipedia.org.* Another example is crowdsourcing, where monetary (finance) is the tangible 'object' being sourced. The crowd creates value for the general public in reaching the set goal for the source. This is one of the most common crowdsources in 'social media. Furthermore, it is notable that crowdsourcing proceeds without a company in the background to make profits or donations only.

2.3 Feature Engineering

Feature engineering transforms raw data into features that represent the underlying problem to the predictive models. It is a representation of real-world modelling problem [26][27][28]. Additionally, feature engineering takes existing domain knowledge and extract characteristics, properties, attributes from raw data to develop a tool, process or framework that could mimic a real-world scenario while employing the machine learning models. The extracted and identified features are used for predictive models. For example, a study by Faris et al., [29] classified the spam email features into three main categories: (i) header features, (ii) payload (body) features, and (iii) attachment features. Later on, this study aimed to develop the spam detection models using the most common spam features and create a simple open-source tool for extracting email features. The importance of that study is that their study showed that the features improved spam detection rates based on various machine learning algorithms. The same approaches have possibilities of being replicated in a forensic crowd-sourcing scenario; however, the authors of this paper acknowledge the insights by the other authors.

3 Related Work

The authors have conducted a study based on previous works that have addressed different contact tracing approaches, and this has been shown in Table 1. These studies have employed various approaches, and in each, a number of challenges exist, which we use against our framework.

As shown in Table 1 various studies explored have addressed the application of feature engineering to information security, teaching and learning and crowd-sourcing. The authors have not explored the options of the Crowd-Sourcing (CRWD) model that identifies vital and relevant digital forensic evidence based on heterogeneous sources. The CRWD model shows that the existing triage of digital forensic evidence soundness features can be possible to use feature engineering approaches coupled with machine learning approaches to solve attribution challenges in post-event response.

Table 1. Related work and challenges of the proposed Crowd-sourcing DFE & Profiling Security Incidents using Feature Engineering Techniques

REF	Focus	Findings	Challenges
[2]	Crowdsourcing Cybersecurity: Cyber Attack Detection using Social Media	The paper proposed the use of social media as a crowdsourced sensor to understand and detects a broad range of cyber-attacks using supervised learning techniques	The use of supervised learning narrows the study
[30]	Detecting Android malware with intensive feature engineering	Detecting android malware that focuses on the executable file (classes.dex), resource and the abstraction of the APK application using these as the feature engineering objects at the single level	Potential of inadequacy inaccuracy and false alarm rate
[31]	A Digital Forensics Triage methodology based on feature manipulation techniques	A methodology to automate the digital evidence of device classification process for crime-related that deals with feature weighting quantified using Kullback-Leibler measure to incorporate Machine Learning principles while improving classification accuracy through feature manipulation.	The solution lack generality as a crime template must be chosen before the device is searched for evidence.
[29]	Improving email spam detection using content-based feature engineering approach	A developed open-source tool that provides a flexible way to extract a large number of features from any email corpus. The tool extracted 140 features of SpamAssassin email corpus.	Spam email features could influence the overall outcome of different spam corpora.
[32]	A Learning to Rank Framework for Developer Recommendation in Software Crowdsourcing	The paper proposed a CRF model to build a successful software crowdsourcing platform. The CRF model effectively uses features that recommend tasks to the crowd workers by extracting feature criteria such as topic-based, skills and locations.	The feasibility and real-world application could prove to be ineffective.
[28]	Data Analytics of Crowdsourced Resources for Cybersecurity Intelligence	The author proposed a methodology that collects data, representing the data using and offer security recommendations based on collective data of social media community to monitor vulnerabilities	The authors identified the challenges associated with using an agile approach to point out cyberattacks risks.
[21]	A natural human language framework for digital forensic readiness in the public cloud	The authors proposed the use of natural human language interaction as a unique feature identifier to detect cybercrime attacks in the public cloud.	Feature identification for in-progress is sometimes hectic to detect cybercrime.

4 CRWD Model

This section gives an illustration of the proposed CRWD model. Firstly, a discussion on the CRWD model assumptions is given, followed by a high-level overview of the model, and later on, a detailed representation of the CRWD model will become apparent.

4.1 Assumptions

The CRWD model uses a coordinated approach to achieve its objectives, where diverse tasks can collectively identify and track potential sources of digital evidence. This paper focuses mainly on showing how the coordination of different tasks can help establish critical crowd-sourcing strategies that can be used to detect potential security incidents. Furthermore, it is essential to mention that this paper is not entirely inclined towards a real-life implementation. However, it gives key conceptual approaches that are needed for the suggested CRWD model. Therefore, the proposed CRWD models make the following assumptions:

- **Assumption 1: -** This study assumes that digital evidence is streaming from diverse locations. The probability of an exact identification or attribution of a digital object to a perpetrator and an origin may require extensive feature engineering approaches to make potential judgments.

- **Assumption 2: -** Another assumption of the study is: Each task accomplished is linked to each other, and the output of one provides a significant input to the other.

Apart from the assumptions mentioned above, based on the digital evidence requirements and guidelines mentioned in the rules of evidence, we consider that the standard process of evidence admissibility only requires integrity to be ensured yet. Table 2 highlights some of the components considered when the forensic sound of potential digital evidence is in question. However, the application of feature engineering approach [29][31] to the information security services consists of the components employed in identifying and upholding the validity of potential digital evidence. From an information security perspective, maintaining the forensic soundness of collected digital evidence is paramount. The collected digital evidence is ascertained to be forensically sound when it is an original, reliable chain of evidence and follows the rule of evidence [33].

Table 2. The Digital forensic soundness components

S/n	DF soundness components	Mechanisms to address (CIAAN)
i	Confidentiality (C)	Symmetric Encryption + Assym Encryption +Username and Password
ii	Integrity (I)	Hashing
iii	Authorisation (A)	Access control (ACL), Role allocation
iv	Authentication (A)	+Username and Password
vi	Non-repudiation	Digital certificate + digital signature + Asymmetric encryption

4.2 Scenario

While COVID-19 has ravaged some sectors, cyber-security related cases have sprung up as a result as well. Basically, in every connected environment, cybersecurity is seen to have mostly negative connotations engulfed with uncertainties, fear and a projection of detrimental effects. We consider a COVID-19 'anti-lockdown demonstration' scenario where adversaries can pose as regular citizens to commit digital crimes. Triangulating to detect such adversaries may be a tedious and time-consuming approach, where one may filter exactly what they are looking for. As a result, it is possible for an adversary posing as a demonstrator to go undetected. In this context, profiling adversaries to collect digital data based on contact-tracing devices, voice, biometrics, facial recognition and other human-based forensic security techniques may provide a step towards sourcing valid digital forensic evidence that may be used for forensic hypothesis creation.

4.3 CRWD Model: High-Level View

Figure 1 illustrates a high-level view proposed CRWD model. The CRWD consists of three main components that are represented as phases (labelled 1 to 3), namely 1) Forensic Data Collection (Phase 1), Feature Engineering and Machine Learning Application (Phase 2) and Assessment and Reporting (Phase 3), respectively. Data collection is a phase used to crowd-source digital evidence from a particular scenario in order to form a forensic dataset. After this, feature engineering techniques are used to map essential aspects needed to judge what may be vital to be trained by machine learning algorithms to extract key events that can be used to form a strong hypothesis.

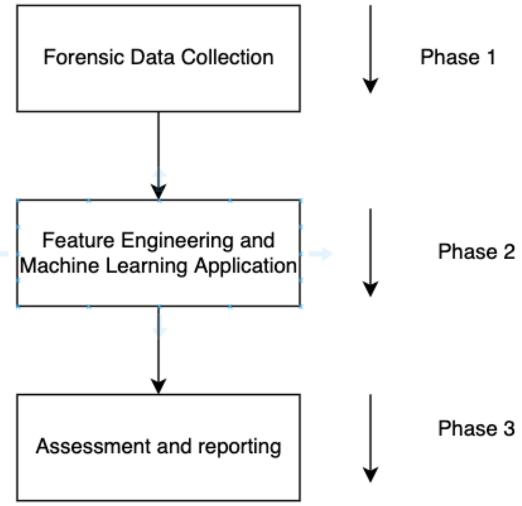

Fig. 1. Overview of combined multiple functionalities

5 All-Inclusive CRWD Model Steps

Phase 1: Forensic Data Collection Figure 2 shows Phase 1 of the CRWD model, which consists of four distinct steps as follows: Scenario (1), Evidence Collection Guidelines (2), Forensic data Collection and Forensics database. Given the complexity involved in crowd-sourcing digital evidence, this phase attempts to achieve investigative approaches by tracking, monitoring, and collecting critical aspects of digital data that can be used to create digital forensic hypotheses. This is mainly a precise process that collects specific data types that can easily be mapped to existing contents in the database. For example, in obtaining data, the following strategies could be used: Facial recognition algorithms, using contact-tracing data from connected mobile devices, mugshots, Closed-Circuit Television (CCTVs), voice recognition, biometrics or other significant human-based traits. The ultimate goal of this phase is to apply scientific-based techniques based on human-based forensics techniques and standardized digital evidence collection approaches to solving critical forensic attribution challenges. It is also observed that this phase has the objective of gathering data with a degree of high accuracy

- **Step 1: Scenario -** Figure 1, the scenario projected in this context has been from a COVID-19 (See Section 4.2) where multiple parties-where target data can be harvested based on evidence collection guidelines, which are discussed in the next phase.
- **Step 2: Evidence Collection Guidelines -** The rules that govern how digital/electronic evidence is collected or handled emphasizes key aspects as

follows: Authenticity, admissibility, reliability, completeness and believability. Other relevant aspects include the expert witness testimony that requires the use of scientific methods that could be used to show knowledge and understanding of matters that deals with digital data from the technical point of view.

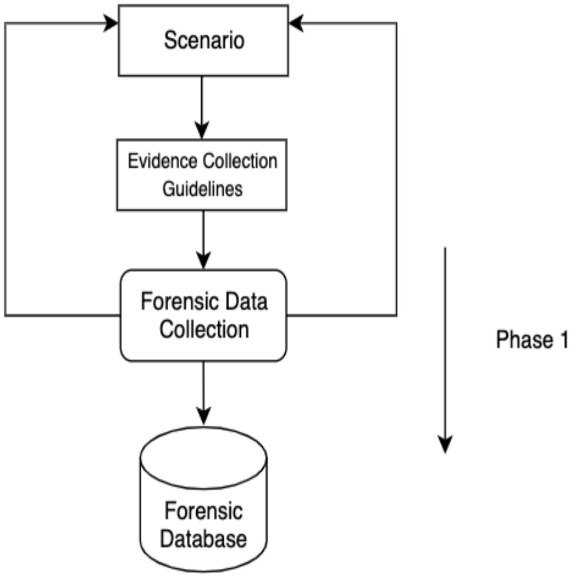

Fig. 2. Depiction of forensic data collection approach-A major process of creating a forensic dataset in a readiness approach

This study explicitly emphasises the need for collecting only data with relevant attributes that can easily be used to profile adversaries. For example, facial recognition, specific proportions, coordinates and contacts by digital devices as highlighted by Baror et al.,[34].

Phase 2: Feature Engineering and Machine Learning Application Strategies that allow security incident detection in this context can be achieved more quickly and optimized minimally, as is shown in Figure 3. This is possible when feature engineering and machine learning methods are applied to the extracted forensic dataset. In the context of this paper, we suggest that it is essential to conduct data preprocessing based on the crowdsourced evidence that is collected in (Phase 1) of Figure 2.

- **Step 1: Feature Engineering -** Basically, this allows a set of essential features to be identified from a given domain in order to make judgement

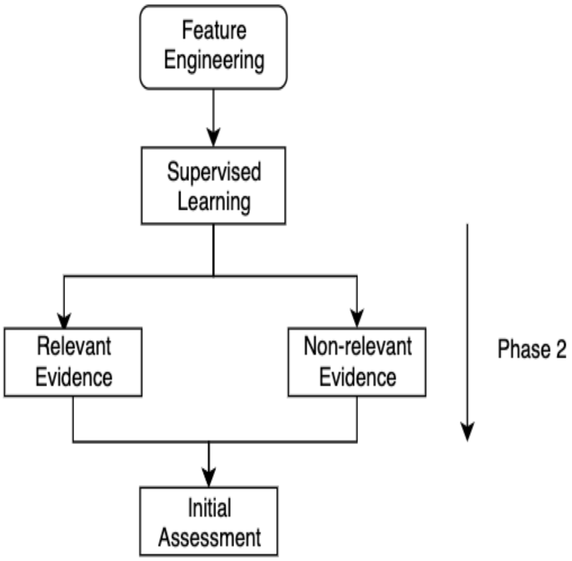

Fig. 3. Applying feature engineering process and machine learning applications to identify the relevance of digital data/forensic evidence

whether the outcome could be trained in order to identify key events that can be used for digital investigation processes. Essential features are adjusted to remove unnecessary variables then standard scaling is applied to normalize the features to generate statistical scores as is shown in the feature engineering pipeline that is shown in Figure 4.

- **Step 2: Supervised learning -** In this study, we consider supervised learning as a suitable approach that could be applied where selected classifiers can be trained in order to assess the behavior that is exhibited based on the identified significant features. Consequently, this study does not explicitly identify specific machine learning algorithms that must be used. However, critical supervised learning algorithms like Support Vector Machines (SVM), Random Forest (RF) or Naive Bayes could be applied. This extends to other key algorithms depending on the features that are identified.

- **Step 3: Relevant and non-relevant Digital Evidence -** To ascertain whether digital evidence has some relevance is a key step to generating stronger forensic hypothesis, forensic soundness and also admissibility. In addition, it is important to create a logical connection on the crowd-sourced evidence so that it can be relied upon to prove or disprove facts. Also, non relevant evidence is valid evidence in all circumstances. However, it may not be admitted as admissible evidence in some circumstances. In order to gen-

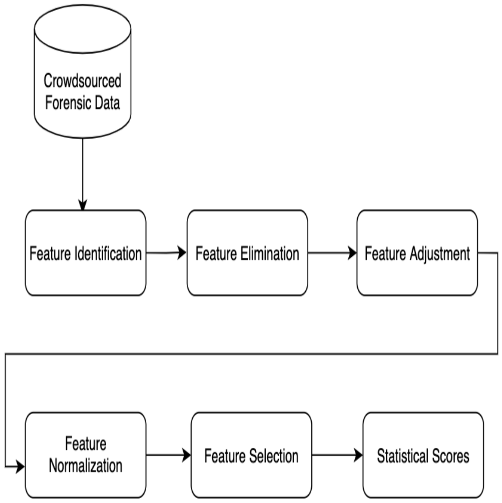

Fig. 4. Feature Engineering Pipeline and actions on crowdsourced Forensic Data

erate approaches that can allows profiling incidents as suggested by Kebande [11][35][6], one needs to consider relevant digital evidence.

- **Step 3: Initial Assessment -** In this phase, key digital forensic activities are examined to see if they can establish facts that can be used in the formation of a forensic hypothesis. Assessment in this context is not considered a closure phase because emergent digital artefacts could be integrated into the process in many circumstances.

Phase 3: Assessment and Reporting Figure 5 illustrated strategies for assessment and reporting. Incident detection and identification process follows the guidelines that have been mentioned in ISO/IEC 27043: 2015 standards [9][36][37]. The key aspects in this context identify sources of digital evidence, which are mainly vital to incident detection. In this study, we have considered incident detection and identification as a post-event response strategy. This context focuses on generating modalities of profiling security incidents, assessing whether those incidents are authentic and reliable.

- **Step 1: Incident detection and Identification -** This strategy is basically an incident response strategy considered to be a process that is used to assess incidents based on the existing standards, policies and procedures and it show if an incident really qualifies to be an incident. In most cases, this could be a security breach, threat or a potential attack.

Step 2: Profiling incidents, Authenticity, Integrity and Reliability Assessments Profiling security incidents entails identifying incidents based on the

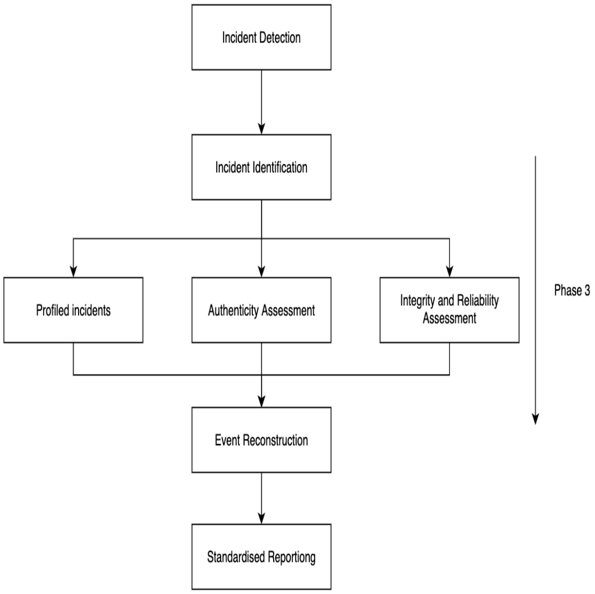

Fig. 5. Assessment and reporting phases

specified features. This is cross-checked based on the severity, impact and linkage metrics that those incidents may have at the time of detection. Modalities on how incidents could be profiled to the respective metrics have been highlighted by Kebande [38] and the aspects of classifying the type and severity of those incidents have also been suggested by studies in [39] where the CVSS scores have been used to assess the type of incidents mainly threats, vulnerabilities or attacks.

In most cases, a potential incident may go undetected if checks as to whether they are authentic or not being met. An examination of whether incidents are genuine or not needs to be accommodated in order for the incident to be linked to the crime or a suspect-also this increases chances of admissibility.

Based on the digital evidence guidelines, rules and practices like ACPO, ISO/IEC 27043 and the Federal Rules of Evidence, it is essential to create cryptographic hashes to retain every bit of collected digital information in it's original form. In order to increase assurance of integrity and reliability, this practice must be adopted.

Step 3: Event Reconstruction - In this phase, which also has been mentioned in the digital forensic practices and models, a timeline of how events transpired is pieced together to allow a proper investigative process like the chain of custody to prevail, which is also a key ingredient for forensic soundness. Event reconstruction is a key process for admitting crowd-sourced digital evidence and the validity of security incidents.

Step 4: Standard Reporting - While this process precedes before investigation closure, it is of importance to generate acceptable and standardized reports. As it has been mentioned by Authors [36][37][11] standardized reporting is an essential aspect that increases the admissibility and acceptability of digital forensic evidence.

6 Discussions

Cybersecurity weaknesses and aggravated cybercrime have recently seen some digital crimes go unsolved, especially with the constantly increasing number of devices and data complexity and primary anti-forensic tools. Based on the scenario that has been explained in this paper. Fighting digital crimes in COVID-19 complicates pre-investigative and post-investigative strategies and extends the attack surface. However, we note that crowd-sourcing digital evidence based on tracking, tracing and identifying human subjects provides a stepping stone towards solving the ever-complex attribution puzzle. We have explicitly specified the need for profiling key security incidents based on feature engineering approaches-which is vital in extracting and mapping events that helps in post-incident/reactive forensics. Based on the assumption of the CRWD model illustrated in Figures 1 to 5, it is essential to note that linkage and data may be the key to proving facts during incident response. The CRWD model stands as a suitable approach. It is a step that encapsulates strategies mentioned in standardized guidelines, including expert witness testimonies, as Daubert vs Merryl principles highlight. Consequently, from a generic view, our propositions forensically collect and allow machine-based learning approaches, which is vital in allowing system models to utilize forensic data to arrive at key decisions. Important to note is also the fact that, even though other models may have a relatively close scope, our proposition stands out because it can easily be extended or applied in practically real-life scenarios if implemented in a pandemic scenario.

Based on the intuitions that have been highlighted in the scenario that has been highlighted in (Section 4.2), the authors deduce that there may be a need to explicitly crowd-source relevant evidence that may attribute a piece of given evidence to a digital crime. In a reactive forensics approach, a given investigation process involving digital forensic experts and law enforcement agencies may find the CRWD model's approach quite relevant when it is imperative to match potential security incidents to the perpetrator. For example, if P is a perpetrator and X has to prove and map that a crime C is entirely attributed to P, it would be necessary for X to align the occurrences based on the sequences (Phase 1 to 3) CRWD model. This process would typically allow the interaction of phases while maintaining the standard investigation guidelines simultaneously. Other concurrent processes mentioned as part of investigation processes, event reconstruction, and how the chain of custody is maintained also holds in this case. And thus, it is worth noting that adopting the CRWD processes may expedite and promptly give to some degree a very effective approach-based alignment with standardized guidelines. It is also less likely that the CRWD model would accelerate the discovery of the incident process, which is a concern that we admit as a constraint although from a preliminary review. However, verification and validation of the CRWD model may prove or disprove this assertion, which the study is keen on. That notwithstanding, it is clear that this constraint may depend on the investigative situation and crime complexity. In the context of our approach, we argue from a generic perspective. Still, it is worth noting that this constraint is currently is viewed from a generic perspective.

7 Conclusion and future work

In this paper, we have proposed a CRWD model, which, based on its representation, conducts crowd-sourcing of digital evidence as a step towards profiling potential security incidents. The suggested approach has three significant steps: Forensic data collection, feature engineering, machine learning application, assessment and standardized reporting. Possible future work for the CRWD model is to test it against possible constraints and develop a verifiable prototype that can conduct the propositions as mentioned earlier in real-time and in an adaptive manner.

References

1. D. Jemielniak and A. Przegalinska, *Collaborative society.* MIT Press, 2020.
2. R. P. Khandpur, T. Ji, S. Jan, G. Wang, C.-T. Lu, and N. Ramakrishnan, “Crowdsourcing cybersecurity: Cyber attack detection using social media,” in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pp. 1049–1057, 2017.
3. I. R. Adeyemi, S. Abd Razak, and M. Salleh, “Understanding online behavior: Exploring the probability of online personality trait using supervised machine-learning approach,” *Frontiers in ICT*, vol. 3, p. 8, 2016.
4. R. Goolsby, L. Shanley, and A. Lovell, “On cybersecurity, crowdsourcing, and social cyber-attack,” tech. rep., OFFICE OF NAVAL RESEARCH ARLINGTON VA, 2013.
5. X. Zhang, J. Tong, N. Vishwamitra, E. Whittaker, J. P. Mazer, R. Kowalski, H. Hu, F. Luo, J. Macbeth, and E. Dillon, “Cyberbullying detection with a pronunciation based convolutional neural network,” in *2016 15th IEEE international conference on machine learning and applications (ICMLA)*, pp. 740–745, IEEE, 2016.
6. V. R. Kebande, N. M. Karie, and S. Omeleze, “A mobile forensic readiness model aimed at minimizing cyber bullying,” *International Journal of Computer Applications*, vol. 140, no. 1, pp. 28–33, 2016.
7. S. Omeleze and H. S. Venter, “Digital forensic application requirements specification process,” *Australian Journal of Forensic Sciences*, vol. 51, no. 4, pp. 371–394, 2019.

8. S. Omeleze and H. S. H. Venter, "Digital forensic application requirements specification process," *Australian Journal of Forensic Sciences*, vol. 51, no. 4, pp. 371–394, 2019.
9. A. Valjarevic and H. S. Venter, "Towards a digital forensic readiness framework for public key infrastructure systems," in *2011 Information Security for South Africa*, pp. 1–10, IEEE, 2011.
10. V. R. Kebande and H. S. Venter, "Novel digital forensic readiness technique in the cloud environment," *Australian Journal of Forensic Sciences*, vol. 50, no. 5, pp. 552–591, 2018.
11. V. R. Kebande and H. S. Venter, "On digital forensic readiness in the cloud using a distributed agent-based solution: issues and challenges," *Australian Journal of Forensic Sciences*, vol. 50, no. 2, pp. 209–238, 2018.
12. J. Sammons, *The basics of digital forensics: the primer for getting started in digital forensics.* Elsevier, 2012.
13. M. M. Houck and J. A. Siegel, *Fundamentals of forensic science.* Academic Press, 2009.
14. S. O. Baror and H. Venter, "A taxonomy for cybercrime attack in the public cloud," in *International Conference on Cyber Warfare and Security*, pp. 505–X, Academic Conferences International Limited, 2019.
15. K. Kent, S. Chevalier, T. Grance, and H. Dang, "Guide to integrating forensic techniques into incident response," *NIST Special Publication*, vol. 10, no. 14, pp. 800–86, 2006.
16. NIST, "Guide to integrating forensic techniques into incident response (nist special publication 800-86)," Tech. Rep. August, ..., 0000.
17. M. G. Sibiya *et al.*, *Digital forensic model for a cloud environment.* PhD thesis, University of Pretoria, 2015.
18. V. Kebande and H. Venter, "A functional architecture for cloud forensic readiness large-scale potential digital evidence analysis," in *European Conference on Cyber Warfare and Security*, p. 373, Academic Conferences International Limited, 2015.
19. V. R. Kebande and H. S. Venter, "Adding event reconstruction to a cloud forensic readiness model," in *2015 Information Security for South Africa (ISSA)*, pp. 1–9, IEEE, 2015.
20. V. R. Kebande and H. S. Venter, "A cloud forensic readiness model using a botnet as a service," in *The international conference on digital security and forensics (DigitalSec2014)*, pp. 23–32, Ostrava: The Society of Digital Information and Wireless Communication, 2014.
21. S. O. Baror, H. S. Venter, and R. Adeyemi, "A natural human language framework for digital forensic readiness in the public cloud," *Australian Journal of Forensic Sciences*, pp. 1–26, 2020.
22. ISO/IEC, "27043: 2015 international standard, information technology — security techniques — incident investigation principles and processes," *ISO.org*, vol. 1, no. 1, pp. 1–30, 2015.
23. J. Howe *et al.*, "The rise of crowdsourcing," *Wired magazine*, vol. 14, no. 6, pp. 1–4, 2006.
24. J. Howe, *Crowdsourcing: How the power of the crowd is driving the future of business.* Random House, 2008.
25. L. Hammon and H. Hippner, "Crowdsourcing," *Business & Information systems engineering*, vol. 4, no. 3, pp. 163–166, 2012.
26. K. Ramasubramanian and A. Singh, "Feature engineering," in *machine learning using R*, pp. 181–217, Springer, 2017.

27. C. Zhang, L. Cao, and A. Romagnoli, "On the feature engineering of building energy data mining," *Sustainable cities and society*, vol. 39, pp. 508–518, 2018.
28. N. Sun, J. Zhang, S. Gao, L. Y. Zhang, S. Camtepe, and Y. Xiang, "Data analytics of crowdsourced resources for cybersecurity intelligence," in *International Conference on Network and System Security*, pp. 3–21, Springer, 2020.
29. H. Faris, J. Alqatawna, A.-Z. Ala'M, I. Aljarah, *et al.*, "Improving email spam detection using content based feature engineering approach," in *2017 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*, pp. 1–6, IEEE, 2017.
30. M. Yang and Q. Wen, "Detecting android malware with intensive feature engineering," in *2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, pp. 157–161, IEEE, 2016.
31. D. McClelland and F. Marturana, "A digital forensics triage methodology based on feature manipulation techniques," in *2014 IEEE International Conference on Communications Workshops (ICC)*, pp. 676–681, IEEE, 2014.
32. J. Zhu, B. Shen, and F. Hu, "A learning to rank framework for developer recommendation in software crowdsourcing," in *2015 Asia-Pacific Software Engineering Conference (APSEC)*, pp. 285–292, IEEE, 2015.
33. R. Hanzlick, *Principles of Evidence.* Juta and Company Ltd, 2006.
34. D. Santos, O. Sergeyeva, A. Boudhir, S. Baror, H. Venter, V. Kebande, *et al.*, "A framework for concurrent contact-tracing and digital evidence analysis in heterogeneous environments," in *Innovations in Smart Cities Applications Volume 4the Proceedings of the 5th International Conference on Smart City Applications*, vol. 183, pp. 1183–1196, 2020.
35. E. Casey, *Digital evidence and computer crime: Forensic science, computers, and the internet.* Academic press, 2011.
36. A. Valjarević, H. Venter, and R. Petrović, "Iso/iec 27043: 2015—role and application," in *2016 24th Telecommunications Forum (TELFOR)*, pp. 1–4, IEEE, 2016.
37. S. Omeleze and H. S. Venter, "Testing the harmonised digital forensic investigation process model-using an android mobile phone," in *2013 Information Security for South Africa*, pp. 1–8, IEEE, 2013.
38. V. R. Kebande, N. M. Karie, R. D. Wario, and H. Venter, "Forensic profiling of cyber-security adversaries based on incident similarity measures interaction index," in *2018 International Conference on Intelligent and Innovative Computing Applications (ICONIC)*, pp. 1–6, IEEE, 2018.
39. V. R. Kebande, I. Kigwana, H. Venter, N. M. Karie, and R. D. Wario, "Cvss metric-based analysis, classification and assessment of computer network threats and vulnerabilities," in *2018 International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD)*, pp. 1–10, IEEE, 2018.