







VITEC: A Violence Detection Framework

Hany Ferdinando^{1,2(✉)} , Tuija Huuki¹ , Liang Ye^{1,3} ,
Tian Han^{1,4}, Zhu Zhang^{1,4}, Guobing Sun^{1,5}, Tapio Seppänen¹ ,
and Esko Alasaarela¹

¹ University of Oulu, Oulu, Finland
hany.ferdinando@oulu.fi

² Petra Christian University, Surabaya, Indonesia

³ Harbin Institute of Technology, Harbin, China

⁴ Harbin University of Science and Technology, Harbin, China

⁵ Heilongjiang University, Harbin, China

Abstract. Hundreds of millions of youths suffer from various violence each year. The negative impacts motivate much research and numerous studies on violence. However, those attempts went their own way, making the achieved results, especially from engineering, not so useful. Based on the Sensor and Social Web (SEWEB) concept, Violence Detection (VITEC) was proposed as a possible framework to facilitate multi-disciplinary researchers in their fight against violence. At its core, it consists of a primary agent, which is violence detection using physiological signals and activity recognition, and a secondary agent, which is violence detection using surveillance video. The second layer of the proposed framework contains a cloud computing service with a Personal Safety Network (PSN) database. The cloud computing service manages all data, notifications, and some more thorough processing. The upper layer is for both observed young persons and members of the PSN. The proposed framework offers business opportunities. The existing school violence/bullying intervention programs can take advantage of VITEC by providing almost instant notifications of violent events, enabling the victims to get immediate help and intensifying coordination among different sectors to fight against violence. In the long run, VITEC may provide an answer related to the vision of having a world free from violence in 2030, as addressed by the UN Special Representative of Secretary-General on violence against children.

Keywords: Violent detection · Bullying detection · VITEC · Surveillance video · Physiological signals · Activity recognition

1 Introduction

Our societies change dramatically and drastically due to advancement in technologies. Information and news spread easily and quickly to enable fast learning and engagement. There are many online tutorials we can find not only in text form but also in audio/video/multimedia format. We can know what happened on the other side of the world almost instantly. Intercontinental video calls are now very affordable, with better quality. However, the negative side is unavoidable. Abuses of power emerge in new and

unexpected ways, such as harassment in social media and cyberbullying. Consequently, feelings of insecurity may arise not only among adults but also among young people.

Indeed, hundreds of millions of youths suffer from different kinds of violence in educational environments [1–3] and on the streets [4] every year all over the globe. Addressing problems of violence in the lives of children has never been more relevant than in 2018. Highly valued research reviews inform us that the consequences of violence are diverse, immediate, and long-lasting. Young persons who experience violence are likely to have stomach pains, headaches, and difficulties eating and sleeping; they may become afraid of attending school, which may in turn interfere with their ability to concentrate in class or participate in school activities. Furthermore, a violent experience can cause depression, loneliness, low self-esteem, suicidal thoughts, interpersonal difficulties, antisocial and criminal behavior, and attempted suicide [5, 6]. Research has also shown how violence slows social progress by generating huge economic costs, hindering sustainable development, and eroding human capital. A study found that the global costs of violence against children could be as high as US \$7 trillion per year [7]. All forms of violence infringe on the fundamental right to a safe life, create unsafe learning environments, and reduce the quality of lives for all children and youths.

Concerning these impacts, United nation children’s fund (UNICEF) launched #ENDviolence in 2013 to bring them to an end. The UN Special Representative of the Secretary-General on violence against children (SRSG-VAC), Marta Santos Pais, emphasized the vision of building a world free from fear and violence as the result of adopting the 2030 Agenda for Sustainable Development in September 2015 [8]. Quoted from Article 19 [9], *“the child must be protected from all forms of physical or mental violence, injury or abuse, neglect or negligent treatment, maltreatment, or exploitation, including sexual abuse, while in the care of parent(s), legal guardian(s) or any other person who has the care of the child”*.

However, the progress on protecting children from violence around the world has been slow, uneven, and fragmented [10]. Furthermore, most of the promising intervention models rely on education systems, schools, and teachers, which often have limited capacities and resources to work efficiently enough to address violence. Coordination between the education sector and other sectors such as health, social service, and child protection, might not work efficiently in practice, making the action too slow. The reporting system must be anonymous and confidential to protect both victims and witnesses.

Recently, it has been discussed whether or not technological solutions might assist in tackling violence in young persons’ lives. Technology has evolved to a level that would have been impossible to achieve a couple of decades ago. Recently, semiconductor technology, including sensor development has enabled systems to become smaller and more powerful than before. A smartphone, its apps and the internet have become part of contemporary life, which promote mobility for the users.

A smartphone can be considered a small computer. It is equipped with a processor, memory, and operating system. Moreover, a smartphone is also armed with several sensors, the most common one being the 3D acceleration sensor. It delivers information about the position of the smartphone, for example detecting if a smartphone is in a

landscape or portrait position, so that the display can be adjusted. It also enables the possibility for developers to write apps, providing users with various applications that can be downloaded.

In terms of protecting young persons against violence, apps such as *STOP!t*, *Speak UP!*, *BullyButton*, and *BullyTag* have been developed to help victims and witnesses report violent events anonymously. They provide a safe way to send audio, video, and text about an acute violent situation to a corresponding database server. Investigation of the database report may lead to new insights about violence. Violence, however, is an iceberg phenomenon, where only limited cases are reported, due to, for example, feeling of ashamed or fear of revenge from the perpetrators.

However, those apps, for example *Stop!t* got about three out of five stars based on the users' rating [11]. Collecting from various apps ratings, a few people reported that they were helpful to report the violent event with no fear and got assistance from the school several days later [11] and one person put his hope in anti-bullying apps [12]. There was also a comment that someone might misuse the apps to make a joke or false claim [13]. In general, all these apps require a manual activation to generate reports. In a threatening situation, most likely the victim has no time, courage, or possibility to tap a button. Also, the bystanders may face similar barriers to reporting, which makes these apps less useful. The apps for reporting violence must not only be anonymous, but also be able to autonomously detect the violent event and automatically perform the required action.

In addition to the apps-based reporting systems, there have been several attempts to detect violent situations using various modalities such as video, audio, and physiological signals. Although they have offered some promising results, no significant impact to reduce violence rate has been delivered, because each modality has worked on its own with no specific framework on which the information and achievements from various resources would be gathered and shared easily. In an attempt to connect different modalities and a workable reporting system, in this paper we propose VITEC, a violence detection framework based on the Sensor and Social Web (SEWEB) concept [14], as a contribution from various disciplines to address violence among young people. To the best of our knowledge, there is not yet available such an interdisciplinary designed, research-based framework to address physical violence targeted at young people.

SEWEB has been offered to protect children and youth from various kinds of dangerous situations, including health issues and violence [14]. It can be extended to not only school-related violence [3] but also street violence [4] and rape [15]. It is about connecting wearable and mobile sensors to social media in human well-being applications. It involves technologies such as smartphone apps, cloud computing services, social media, wearable safety devices to process the activity, location, voice, health-related and well-being information of the subjects. The system identifies and categorizes phenomena such as stress level, emotional and mood states, physical activity level in various modalities, and location. The results are reported to some pre-defined contacts, for example parents, school staffs, and authorities.

2 Research in Violence Detection

In general, two approaches or sets of system exist in violence detection research. Firstly, the systems work as a witness to violent events. Included in this approach are violence detection using surveillance video signals [16–18] and physiological signals measured from the subjects who were watching the violent scenes, e.g. EDA [19], EEG [20]. Secondly, the systems use signals measured from the victims directly, for example, their human body movement [21] and ECG signal [22].

Nam, Alghoniemy, and Tewfik pioneered a study on violent scene characterization based on flame and blood detection [16]. It opened a research pathway on violence detection using video. In general, studies using video to detect violence can be separated into two groups based on the source of video signals, either from movies or surveillance cameras. The main differences reside in the availability of sound and colour, because videos from surveillance cameras have no sound and are in a grey scale format. Video from surveillance camera usually has low resolution. The VITEC framework takes more advantages from the studies using surveillance videos. Table 1 provides the most promising attempts to detect violence using surveillance video. Various methods have been proposed with encouraging results and contributions.

Table 1. Attempts in violence detection using surveillance-like video

Authors	Results/Contributions	Note
Souza et al. [25]	Accuracies were up to 99%, depending on the methods	It used the collected grey scale videos from various sources without audio
Nievas et al. [26]	Accuracies were up to 92%, depending on the methods	It ignored color and audio to detect either fight or non-fight scene
Bilinski et al. [27]	Accuracies were up to 99%, depending on the methods and data sets. The proposed method offered a fast-computational algorithm	Videos were from various data sets: Violent Flow, Hockey Fight, and Movie
Zhang et al. [28]	Accuracies up to 87%, depending on the data sets. The proposed method offered a moderately fast computational algorithm	Surveillance videos were from various data sets: BEHAVE, the CAVIAR, and Crowded Violence
Ribeiro et al. [17]	Accuracies were around 90% using various databases. The system can be used in various contexts	Videos were from both simulation and real surveillance

The relationship between electrodermal activities (EDA), which is represented as either skin conductance or resistance, and motion picture violence has been studied as a pilot several decades ago [23]. A correlation was found between EDA and reaction to violent scenes. Lorber conducted a meta-analysis of 95 studies to investigate the relationship of EDA and heart rate (HR) – the number of heart beats in a minute – with aggression, psychopathy, and conduct problems [24]. It was suggested that both HR and EDA were reliably though modestly associated to each other in many cases.

Furthermore, violent scene annotation using EDA signal was proposed based on the hypothesis that high-level information in EDA may reveal the affective state of the audiences, especially their reactions to violence [19]. This study found that EDA offers objective, reliable, and robust measurements of user reaction to help in violent scenes annotation. However, since the signal is sensitive to any affective stimulation, e.g. pleasure, it was recommended to be used with other modalities.

Violent events could also be detected using brain signals (EEG) measured from subjects watching violent scenes [20]. It was a breakthrough to address the problem of violence, as compared to the usual questionnaires, interviews, and psychological tests. It achieved an accuracy up to 98.7% and strengthened the idea of using physiological signals to detect violent events.

Another study found that the highest average HR and EDA changes occurred in subjects who played a violent video game, as compared to watching others playing a violent video game or watching a violent movie [29]. Although EEG was not included in this study, perhaps this type of signal also provides the same response. This work empirically revealed that two approaches to attack violence were available, one from the witness' and the other from the victim's point of view. Research on violence detection using surveillance video belongs to the first approach, as well as those using physiological signals measured from the subjects watching violent scenes.

Attempting to detect violence from the victim's point of view, Ye et al. used 3D acceleration and 3D gyroscope signals measured from the subject involved in violence simulations to detect violence based on their body movements [21, 30]. The study relied on the main idea that during a violent situation the victim's body movement are different than during anon-violent situation. The proposed system successfully distinguished some normal daily life body movements from the ones related to violence, achieving an average accuracy of 92%.

Using single channel ECG (electrical activity of the heart) signals measured in pupils involved in violence simulation, violent events could also be identified [22]. This work also emphasized on the violence detection from the victim's point of view. It required a wearable sensor able to measure ECG signals. Accuracies to classify violent and non-violent events using 6-s ECG signal were up to 87%.

3 VITEC Framework

Based on the SEWEB concept, an online multi-modal **VI**olence de**TE**ction (VITEC) framework is proposed using several components working together to provide reliable performance: (1) violence detection system based on the witnesses' and victims' point of view, (2) geographical location acquired from the GPS of a smartphone, (3) indoor location acquired from an RFID-based system, (4) reliable communication channel, and (5) processing unit: both local and cloud computing services.

Table 2 compares the original SEWEB concept to VITEC, providing general overview of the proposed framework in relation to the SEWEB concept. Embedded sensors used in the proposed framework are a 3D accelerometer and 3D gyroscope [21] to measure body movement or activity. This choice is a bit problematic, because not all smartphones have a gyroscope sensor. So, research on body movement recognition

using an accelerometer only is recommended. Ambient information is not relevant, but the physiological signals of the subject measured using a wireless wearable sensor are needed. The SEWEB used no CCTV, but VITEC takes advantage of research on violence detection using surveillance video signals. Both SEWEB and VITEC need location data from either the GPS or RFID-based system.

Table 2. Comparing the SEWEB's and VITEC's component side-by-side

	SEWEB	VITEC
Technologies used to identify and measure relevant signals		
Embedded sensors of smartphones	✓	✓
Microphone of smartphones	✓	
Cameras of smartphones	✓	
Processing capacity of smartphones	✓	✓
Wearable sensors	✓	✓
Ambient sensors	✓	
CCTV		✓
Data measured and collected by sensors		
Activity data	✓	✓
Location data	✓	✓
Voice data	✓	
Health-related data	✓	✓*
Well-being data	✓	
Video data		✓
Technologies used to collect, process, and share/display data		
Smartphone apps	✓	✓
Cloud computing service	✓	✓
Social media solutions	✓	✓
Wearable safety devices	✓	✓
Phenomena to identify, measure, and categorize		
Change in stress level	✓	
Emotional state changes	✓	
Fast and slow changes in mood	✓	
Changes in physical activity level and modality	✓	✓
Detailed location information	✓	✓

* Physiological signals such as ECG, EDA, EEG, temperature, etc.

Although we cannot avoid verbal bullying, voice data is not relevant to VITEC because recording voice outside an isolated area is problematic. When someone is shouting the voices are recorded through all surrounding microphones, and the traceability of the voice data to a specific subject is questionable. VITEC does not use the smartphone camera, because users have to direct the camera manually. It is unrealistic to assume that the victim would have the possibility to direct the camera towards the perpetrator(s) or himself/herself.

VITEC does not identify stress level and emotional/mood changes. Psychologists argue about what kind of emotions/moods emerge during violent events, but to the best of our knowledge no empirical study has yet been found, which map those changes to violent events. Perhaps in the future these modalities can be added.

The four main components that make up the VITEC framework are a primary and secondary detection system, a Personal Safety Network, and a cloud computing service. The primary detection system is a collection of numerous primary agents, which work on the young people and detect violence based on the victim's point of view. The secondary detection system contains several of CCTV cameras connected to local video processing that detect a violent event based on the witness' point of view. They serve as secondary agents. Figure 1 shows the connectivity diagram of the VITEC framework.

As mentioned above, pupil welfare systems and other child protection services have limited capacities to work efficiently to address violence. The primary and secondary systems together with the cloud service are able to track and report violence automatically, delivering an additional resource for child safeguarding and combating violence. The VITEC framework also provides an automatic reporting feature. Both primary and secondary agents initiate the report by sending a message about suspicious events to the cloud service. When the cloud service confirms that a violent act, that event is recorded in the database together with the ID of all agents, the geographical location, date, and time. Thus, the number of unreported violent events can be reduced. Furthermore, the database in the cloud service can be used to study phenomena related to violence, which can benefit schools, other child welfare services, policy makers, and parents. It also enables better coordination between education and other child welfare sectors, such as health and social services and child protection. The primary and secondary systems depend upon various research activities, and so does the cloud service. As a result, the VITEC framework arranges such spaces for researchers from various fields to work together to fight against violence. It also encourages more collaborative work to deliver more valuable results.

3.1 Primary and Secondary Agents

Each primary agent consists of a wireless wearable multi-modal sensor that measures signals from the body of a young person and regularly sends appropriate signals to their smartphone for further processing. Each young person in VITEC is equipped with a primary agent. The smartphone employs a pre-detection algorithm, which must be simple but powerful enough to run within the limited smartphone environment and resources, to identify whether or not a suspicious event has occurred. Once a suspicious event is identified, the primary agent sends signals carrying that suspicious event along with the primary agent's ID and geographical location acquired from GPS to the cloud for further processing. The contribution of detecting violence using human body movements [21, 30] and physiological signal [22] fit for this role.

The secondary system contains several CCTV cameras connected to local video processing. CCTV cameras are security cameras that monitor the area of interest, for example a school yard, parking area, or street corner. Each CCTV sends recorded surveillance video with its ID to a local computer server for further processing.

This local server sends the detection result to the cloud. Results presented in Table 1 are promising for this role, because installing CCTV cameras at every geographical spot is impossible. According to the SEWEB concept, the secondary system is an additional system.

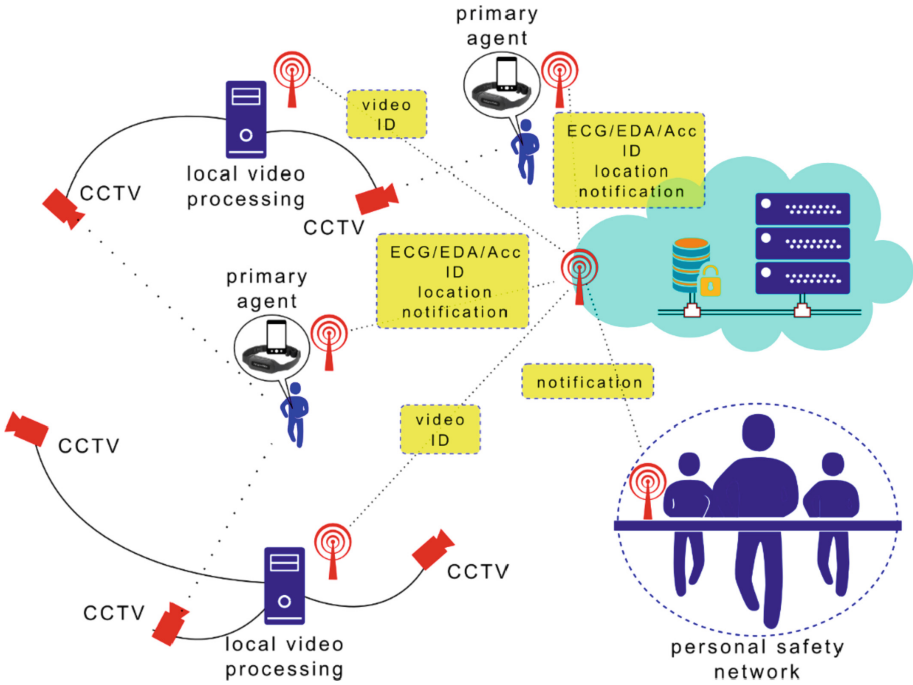


Fig. 1. VITEC, an online multimodal framework, consists of several primary agents (a smartphone and wearable sensors) and a number of secondary agents (CCTV connected to local video processing), and specific members of the Personal Safety Network connected to a cloud computing service.

Consequently, in the cloud service there are three possible cases based on the availability of the primary and secondary agents at a certain location. Upon receiving data from the primary agents, the cloud service must clarify if a violent event has occurred. It starts identifying the violent event using the signals from the primary agents. If a violent event is confirmed, the cloud service seeks for some notifications from the secondary agents in the area based on the location sent by the primary agents. The cloud services then decide which notifications should be fused to the result from the primary agents. This is the case when both agents are available. When there is no secondary agent available in the area of interest, the end result depends on the primary agents only. Another case is when there is no notification from the primary agents, but the secondary agents report a suspicious event at a certain location. The reason for this might be that the victim is one of the primary agents, but the battery power is low or

that victim is not in the system yet. In this case the cloud service sends notification of the event to local authorities, for example the nearest police station, so the victim gets help immediately.

The primary agent solves problem about manual activation in the existing anti bullying apps. With automatic activation, the victims do not have to remember to press certain button when they experience violence. Furthermore, the system also sends notification to the authority on the victims' behalf. Some victims decide to report the violence events later, but most of them cancel or forget it.

3.2 Personal Safety Network

The Personal Safety Network is defined as the network connections a user wants to be in touch with in case of emergency or a socially challenging situation [14]. In other words, to whom should the notification be sent, depending on the context, e.g. location, type of challenging situation, time, type of help, etc. Those determine the members of the Personal Safety Network, so it is dynamically changing. For example, when a young person experiences violence on the street or on the way home from school, parents and the nearest authority can be notified. So, VITEC includes parents in all cases by default and dynamically changes other members of the Personal Safety Network based on the context.

3.3 Notifications

Notifications are sensor-initiated and user-initiated [14]. Within the VITEC framework, the sensor-initiated notification is a notification generated by the cloud services, because it is the end result after processing signals from sensors. The cloud sends a notification about a suspicious event to the members of the Personal Safety Network retrieved from the database as an alarm. The corresponding young person also receives a confirmation notification, indicating that the cloud has acknowledge the event. When a young person thinks that a confirmation notification should be received but he/she receives none, a user-initiated notification can be sent to the cloud manually, similar to the existing smartphone apps related to violence. Perhaps some situations cannot be measured by sensors, so that no notification is generated by the cloud services. This kind of situation is defined as a challenging situation. VITEC also sends notification about important messages, e.g. low battery power. Notification can be in a text message and/or as a status displayed in social media.

3.4 Context and Sensor Data

According to SEWEB, sensor data must be interpreted based on the context [14]. For example, in violence detection a primary agent interprets a violent event on the basis of pushing or tackling, also in cases of equally rough and tumble play. To decrease these kinds of false alarms, a multi-modal approach – including for example EDA and ECG – is needed to aid in violence detection. Both physiological signals are sensitive enough to affect stimulation, such that it may indicate that the subject is not in danger but having fun. Ferdinando et al. (2017) showed that it is possible, although not always,

to separate some fun activities denoted as non-violence, e.g. playing, from simulated violence, e.g. pushing and hitting. It will require deeper investigation to get a more accurate algorithm to extract relevant information from the raw signals.

Figure 2 displays the position of each component in the VITEC framework in a layer diagram. This is a general layer diagram, as detailed layers on communication channels between layers are not provided. However, these follow a general data communication layer, consisting of certain standard internet protocols between clients and servers. The core of VITEC, as shown in Fig. 2, contains the primary and secondary agents, which detect violent events from the victims' and witnesses' points of view, respectively. Each agent only sends a message to the cloud when a suspicious event occurs.

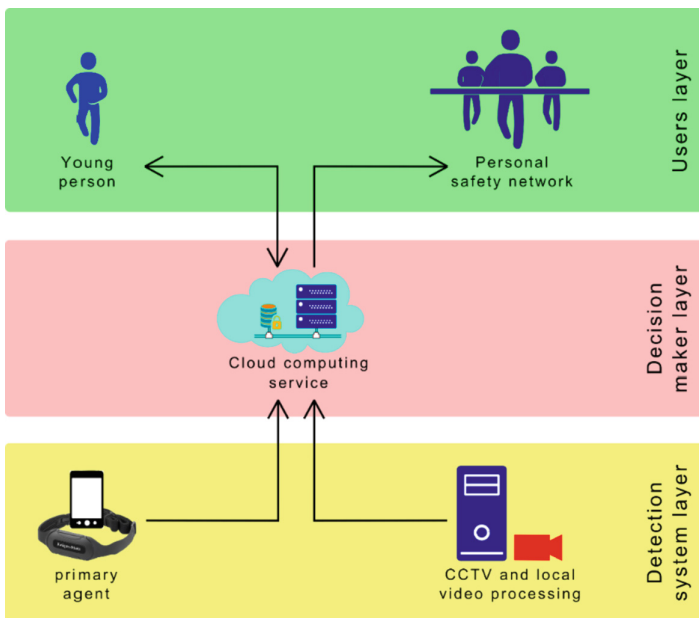


Fig. 2. Layer diagram of the VITEC framework presenting the position of each component and how each component interacts with the others.

The cloud computing service analyses the signals sent by primary agents more thoroughly and then combines with information from secondary agents to decide whether a violent event is occurred or not. If the violent event happens, the cloud service send notification to contact retrieved from database based on the context of the young person. A double arrow between the young person and the cloud represents a confirmation notification sent by the cloud to the young person and a user-initiated notification from the young person to the cloud, described above as a challenging situation.

4 The VITEC Service Platform and Business Concept

The VITEC service platform and business concept utilises a cloud computing service. The VITEC service platform can use either *Software as a Service* or *Platform as a Service*. If the system only provides services for end-users, then the *Software as a Service* should be used, where the vendor takes responsibility to maintain and develop the applications, and the users only provide data to the cloud. If there is the possibility to extend the services to other developers, then the right choice is the *Platform as a Service*. Instead of delivering service to end-users only, the *Platform as a Service* also provides services to developers who can create the applications for end-users. It allows for more collaboration among developers and expands the capability of the system to handle more complex problems.

Adopting the SEWEB business concept, VITEC offers a safety solution for young persons, location monitoring, and notifications for specified members of the Personal Safety Network to enable security knowledge with a potential market of around 10 million families [14]. The advantages offered by this business concept are the combined use of social media and sensor with shareable data for school or other related authorities. Some key operations within this business concept are apps and service developments, R&D, and marketing via various anti-bullying/anti-violence associations and school.

5 Social Contributions of VITEC

VITEC is a novel, innovative measure, designed through the multidisciplinary collaboration of academic scholars coming from the fields of engineering, psychology, and educational sciences for immediate recognition of acute physical violence among young persons. The affordances of VITEC are multiple across various age groups and across a range of spaces, places, geographical locations, and cultures. VITEC offers many ways forward for parents and other educators and public authorities, such as, school personnel, pupil welfare staff, social workers, and police by enabling efficient, immediate intervention when an acute violent situation is at hand. For the targets of violence, VITEC forms a unique personal alarm system that automatically contacts specified members of the social welfare network in case of a violent situation. It potentially increases the safety of not only young persons but also other potential targets of violent assault in locations and time of intensified risk for violence such as underpasses, empty or remote streets, on the way home from school, and at night-time. As for rape, there is an insistent myth (with hardly any supporting evidence) that girls and women lie about rape [31]. An automatic digital reporting system can potentially reduce the belief that rape is falsely reported, and so contribute to wider social change to address how rape is understood and responded to, and thus support in reducing sexual violence. Overall raised awareness of available digital applications for immediate intervention with increased risk of getting caught can make rape and other physical assaults less attractive to perpetrators.

6 Potential Future Research

The VITEC framework might open potential research in various fields. Machine learning has a wide range of research opportunities. For example, employing primary and secondary agents, the optimum way to fuse the results from both agents is still an open question. It ranges from the simplest method, e.g. a linear fusion with or without weight, to more complex methods, e.g. a Bayesian Network [32]. Another important example would be algorithms to choose the appropriate secondary agents based on the location. It is possible to get several secondary agents, which can contribute to detecting violent event, but perhaps not all agents are useful.

Researchers in machine vision have developed many useful methods to extract features from surveillance videos, and more advanced algorithms are coming to improve the present ones. Various scenarios may appear to verify the proposed methods. One possible scenario is when the recorded event is partially visible in a corner of the video frame only and no one knows whether a violent event has happened or not. The system must be able to decide whether such scenes are discarded or not. If motorized CCTVs are installed, the system may rotate the CCTVs' position when some suspicious events are detected at a street corners, for example, to make those scenes appear more clearly.

Perhaps what the primary agent requires most is a robust feature extraction method. ECG, for example, requires methods able to obtain powerful features from only several seconds signal, which is very challenging. Methods to harvest features from acceleration and gyroscope signals are on the horizon, and advancement in mathematics and programming languages supports them. Included in this part is a pre-detection algorithm running in the smartphone.

Presently, the primary agent relies on ECG signals and acceleration/gyroscope signals. To the best of our knowledge, there is no wearable sensor able to measure these three signals synchronously. Adding new modalities may require new sensors, which are not available yet. For example, using EEG signals require a new model of sensor, because wearing the available EEG sensor outside a laboratory or medical facility is not an option. This should motivate research on sensor development, e.g. a multi-modal wearable sensor with wireless connectivity, a wearable sensor with low energy consumption, etc.

VITEC depends on a reliable network, so data transmission should be made as efficient as possible. The primary agent sends data such as ID, location, raw signals, and notification of a possible violent event to the cloud service for further processing. The data transmission is only enabled when the primary agent receives a suspicious result, assuming that the pre-detection algorithm is mature enough. It allows saving the power and resources of the smartphone. The secondary agents also send data, e.g. detection result and ID to the cloud service. It means research on lossless data compression is in demand.

The proposed platform also drives research on apps development within the ubiquitous computing community. It includes but is not limited to (1) handling such a large number of nodes within the network communication to provide reliable operation; (2) dealing with resource constraint of the smartphone; (3) managing data transmission security and cryptographic protocol.

Research on violence detection using surveillance video also demands studies related to the optimum numbers and positions of cameras. It should minimize the number of cameras but offer large observation areas. This influences the number of dedicated local computers processing the data. Furthermore, recognition results from several angles or views that overlap must be fused carefully to provide reliable results. In the field of optimization, some strategies must be found to get the most appropriate number of CCTVs used in a certain area. This starts by defining some required parameters, e.g. possibility of occlusion, coverage area, percentage of overlapping areas, type of CCTV camera, etc. and continues to algorithm development and performance evaluation.

7 Discussions and Conclusions

Based on the SEWEB concept, the Violence Detection Framework – VITEC – was proposed as a response to problems related to violence among young people. It utilises sensors installed on the body of a young person and CCTV cameras to collect data to detect acute events of physical violence. It consists of a primary detection system (based on the victim’s point of view) and secondary detection system (based on the witness’ point of view). It also provides a framework to enable many researchers working on violence-related projects to work hand in hand to fight against violence.

VITEC places research on violence detection at its core to tackle the problem of school violence, providing researchers the means to transform their promising results into an applicable system. The primary agents detect violence using physiological and activity signals measured from the young persons. Although research on violence detection from the victim’s point of view is not mature yet, advancements in mathematics, signal processing methods, computer technologies, and sensors development indicate that it is not beyond reach. Similarly, methods used in surveillance video-based violence detection in the secondary agent also take advantage of those advancements. The ultimate goal of these parts is to find a general model for both agents.

VITEC can be combined with existing school violence/bullying intervention programs to provide another way to report violent events. It also provides almost instant violent event notifications with location information. Furthermore, it may provide the answer of a world free from violence, as addressed by the UN Special Representative of Secretary-General on violence against children. VITEC is a kind of unique personal alarm system that automatically contacts specified persons.

Acknowledgment. This research and publication were supported by the Finnish Cultural Foundation, North-Ostrobothnia Regional Fund 2017.

References

1. United Nations: Protecting children from bullying - Report of Secretary-General. <https://violenceagainstchildren.un.org/content/protecting-children-bullying-report-secretary-general>. Accessed 2017
2. Sunnari, V., Kangasvuo, J., Heikkinen, M., Kuorikosi, N.: Gendered and Sexualized Violence in Educational Environments. Oulu University Press, Oulu (2002)

3. Barter, C., Berridge, D.: *Children Behaving Badly? Peer Violence Between Children and Young People*, pp. 1–262. Wiley-Blackwell, Chichester (2011)
4. Young, T., Hallsworth, S.: Young people, gangs, and street-based violence. In: Barter, C., Berridge, D. (eds.) *Children Behaving Badly? Peer Violence Between Children and Young People*. Wiley-Blackwell, Chichester, Chapter 5, pp. 59–69 (2011)
5. United Nations Educational, Scientific and Cultural Organization (UNESCO): *School Violence and Bullying: Global Status Report*, pp. 1–56. Unesco, Paris (2017)
6. United Nations Children’s Fund: “UNICEF #ENDviolence,”. <https://www.unicef.org/end-violence>. Accessed 2018
7. Pais, M.S.: The economic costs of violence against children, 13 July 2015. https://violenceagainstchildren.un.org/economic_costs_of_vac_viewpoint. Accessed 2018
8. Pais, M.S.: SRSRG Santos Pais viewpoint: the 2030 agenda-implementation, follow up and review. http://srsrg.violenceagainstchildren.org/viewpoint/2016-08-02_1478. Accessed 2018
9. Office of the United Nations High Commissioner for Human Rights: *Convention on the Rights of the Child*. <http://www.ohchr.org/EN/ProfessionalInterest/Pages/CRC.aspx>. Accessed 2018
10. SRSRG on Violence against Children: *Toward a World Free from Violence - Global Survey on Violence against Children*, pp. 1–160. Unicef, New York (2013)
11. Google: STOPit - Apps on Google Play. <https://play.google.com/store/apps/details?id=com.stopitcyberbully.mobile&hl=en>. Accessed 2018
12. Google: Speak Up! (R) For Someone - Apps on Google Play. <https://play.google.com/store/apps/details?id=com.rule14.socialvision>. Accessed 2018
13. Apple Inc.: STOPit on the App Store. <https://itunes.apple.com/us/app/stopit-app/id719179764?mt=8>. Accessed 2018
14. Kinnunen, M., et al.: Wearable and mobile sensors connected to social media in human well-being applications. *Telematics Inform.* **33**, 92–101 (2016)
15. Phipps, A., Ringrose, J., Renold, E., Jackson, C.: Rape culture, lad culture and every-day sexism: researching, conceptualizing and politicizing new mediations of gender and sexual violence. *J. Gend. Stud.* **27**, 1–8 (2018)
16. Eyben, F., Weninger, F., Lehment, N., Schuller, B., Rigoll, G.: Affective video retrieval: violence detection in hollywood movies by large-scale segmental feature extraction. *PLoS ONE* **8**, 1–9 (2013)
17. Ribeiro, P.C., Audigier, R., Pham, Q.C.: RIMOC, a feature to discriminate unstructured motions: application to violence detection for video-surveillance. *Comput. Vis. Image Underst.* **144**, 121–143 (2016)
18. Zhang, T., Jia, W., Yang, B., Yang, J., He, X., Zheng, Z.: MoWLD: a robust motion image descriptor for violence detection. *Multimed. Tools Appl.* **76**, 1419–1438 (2017)
19. Fleureau, J., Penet, C., Guillotel, P., Demarty, C.H.: Electrodermal activity applied to violent scenes impact measurement and user profiling. In: *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, pp. 3310–3315 (2012)
20. Baltatzis, V., Bintsi, K., Apostolidis, G.K., Hadjileontiadis, L.J.: Bullying incidences identification within an immersive environment using HD EEG-based analysis: a Swarm Decomposition and Deep Learning approach. *Sci Rep.* **7**, 1–8 (2017)
21. Ye, L., Ferdinando, H., Seppänen, T., Alasaarela, E.: Physical violence detection for preventing school bullying. *Adv. Artif. Intell.* **2014**, 1–9 (2014)
22. Ferdinando, H., et al.: Violence detection from ECG signals: a preliminary study. *J. Pattern Recogn. Res.* **12**, 7–18 (2017)
23. Kleinkopf, H.F.: *A pilot study of galvanic skin response to motion picture violence*. Thesis, Texas Tech University (1975)

24. Lorber, M.F.: Psychophysiology of aggression, psychopathy, and conduct problems: a meta-analysis. *Psychol. Bull.* **130**, 531–552 (2004)
25. de Souza, F.D.M., Chavez, G.C., Valle Jr., E.A., de Araujo, A.A.: Violence detection in video using spatio-temporal features. In: 23rd SIBGRAPI Conference on Graphics, Patterns and Images, pp. 224–230 (2010)
26. Bermejo Nievas, E., Deniz Suarez, O., Bueno García, G., Sukthankar, R.: Violence detection in video using computer vision techniques. In: Real, P., Diaz-Pernil, D., Molina-Abril, H., Berciano, A., Kropatsch, W. (eds.) CAIP 2011. LNCS, vol. 6855, pp. 332–339. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-23678-5_39
27. Bilinski, P., Bremond, F.: Human violence recognition and detection in surveillance videos. In: 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 30–36 (2016)
28. Zhang, T., Yang, Z., Jia, W., Yang, B., Yang, J., He, X.: A new method for violence detection in surveillance scenes. *Multimedia Tools Appl.* **75**, 7327–7349 (2016)
29. Jung, Y., Skoric, M., Kwon, J.H., Detenber, B.: Watching vs. playing: effects of violent media on presence, physiological arousal and aggressive cognitions. In: Proceeding of the International Society for Presence Research Annual Conference, pp. 1–7 (2011)
30. Ye, L., Ferdinando, H., Seppänen, T., Huuki, T., Alasaarela, E.: An instance-based physical violence detection algorithm for school bullying prevention. In: Proceeding of 2015 International Wireless Communications and Mobile Computing Conference (IWCMC), pp. 1384–1388 (2015)
31. McMillan, L.: Police officers’ perceptions of false allegations of rape. *J. Gend. Stud.* **27**, 9–21 (2018)
32. Holmes, D.E., Jain, L.C.: Introduction to Bayesian networks. In: Jain, L.C. (ed.) *Innovations in Bayesian Networks. Studies in Computational Intelligence*, vol. 156, pp. 1–5. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-85066-3_1