A Generalized Mechanism beyond NLP for Real-Time Detection of Cyber Abuse through Facial Expression Analytics

Atanu Shome
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
atanu.cse.ku@gmail.com

Md. Mizanur Rahman
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
engr.mizanbd@ymail.com

Sriram Chellappan
University of South Florida
Tampa, USA
sriramc@usf.edu

A. B. M. Alim Al Islam
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
alim_razi@cse.buet.ac.bd

ABSTRACT

Abuse in cyber space is a problem requiring immediate attention. Unfortunately, despite advances in Natural Language Processing techniques, there are clear limitations in detecting instances of cyber abuse today. Challenges arising due to different languages that teens communicate with today, and usage of codes along with code mixing and code switching make the design of a comprehensive approach very hard. Existing NLP based approaches for detecting cyber abuse thus suffer from a high degree of false negatives and positives. In this paper, we investigate a new approach to detect instances of cyber abuse. Our approach is motivated by the premise that abusers tend to have unique facial expressions while engaging in an actual abuse episode, and if we are successful, such an approach will be language-agnostic. Here, using only four carefully identified facial features without any language processing, and realistic experiments with 15 users, our system proposed in this paper achieves 98% accuracy for same-user evaluation and up to 74% accuracy for cross-user evaluation in detecting instances of cyber abuse.

CCS CONCEPTS

• **Human-centered computing** → *User studies.*

KEYWORDS

Human Computer Interaction (HCI); Cyber Bullying; Machine Learning; Facial Expression

ACM Reference Format:
Atanu Shome, Md. Mizanur Rahman, Sriram Chellappan, and A. B. M. Alim Al Islam. 2019. A Generalized Mechanism beyond NLP for Real-Time Detection of Cyber Abuse through Facial Expression Analytics. In *16th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (MobiQuitous), November 12–14, 2019, Houston, TX, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3360774.3360793

MobiQuitous, November 12–14, 2019, Houston, TX, USA
© 2019 Association for Computing Machinery.
ACM ISBN 978-1-4503-7283-1/19/11…$15.00
https://doi.org/10.1145/3360774.3360793

1 INTRODUCTION

Cyber abuse is amongst the biggest challenges to young people as a result of time spent online. It creates an environment that can yield significantly negative short-term and long-term impacts for victims [1, 2]. Furthermore, cyber abuse is not a problem only for young people, or for people in a particular society, but is increasingly affecting citizens across all ages across the globe today. Detecting (which will potentially lead to combating) instances of cyber abuse is thus a topic of urgent interest today.

1.1 State-of-the-Art in Detecting Cyber Abuse

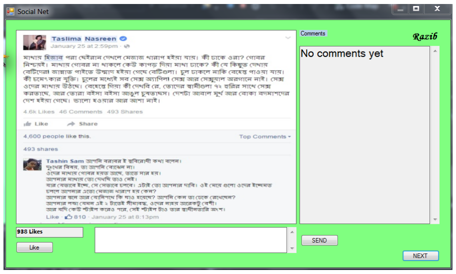
a) Crowd-Supported Detection: Popular social media platforms such as Facebook, Twitter, YouTube, etc. use crowd-sourced reporting features to identify actions indicative of cyber bullying and/or abuse. However, this approach is problematic and perhaps exhibits limited appeal, since real instances of cyber abuse often go un-reported [3], and even more dangerous are situations when users flag someone as abusive when they actually are innocent. The key challenge here stems from most users lacking a robust understanding of what constitutes cyber abuse due to complexities of interpreting language, acronyms, codes, contexts, emotions, and more. Furthermore, setting metrics and thresholds for social media platforms to actually flag someone as a bully is not at all easy today considering diversification in social media usage [4].

b) Natural Language Processing (NLP) Based Techniques: There is a significant body of work in the academia to detect cyber abuse from an NLP perspective [1, 3, 5–8]. Using a combination of words, sentiment analysis, and sequences of words (by carefully using a database of suspicious words), the idea is to design learning algorithms to flag text as abusive or otherwise. Unfortunately, complexities in text related to code mixing [9], code switching [10], and significant variations across numerous (around 6500) languages around the world [11]) result in a large number of errors and false alarms. Furthermore, since abuse is not necessarily textual, but also can include content arising from images and videos, NLP techniques have limited applicability in real-world settings today.

c) Approaches Taken by Parental Control Apps in the Market Today: There is now a lot of demand for apps that parents can install on devices of their children that flag inappropriate content. Upon research, we find that existing apps primarily rely on only identifying suspicious key words (e.g., “die", “hate", “drugs", “abuse", etc.), or inappropriate content in images (e.g., a bottle of alcohol, or a knife, or a gun) to flag them. Very minimal NLP or Image Processing is accomplished in this regard. Naturally, the false positive rate is too high, and after sometime, interests of parents wane out due to many false alarms.

1.2 Our Novel Approach to Detect Cyber Abuse

In this paper, we envisage a novel approach to detect instances of cyber abuse via monitoring unique facial features of a perpetrator during an actual episode of abuse. It is accepted that human beings exhibit six different kinds of dominant emotions via their faces that are recognizable by other humans [12]. These are anger, happiness, surprise, disgust, sadness and fear. We are intrigued by the possibility that perpetrators of cyber abuse, when they actually engage in the process of abuse will have unique expressions on their faces that can be an indicator of abuse. In a very recent study [13] done by Pozzoli et al., some interesting aspects of emotion and cyber abuse are presented. The most interesting and relevant findings from [13] are that a) those that can detect "fear" in others tend to be better at finding potential victims for abuse; while b) those that can detect "anger" are better at avoiding victimization by bullies. In parallel, we are aware of recent trends in detecting fear and other emotions via processing facial expressions [14–16]. Thus, with our intuition, and findings from recent studies, the premise of designing a system that is able to recognize subtle facial expressions of bullies during an on-going episode becomes strong. Such a system if successful is language agnostic.



Taslima Nasreen
January 25 at 2:59pm

মাথায় হিজাব পরা ছেইরান দেখলে মেজাজ খারাপ হইয়া যায়। কী ঢাকে ওরা? গোবর নিশ্চয়ই। মাথায় গোবর না থাকলে কেউ কাপড় দিয়া মাথা ঢাকে? কী যে কিম্ভুত দেখায় বেটিদের! জান্নাত পাইতে উন্মাদ হইয়া গেছে বেটিগুলা। চুল ঢাকলে নাকি বেহেস্ত পাওয়া যায়। কী চমেৎকার যুক্তি। চুলের মধ্যেই সব সেক্স অ্যাপিল! সেক্স আর সেক্সুয়াল অরগানে নাই। সেক্স ওদের মাথায় উঠছে। বেহেস্তে গিয়া কী দেখবি রে, তোদের স্বামীগুলা ৭২ হুরির সাথে সেক্স করতাছে, আর তোরা বইসা বইসা আঙুল চুষতাছস। দেশটা আবাল মূর্খ আর বোকা বদমাশদের দেশ হইয়া গেছে। ভালো হওয়ার আর আশা নাই।

4.6k Likes 46 Comments 493 Shares

Like Share

4,600 people like this. Top Comments

493 shares

Tashin Sam আপনি বরাবর ই স্ববিরোধী কথা বলেন।
দুঃখের বিষয়, তা আপনি বোঝেন না।
ওদের মাথায় গোবর হয়ত আছে, তাতে সার হয়।
আপনার মাথায় তো দেখছি তাও নেই।
যার যেভাবে ইচ্ছে, সে সেভাবে চলবে। এটাই তো আপনার দাবি। ওই মেয়ে গুলো ওদের ইচ্ছেমত চললে আপনার এতো মেজাজ খারাপ হয় কেন?
আপনার স্তনে আর যোনিপথে কি ঘাও হয়েছে? আপনি কেন তা ঢেকে রেখেছেন?
আপনার লজ্জা যেমন এই ২ টাতেই সীমাবদ্ধ, ওদের নাহয় আরেকটু বেশী।
আর যদি কেউ স্টাইল করেও পরে, সেই স্টাইল টাও তার স্বাধীনতারি অংশ।
Like · 810 · January 25 at 8:13pm

No comments yet

938 Likes

Like SEND NEXT

Figure 1: Screen-shot of our customized social networking platform "Social Net"

1.3 Our Contributions

In this paper, we aim to design a system that is capable of detecting intentions of abuse by a perpetrator via analyzing subtle changes in facial expressions recorded from a camera in the device the person is typing from. Our specific contributions are as follows.

a) Designing a Customized Social Network Platform: We develop and deploy a customized social networking platform named "Social Net" in this paper (Figure 1). It is a desktop-based platform for interacting with social media posts, with an added feature that the application running on the platform captures seven photos of a user every second. The images are sent to a server in *real-time* for processing.

b) Real Experiments on Abuse in Our Platform: Based on extensive user interviews, mining existing social media usages, and taking local cultures (Bangladesh, in our study) into account, we then identify a series of topics that participants were asked to engage others on. These topics had a high change to trigger abuse-related conversations in a social media platform such as ours. Having a total of 15 participants active in our platform for a week, we identified 5925 and 25575 sample points indicative of abusive and non-abusive content respectively. Recall that throughout the experiment, the images of subjects were being recorded from the device. Initially, we collected images from our participants without prior notifications to avoid biased inputs. At the completion of the study, we notified them about the image capturing, and they agreed to let us use the images in our research study, as long as identities were not revealed.

c) Feature Extraction, Machine Learning, and Evaluations: Then, after carefully tagging the images as indicative of ones with abusive intent or otherwise, we identify a total of 15 facial features, and finally narrow down to four that clearly separate one class from the other. At a high level, the features we extract indicate emotions most related to *anger.* Using these four features, and a Random Forest based classification algorithm, we achieve 98% accuracy for same-user evaluation and up to 74% accuracy for cross-user evaluation in detecting cyber abuse.

d) Real-Time Detection Module: Finally, we develop an application to indicate presence of abusive intent in cyber communications. Since real-time intervention is our ultimate concern, we include a real-time alarm generator module within "Social Net" for detecting abusive intent, and generate an alarm when that happens. We conduct user evaluations of our application in real settings with five participants and achieve up to 72% accuracy (with an average of 69% accuracy). Usability study on the same set of participants shows that despite the fact our solution captures images from users, they think the solution is effective and will discourage abusers. We explain our platform, questionnaire, posts, and other details of this experiment later in this paper in Section 4.

2 RELATED WORK

In this section, we present a review or related literature. We start with the research studies done so far for detecting cyber abuse. Finally, we show context that support our proposed system.

Detecting cyber bullying and abuse has been rigorously investigated over the last decade. Researchers here mostly focus on textual data analysis [6–8, 17, 18]. For example, Nandini et. al., attempt to use Fuzzy logic to filter the input text to prepare for classification, and detect cyber abuse using genetic algorithms [8]. Reynolds et. al., worked with machine learning on textual data and show an accuracy of 78.5% in detection of cyber abuse using C4.5 Decision tree learner and instance based learner [17]. They use "Amazon's Mechanical Turk" web service to label the textual data as related to cyber bullying or not. Dinakar et al., apply common sense reasoning on textual data for the purpose of abuse detection [18]. They have constructed a common sense knowledge base "BullySpace" and utilized their proposed technique "AnalogySpace" for applying common sense reasoning to detect abuse.

Chen et al., perform another investigation through incorporating abusers' personal context with textual data such as writing style (e.g., number of capital and small letters used, ratio of short sentences, etc.) [19] to detect abuse. They propose a Lexical Syntactic

Feature (LSF) architecture to do so. However, this work also uses textual analysis along with personal feature based separation.

The problems with existing NLP approaches are that they are trained for specific languages only (primarily English). There are more than 6500 languages in this world [11]. Thus, it becomes impossible for social networks to generalize abuse/bully detection. Moreover young people today tend to use different short message forms [20], which makes text-based solutions much harder to succeed. Code mixing [9] and code switching [10], which are common practices today, stretch the extent of difficulty in using textual data analysis for real-time cyber abuse detection. Nonetheless, Sari et al., observe that textual data from both cyber abuse and aggressive jokes are often correlated [21], which add yet another complex dimension in textual analysis for cyber abuse detection.

Considering all these challenges in text based detection of cyber abuse, we are intrigued by the possibility of using real-time facial analytics to generalize abuse detection. Computing techniques in the realm of facial expression is a well studied topic [22–24]. For example, Barlett et al., achieve 85% accuracy in detection of pain by measuring facial movements and pattern recognition using computer vision [22]. Bonanno et al., in a psychological study, examine disclosure-nondisclosure of childhood sexual abuse in relation to nonverbal expressions of emotion in faces of subjects [25]. They find dissimilarities in expressions between abused and non-abused persons. Besides, Christani et al., employ Social Signal Processing (SSP) based on video surveillance for identifying non verbal cues such as face expressions, gazing, and body postures to relate them with context dependent activities [26]. Garcia et al., show that people exhibit emotional state changes while interacting online [16].

In the realm of connecting abusive behaviour with human emotions, studies show that there are strong correlations between anger and intent to abuse/ bully [27–31]. In other words, when a subject is intending to abuse or bully someone, then the subject is most likely to express the emotion of anger. Wang et. al., experimented with 464 young Chinese adults and applied social-cognitive model along with general aggression model [29]. The study concluded that cyber abuse is positively and significantly associated with anger. Bosworth et al., in another study observe that anger was a powerful predictor of abusive behavior [28]. They conclude that high levels of anger are associated significantly with the highest levels of abuse. Another study by Hussain et al., show that victims of abuse exhibit anger-in phenomenon as opposed to anger-out by perpetrators of abuse. These works establish the fact that abusive behavior can be revealed through facial expressions, and is worthy of investigation. Based on success of peer research in the space of showing the feasibility of emotion recognition through facial expression analytics [13, 32, 33], we envision detecting complex facial expressions that could indicate intent to abuse, hence paving the way for a language agnostic technique.

In terms of other related work, a compelling study done by Gheiratmand et al., show that Schizophrenia can be predicted by large-scale data analysis using functional Magnetic Resource Imaging (fMRI) [34]. They achieve 74% accuracy in generalized detection of Schizophrenia using a combination of imagery from a subject's physical biometrics and brain imagery. Researchers have also worked on finding unique signatures from keystrokes for authentication purpose [35], which is also a behavioral marker.

To summarize, our motivation in this study is to analyze facial expressions of persons that intend to abuse in cyber space. To the best of our knowledge, facial expression based cyber abuse identification in real time, is yet to be investigated in the literature and our study is the first one in this field.

3 PROPOSED ARCHITECTURE

Figure 2 shows our proposed facial-analytics based architecture to detect if a subject is engaging in an instance of perpetrating cyber abuse. Our proposed architecture is divided into two major parts. First is the surveillance, which keeps capturing photos from users and extracts facial feature dynamics. Second is matching those extracted dynamics with pre-defined patterns (using machine learning) to detect intent to abuse in real-time.

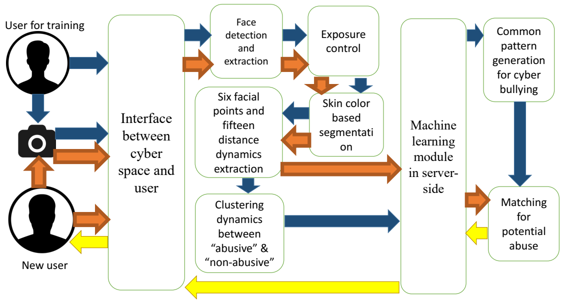

User for training
Interface between cyber space and user
Face detection and extraction
Exposure control
Skin color based segmentati on
Six facial points and fifteen distance dynamics extraction
Clustering dynamics between "abusive" & "non-abusive"
Machine learning module in server-side
Common pattern generation for cyber bullying
Matching for potential abuse
New user

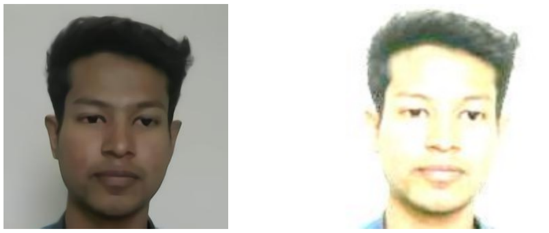
Figure 2: Proposed Architecture

3.1 Surveillance and Facial Expression Analytics Generation

At the outset, to solve our problem of detecting patterns of abusive intent from facial expressions, we perform a series of techniques - face detection, image manipulation, dark spot separation, and so on. First step towards finding facial feature dynamics requires identification of face, which we have done using OpenCV [36].

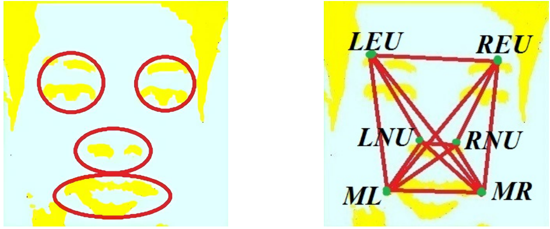
a) Face Extraction using OpenCV: We have used OpenCV [36] to identify a face and crop the identified face (Figure 3a). Haarcascade classifiers were used to identify the facial shape from the gathered photos. It is reported that 95% accuracy can be achieved for face detection using OpenCV [36]. However, in the case that multiple faces are detected inside a single photo, we ensure that only the center face is processed for our purpose. This is most reasonable for our situation. The face extracted as such was resized and saved into fixed 300 × 300 pixels (Figure 3a). This ensures consistent measurement of features irrespective of distance from camera.

b) Image Manipulation & Skin Color based Segmentation: To extract facial features indicative of abusive intent, we need to identify dark spots, and their separation using skin color based manipulation. Every face contains certain dark spots such as eyebrow,



(a) Initial face detection and cropping using OpenCV [36] **(b) Image processing: exposure control and segmentation [37]**

Figure 3: Face extraction and image processing



(a) Dark spot separation **(b) Proposed 6 facial *points* and 15 facial *distance* features**

Figure 4: Facial feature dynamics extraction

pupil, nostril, mouth etc. These are the ones that change based on changing facial expressions, which we need to capture in our study. To identify the dark spots, first we perform image manipulation to highlight those dark areas (Figure 3b). This is a two-step process. First, we perform exposure control to enable us highlight the dark spots. OpenCV permits us to control the exposure of the image. We use two parameters namely contrast (α) and brightness (β) in this regard. The following equation shows the pixel manipulation where $g(i, j)$ and $f(i, j)$ represents pixel value

$$g(i, j) = \alpha \times f(i, j) + \beta. \tag{1}$$

Next, we perform color-based segmentation [38, 39]. Here, we adapt default skin color tone as Caucasian RGB (239, 208, 207). We convert all pixels into two colors. First, pixel colors that are greater than RGB (45, 45, 45) are converted into the color RGB (224, 255, 255). Rest of the pixel colors are converted into RGB (255, 255, 0), which represents those darker spots.

c) Facial Feature Points and Distance Dynamics Extraction: We extract six facial feature points from the dark regions. These are topmost points of left eye and right eye, topmost points of left nostril and right nostril, and left most and right most points of mouth (Figure 4b). Besides, to explore patterns while being involved in cyber abuse, we utilize feature distance dynamics between each pair from these six points. Thus, we consider a total of ${}^6C_2 = 15$ feature distances at the beginning. Here, we use Euclidean distances. For each photo, we calculate these fifteen feature distances (Figure 4b). The reason for exploring only fifteen features (more specifically feature distances) are mainly because of maintaining the criteria of making a real-time application for abuse detection. More features will require more computational time, which might limit the possibility of our solution to perform detection of cyber abuse in real-time. Note that the feature points under our exploration have been frequently adopted or realized by several existing research studies for emotion recognition [12, 33, 40–44]. Also, many research studies have specifically looked into Euclidean distance between points in faces for feature selection for emotion recognition/ facial expression [32, 45–47].

3.2 Real-Time Cyber Abuse Detection

The final step in our system is detection of an instance of abuse. In training our system, after analyzing all the photos based on our features, we generate a two dimensional matrix. Here, each row represents one photo, and the fifteen columns in the row represents 15 distance features. After the generation of the matrix, we mark each row by one of the two possible labels - "abusive" and "non-abusive". This marking procedure is described in the following section. After marking all the rows, we apply machine learning algorithms to train our system for learning a pattern. We find that Random Forests [48] provide the best accuracy of 98% for our purpose while having our dataset divided into 80% for training and 20% for testing. We elaborate the complete process for finding individual and generalized patterns in the next section.

4 SYSTEM DEVELOPMENT AND IMPLEMENTATION ASPECTS

A critical challenge in our experiments was collecting data at the moment of abuse. In our initial experimentation we thought about developing a module that could capture photos while users use social networking sites such as Facebook, Twitter, etc. However, this could not ensure providing us information about which photo is related to actual cyber abuse episode as there is no way to determine or interpret the activities of a user. However, in order to tag any ground truth data as "abusive" or "non-abusive", we need to know exactly which photos are captured from user's face during a particular comment, and what the user's intentions were then. Thus, we need to have a mechanism for relating every photo with corresponding comments and emotions. Therefore, we develop a customized application called "Social Net" with photo capturing capability (Figure 1). While a user engages with "Social Net", photos are taken using a webcam and corresponding features are saved pertinent to each of those photos. Thus, a sample point represents a single photo, associated features, timing of that photo, related post number, interaction type of the user at that moment. Note that our system captures photos without any action or notification to the user. We do so as people could get conscious with the knowledge of capturing photos, and therefore, might provide non-spontaneous fabricated expressions [49].

All captured photos are processed for analysis in our scheme. Note though that in real life implementations of our approach, only feature data and associated algorithms will be saved in the server end for analyzing. The process of image generation, pre-processing and feature extraction from users will be performed in the user-end. Therefore, in such a real-life implementation, privacy will not be impacted, since raw images will not be shared with server.

Now, in order to experiment with our "Social Net" system for our problem, we design it like a honey trap, and present a certain degree of provocative contents to induce a certain degree of abusive actions. These contents were collected from Facebook. We have applied brainstorming sessions, while emphasizing on culture, norms, and thought process of potential survey participants to gather these contents, while respecting sensitivities (i.e., avoiding content that can be deemed too provocative). We have gone through popular groups in Facebook and collected controversial posts. We also provided normal posts such as jokes, memes etc., inside “Social Net" to invoke non-abusive actions. One thing to note is that since, this paper focus on the perspective of the person engaging in abuse, only his/her facial expressions are captured. The party at the other end of this communication in our system were adult volunteers and they were notified of the study details.

We find that participants did use abusive language and intended to abuse in our system during our experimentation. Throughout each session, we capture 1000-4000 photos per person while they were using “Social Net" at the interval of 150 ms per image, which is equivalent to seven shots per second, so that we do not miss any subtle change in facial expressions. We collected a total of 31500 photos from 15 adults (Table 1). Besides, users are offered the opportunity to interact with the system using “like" and “comment" options. These comments are subsequently used to label the feature data between two clusters of actions - “abusive" and “non-abusive". We invited ten separate volunteer to mark the comments as “abusive" or “non-abusive" after careful reading the posts. One very interesting thing we observed was that the language used in our system was a mixture of multiple languages (primarily, Bengali and English), and there were significant use of acronyms, local slang and codes. Like we mentioned earlier, NLP based approaches are too complex to design in such cases, hence further validating the motivation for our work to glean abuse intent from facial expressions. The demographics of our 15 participants is presented in Fig 7.

Table 1: Count of abusive and non-abusive sample points

Cluster	Total no. of photos	% of photos
Abusive	5925	19%
Non-abusive	25575	81%

Note that we maintain a separate file that keeps track of the photos captured during each and every comment. We mark the feature data extracted from those photos that were captured during an abusive comment as “abusive". Together, they make “abusive" cluster data. All other feature data are in “non-abusive" cluster data. We feed feature data dynamics into “WEKA", a machine learning tool, used to analyze using various machine learning algorithms [50]. We use several machine learning classifier algorithms (namely Naive-Bayes, Multilayer Perception, Bagging, Random committee, Random Forest, Random Tree, etc.) to find the best possible algorithm for our purpose. In our first machine learning experimentation phase we train the machine using a set of individual’s own feature data and test on another set of that user’s feature data. To do so, we divide the data set for each of the 15 individuals into 80% for training and 20% for testing inside “Weka" at its choice. Besides, we divide 80% consecutive rows separately and used for training and last 20% consecutive rows for testing. We find that Random Forest algorithm provides the best results for our classification task with an accuracy up to 100% having an average of 98%.

4.1 Individual Abusive Pattern Recognition in Facial Expression

In all experiments presented from now on, we use the Random Forest algorithm, as we find it as the most effective one in our experiments. Here, we divide each individual’s data between training and testing datasets (90% training and 10% testing dataset), which provides an average accuracy of 99%. This accuracy demonstrates that our approach is highly capable of detecting cyber abuse while trained with data from the same person.

4.2 Common Abusive Pattern Recognition in Facial Expressions

This approach applied on the same individual may not work for finding common patterns, as distances between facial aspects can be different for each person. Thus, we need a normalization of distance dynamics to make them more generalized. We use the following equation to normalize the distance dynamics.

$$Normalized\_distance = \frac{distance\_between\_any\_two\_points}{(\frac{((LEU-MR)+(REU-ML))}{2})} \quad (2)$$

We perform the normalization task through dividing all distance values by the average of distances - 1) from top most left eye point to right most mouth point, and 2) from top most right eye point to left most mouth point. We compute fifteen normalized distance dynamics for all the datasets and train up the machine with multiple persons’ data to eventually test on another person’s data. We randomly take 6, 8, 10, and 12 peoples’ data for training, and test over on other persons to see if abusive behavior can be identified in this cross-person manner. This normalization and cross platform testing provides only 10% to 20% accuracy. Therefore, to achieve a better accuracy, we narrow down to few core facial dynamics that mostly contribute to the changes in facial expressions pattern at the time of abuse.

Extracting Core Facial Expression Dynamics: We plot all the fifteen distance dynamics for all individuals and attempt to identify the dynamics that have the most influence on the change in expressions. Here, we perform a centroid based analysis. We calculate centroid for the “abusive" and “non-abusive" clusters respective to all fifteen dynamics. We observe that there are four core facial expression dynamics, which have exhibit most changes for these two clusters than the remaining other eleven facial expression dynamics. The four core facial expression dynamics are: a). LEU-ML, b). LNU-ML, c). REU-MR, and d). RNU-MR (Figure 5). Here, LEU means Left Eye Uppermost Point; LNU means Left Nostril Upmost Point; REU means Right Eye Uppermost Point; RNU means Right Nostril Uppermost Point; ML/ MR means Mouth-Left Most Corner/ Mouth Right-Most Corner; and the symbol "-" means distance between. These are highlighted in Fig 4. Next, we investigate generation of common patterns using these four dynamics with normalized values. Note that dealing with more feature point distances demands more computational power, and therefore, our goal was to explore with a lower number of feature points or distances to keep resource overhead as small as possible. Accordingly, we train up our machine learning algorithm with multiple persons’ data covering the four normalized dynamics and tested on others based on the same four dynamics. This gives us an accuracy of up to 58% and also holds a result for individual pattern finding with an accuracy of 94%. Thus, we come to two realizations up to this point. First, all persons' facial expressions vary in different ways. Here, only few expression dynamics contribute to the change in expressions while performing cyber abuse. Second, the analytics that increase the accuracy for common pattern finding, degrade the accuracy for individual pattern finding.

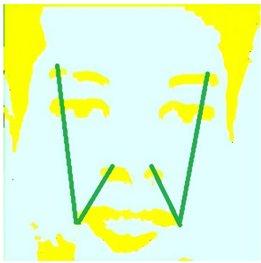


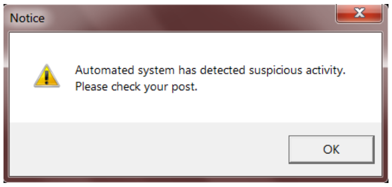
Figure 5: Our proposed four distance dynamics

Noise Reduction from Clusters: The results obtained so far exhibit potential for substantial further improvement. It happens due to the fact that while marking clusters we mark all the photos of a corresponding post to a single cluster as per its context (i.e., “abusive" or “non-abusive"). However, it is a reality that expressions reflected for abusive activity or cyber abuse may not occur for the entire interaction time for that post, and it may occur only for a small time frame within that episode. Accordingly, we need to target only those photos that actually exhibits the distinctive facial feature corresponding to the abusive activity. All other photos or expressions are the normal expressions. To enable better ground truth, we employ a filtration procedure wherein the features we identify earlier are compared as they are generated for an episode of abuse and otherwise. Only those features (captured during an abusive episode) that are distinct from the ones captured during a normal episode are used to train to detect abusive patterns. With this procedure, we train with multiple persons' data with four normalized distance dynamics with Random Forest classifier algorithm after incorporating the filtering task, and then test on others for detecting abusive activities Now, the accuracy increases up to from a minimum of 61% to a maximum of having an average accuracy of 67%. These results exhibit substantially better performance than the earlier ones.

5 REAL-TIME DETECTION MODULE

Upon performing the above off-line analysis we integrate our training set inside “Social Net" to build a real-time cyber abuse detection module. To perform the machine learning in real-time, we use machine learning module from OpenCV [51]. We implement Random Forest algorithm for the purpose of training using the method CvRTrees::train using four normalized facial dynamics training data, which are divided into “abusive" and “non-abusive" clusters. As per our proposed method, only OpenCV will run at the user-end to extract facial features (a few bytes in size), which will then be sent to social networking servers to enable Random Forest algorithm for the cyber abuse identification task. As the social networking servers are generally equipped with high-end resources, running the algorithm should not be too heavy for it. Besides, real-time image processing applications similar to the underlying operation of OpenCV already exist in smartphone applications such as Snapchat [52].

In our real-time module, we first capture photos using the user device and then send the captured photos to real-time detection module operating in the sever. In the server, upon extracting and normalizing the facial distance features, we perform the prediction task to identify an abusive activity. Upon such identification, we generate an alarm and deliver that to the GUI of “Social Net" (figure 6). To avoid users' discomforts for getting too many frequent alarm messages, we present only one alarm message into GUI in case of identification of an abusive activity. Alongside, we also store information about such abusive activity happening in the server for social networking site administrators to monitor the activities of the perpetrator. We present an usability study of such information later in this paper.



Automated system has detected suspicious activity.
Please check your post.

Figure 6: Alarm or warning for a suspicious abusing activity generated in real-time

5.1 Privacy Issues and Human Factors

In our study, we have collected images from our participants without prior notifications. However, afterwards, we notified them about the image capturing and they happily agreed to let the images use in our research studies.

In case of real-time deployment, privacy can be pointed as an important issue. However, note that different pervasive devices nowa-days such as CCTV cameras also capture people's movement (without their consent) [53]. Nonetheless, if our proposed approach would be adopted by any live system, an explicit notification on image capturing could be provided to the users, as such permissions might be required to enable their cameras. Social networking services also adopt such policies to upgrade their services with prior notifications to the users [54]. Collecting human behavioral dynamics in this manner is nothing new as already elaborated in Section 2. For example, existing studies experimented with human physical analytics to develop a relationship between physical and behavioral attributes [16, 27, 35]. In our case, we develop such a relationship between facial features and abusive activities. Here, to ensure privacy, we propose to extract data in the user-end and discard the captured photos immediately after getting the facial features. Besides, we admit that a clever user can bypass our system with conscious avoidance or via fabricated expressions. We plan to explore these issues in future through experiments with more users to see whether we can detect abusive activities done by users even if they know that their facial expressions are recorded.

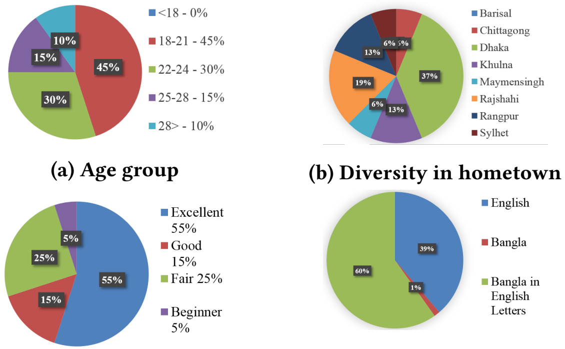
6 PERFORMANCE EVALUATION AND RESULTS

We evaluate performance of our proposed approach through real experimentation. Here, first, we present demography of the users to show an overview on variation in them that eventually indicates inefficiency of textual analysis based cyber abuse detection.

6.1 Demography on Users and Interacted Texts

Figure 7 shows the user demography of our participants. Here, we experiment with fifteen participants and let them use our customized application “Social Net". Later, we included five more adults to use our real-time implementation. We first present a brief overview to participants on how to use “Social Net" without telling the motive of our research as mentioned earlier. We set up an experimental environment inside a laboratory and we invite the participants to use the "Social Net" application individually.

It is worth mentioning that there remains a diversity in hometowns shown in Figure 7b. This diversity results in varieties over the usages of language and choices of words. Here, we find that the participants mostly tend to use native language Bengali and react to contents written in Bengali. However, they are mostly reluctant to use Bengali letters and mostly used English letters even to write comments in Bengali. In our experiment, we observed that 173 Bengali words are written in English letters and 112 words are written in English. Moreover, only 4 words are written in Bengali letters by a single user. We have also observed a fair amount of spelling mistakes done by the participants. Code mixing [9] is done by 12 participants. All these events show how inefficient it will be if only textual analytics are used for cyber abuse detection. These variations make it near-to-impossible to devise a generalized solution using textual analysis. Nonetheless, we had to label the collected data manually (along with ten volunteers used in the labeling task) to divide the interacted contents into the two clusters for these reasons.



(a) Age group

(b) Diversity in hometown

(c) Computer skill

(d) Used languages while interaction

Figure 7: Demography of Users

6.2 Experimental Settings

Users interact with “Social Net" through giving likes, writing comments, etc. While doing so, our application capture photos at the interval of 150 milliseconds. We collect 31500 sample points (Table 1) with recordings of their photos, interactions, timing and comments. Each sample point includes 15 distance vectors (as mentioned in Section 3.1). Among all these sample points, 5,925 sample points got marked as “abusive" and 25,575 sample points were marked as “non-abusive".

6.3 Experimental Results

First, we explore with machine learning algorithms and find the best fit for our purpose. For running the machine learning algorithm, we divide each person’s data into 80% training and 20% testing datasets. We utilize “WEKA" [50], a machine learning tool, to apply several machine learning algorithms in this way. We get the best performance of 98% accuracy with Random Forest algorithms (Table 2).

Table 2: Accuracy (%) of machine learning (ML) algorithms with our system (minimum, maximum, average, and Standard deviation)

ML algorithms	MIN	MAX	AVG	STD
Random Forest [48]	96	100	98	2
Random Committee [55]	95	99	98	2
Random Tree [56]	93	99	96	2
Multilayer Perception [57]	70	84	77	5
Naive-Bayes [58]	61	68	65	3
Bagging [59]	92	96	95	2

Henceforth, we will use Random Forest algorithm for machine learning. Here, we perform our experiments in two ways. First, we feed individual dataset to “Weka". We obtain 98% average accuracy (a maximum of 100%) while training with individual’s own dynamics and testing on the same person (Figure 8).

To calculate these values, we combine both accurate predictions of “abusive" and “non-abusive" cases and took average of all users. Table 3 presents statistical measures related to these.

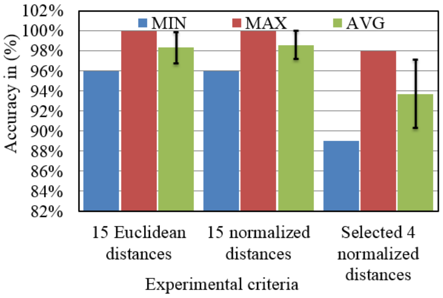


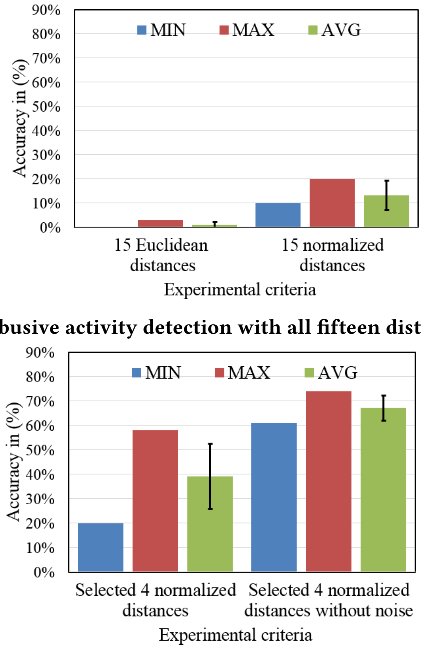
Figure 8: Results of abusive activity detection for training and testing over the same person

Next, we took a step further towards finding the common pattern, i.e., a generalized solution that would work in cross-person manner. First, we combine eight persons’ data and test on other persons separately. Results show that testing on other persons’ data exhibit an accuracy of 10% to 20% with 15 distance dynamics without selecting four key dynamics as explained earlier (Figure 9a).

Subsequently, we perform normalization as presented earlier, and

Table 3: Measures of relevance for abusive activity detection for the same person

Measure	Fifteen normalized features	Four normalized features
Precision	0.9787	0.9025
Recall	0.9857	0.9025
False-positive rate	0.0053	0.0214
False-Negative rate	0.0143	0.0975
F1-score	0.9821	0.9025



(a) Abusive activity detection with all fifteen distances

(b) Abusive activity detection with 4 normalized distances

Figure 9: Results of abusive activity detection for training (over 12 people) and testing over different sets of persons

then test with four normalized facial features. Here, the results improves up to 58% (Figure 9b). Then, we perform noise reduction as presented earlier and conduct the analysis again. Here, the results improve up to 74% (Figure 9b) and an average of 67%. Table 4 summarizes statistical measures of this case. Here, we train with two different numbers of persons (8 and 12). The results demonstrate that the detection performance improves with an increase in the number of persons under training.

Next, we integrate our best-achieved alternative (training on twelve persons' data with noise elimination) in our real-time detection module inside "Social Net". We conduct an experimentation on five persons with the real-time detection module. Here, we get a total of 57 comments out of which 31 are found abusive. In this case, we present more provoking posts, and this is why we could collect more abusive comments. Table 5 shows accuracies for the five participants in this experiment, which demonstrates an average accuracy of 69%. This average accuracy is close to that we have found in the last case, i.e., the average accuracy of 67%. Table 6 presents statistical measure corresponding to this experiment.

Table 4: Measure of relevance for abusive activity detection while training and testing over two different sets of people

Measure	Eight person's data in training	Twelve person's data in training
Precision	0.2559	0.5169
Recall	0.1506	0.3966
False-positive rate	0.1608	0.1725
False-Negative rate	0.8494	0.6034
F1-score	0.1896	0.4488

Table 5: Accuracy (%) of our real-time detection module

Participant	P1	P2	P3	P4	P5
Module accuracy	71%	63%	70%	72%	69%

We also calculate time requirement for abusive activity detection by our system. We find that our system takes only 40 seconds to train and build the model using Random Forest based classification algorithm on an average with 12 persons' data. Since, the training of the system will be done one time at the beginning and social media server computers are equipped with high configuration, it will not have any notable performance issue on the real-time prediction. Thus, remaining concern is the prediction time. With our proposed solution, it takes only 0.35 seconds on an average to predict an activity as abusive or non-abusive. Thus, it can be implemented in real-time application, which will help to prevent abuse rather than detect afterwards.

Table 6: Measures of relevance for abusive activity detection using our real-time cyber-abuse detection module

Measure	Value
Precision	0.7500
Recall	0.6774
False-positive rate	0.2692
False-Negative rate	0.3226
F1-score	0.7119

6.3.1 Usability Study. Upon completion of the system evaluation study we want to know about the usability status of our proposed real-time cyber abuse detection module. To do so, we prepare a questionnaire that we present to the five participants after they used our "Social Net" application with real-time cyber abuse detection module. Table 7 summarizes the questions that we present to the participants. These questions focus about the user interaction experience with our system and comfortability on future use of this proposed approach for cyber abuse detection in real-time. We also present a comment box to the participants to express themselves freely about our proposed system's usefulness.

Table 7: Outcomes of usability study on 5-point Likert scale

No.	Questions	P1	P2	P3	P4	P5
Q1	Do you think that our cyber abuse detection was accurate?	5	3	5	5	4
Q2	Do you think that our application is very natural to use?	5	3	5	4	4
Q3	Do you think that our application does not create any discomfort while using?	4	3	5	4	4
Q4	Will it be okay to use our application knowing that such detection module is in operation?	5	4	5	5	5
Q5	Do you think that this approach has an effect on the mind to discourage cyber abuse?	3	2	5	5	4

We let the participants answer or express their opinion based on the 5-point Likert scale. Here, '1' refers to 'strongly disagree', '2' refers to 'disagree', '3' refers to 'undecided', '4' refers to 'agree', and '5' refers to 'strongly agree'.

We find from this study that participants agree to use a social network service even if they are aware that a system like ours is present to detect abuse. In the comment box, which was included with the questionnaire, P4 said "I felt guilty for the words that I used in my comment after viewing the notification." P1 said "It forced me to think do I really need to write harsh thinks to anyone." Thus, we expect our proposed architecture to play a vital role in cyber-abuse detection in real-time and help combat it.

6.3.2 Experimental Findings. A closer look at our experimental results reveal noteworthy findings. We present some of them below:

- Our proposed architecture can achieve an accuracy up to 100% (an average of 98%) for identifying abusive behavior while being trained by the same person's data.
- In case of training with a dataset of multiple persons' data and then testing on others (not participated in training set), we achieved accuracy up to 74% (an average of 67%).
- Considering all fifteen dynamics provides the best results for individual cases (training and testing on the same person), however, considering only four specialized dynamics provides best results for collective cases (training and testing over different sets of people).
- Intervention mechanisms can be implemented based on our proposed architecture both for warning the user and storing abusive activity at the back-end for administrative action. Our implemented real-time cyber abuse detection system gives a maximum accuracy of 72% (an average accuracy of 69%) that closely matches with our previous findings for abusive activity detection with our proposed four facial feature dynamics. A usability study also confirms user acceptance of our system.

7 HCI THEORETICAL ADVANCEMENT THROUGH OUR WORK

First of all, the core field that we work on is to establish the relation between behavioral aspects and facial features. Here, we find correlations between facial expressions and complex human behavioral aspects specifically for cyber abusive behavior shown by a person. We reveal correlations using image processing and machine learning. We also identify a commonality to a certain extent over human facial characteristics while exhibiting abusive behavior on social media. We find that machine learning can achieve immense success to extract such human behavior. Finally, through a real-time module experimentation, we observe that even when participants know their activities are being monitored, they can engage in abusive activities. This characteristic also agrees with current system in social media where people do abuse, despite existence of manual reporting options. Nonetheless, participants also indicated that their abusive behavior may change upon a notification.

8 CONCLUSION

Cyber abuse is an imminent threat to both children and grown-ups on social media, which can cause imbalance in sustainability towards social and technological growth. Its impact has been very bad in recent times. Therefore, a generalized solution going beyond contemporary specialized NLP based solutions for detecting cyber abuse becomes necessary. Therefore, in this paper, we propose a new methodology using facial feature points that exhibit an ability to detect cyber abuse activities in general. Our exploitation of facial expression dynamics for the purpose of detecting cyber abuse is the first of its kind in the literature to the best of our knowledge. Our proposed methodology achieves an integrated real-time detection module with an average accuracy of 98% through training by the same user's facial dynamics and an accuracy up to 74% for generalized pattern recognition through training by other users' facial dynamics. We demonstrate the performance through real experimentation and user evaluation. In future, we plan to perform an empirical study on the users' experience to observe the effectiveness of interventions using our proposed methodology.

ACKNOWLEDGEMENT

This work was supported in part by US National Science Foundation (Grant # 1718071). Any opinions, findings and conclusions are those of the authors alone, and do not reflect views of the funding agency.

REFERENCES

[1] K. Dinakar, B. Jones, C. Havasi, H. Lieberman, and R. Picard. Common sense reasoning for detection, prevention, and mitigation of cyberbullying. *ACM Trans. Interact. Intell. Syst.*, 2(3):18:1–18:30, September 2012.

[2] Effects of cyber bullying, 2013. Last accessed on 10 January, 2019 https://www.bullying.co.uk/cyberbullying/effects-of-cyberbullying/.

[3] R. Basak, N. Ganguly, S. Sural, and S. K. Ghosh. Look before you shame: A study on shaming activities on twitter. In *Proceedings of the 25th International Conference Companion on World Wide Web*, WWW '16 Companion, pages 11–12. International World Wide Web Conferences Steering Committee, 2016.

[4] How many reports are needed to close a fake account?, 2017. Last accessed on 10 January, 2019 https://www.facebook.com/help/community/question/?id=591712690863574.

[5] R. Zhao, A. Zhou, and K. Mao. Automatic detection of cyberbullying on social networks based on bullying features. *âĂĆIn Proceedings of the 17th International Conference on Distributed Computing and Networking (ICDCN '16) ACM, New York, NY, USA*, 2016.

[6] H. Hosseinmardi, R. I. Rafiq, R. Han, Q. Lv, and S. Mishra. Prediction of cyberbullying incidents in a media-based social network. In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 186–192, Aug 2016.

[7] Q. Huang, V. K. Singh, and P. K. Atrey. Cyber bullying detection using social and textual analysis. In *Proceedings of the 3rd International Workshop on Socially-Aware Multimedia*, SAM '14, pages 3–6, New York, NY, USA, 2014. ACM.

[8] B. S. Nandhini and J. I. Sheeba. Online social network bullying detection using intelligence techniques. *Procedia Computer Science*, 45:485 – 492, 2015.

[9] A. Das and B. Gamback. Code mixing in social media text. the last language identification frontier? *Traitement Automatique des Langues*, 54:41–64, 2013.

[10] Davies E. E. Bentahila, A. The syntax of arabic-french code-switching. *Lingua*, 59:301–330, 1983.

[11] How many languages are there in the world?, 2016. Last accessed on 10 January, 2019 https://www.ethnologue.com/guides/how-many-languages.

[12] A. Fnaiech, M. Sayadi, and P. Gorce. Feature points tracking and emotion classification. In *Advanced Technologies for Signal and Image Processing (ATSIP), 2016 2nd International Conference on*, pages 172–176. IEEE, 2016.

[13] T. Pozzoli, G. Gini, and G. Altoe. Associations between facial emotion recognition and young adolescents behaviors in bullying. *PLoS one*, 12(11):e0188062, 2017.

[14] F. W Smith and S. Rossit. Identifying and detecting facial expressions of emotion in peripheral vision. *PloS one*, 13(5):e0197160, 2018.

[15] S. M. During and R. J. McMahon. Recognition of emotional facial expressions by abusive mothers and their children. *Journal of Clinical Child Psychology*, 20(2):132–139, 1991.

[16] D. Garcia, A. Kappas, D. Küster, and F. Schweitzer. The dynamics of emotions in online interaction. *CoRR*, abs/1605.03757, 2016.

[17] K. Reynolds, A. Kontostathis, and L. Edwards. Using machine learning to detect cyberbullying. In *2011 10th International Conference on Machine Learning and Applications and Workshops*, volume 2, pages 241–244, Dec 2011.

[18] K. Dinakar, R. Reichart, and H. Lieberman. Modeling the detection of textual cyberbullying. In *The Social Mobile Web*, 2011.

[19] Y. Chen, Y. Zhou, S. Zhu, and H. Xu. Detecting offensive language in social media to protect adolescent online safety. *In Privacy, Security, Risk and Trust (PASSAT), International Conference on and 2012 International Confernece on Social Computing (SocialCom)*, pages 71–80, 2012.

[20] S. Thurairaj, E. P. Hoon, S. S. Roy, and P. W. Fong. Reflections of students' language usage in social networking sites: Making or marring academic english. *Electronic Journal of e-Learning*, 13:302–316, 2015.

[21] S. V. Sari. Was it just joke? cyberbullying perpetrations and their styles of humor. *Comput. Hum. Behav.*, 54(C):555–559, January 2016.

[22] M. S. Bartlett, G. C. Littlewort, M. G. Frank, and K. Lee. Automatic decoding of facial movements reveals deceptive pain expressions. *Current Biology*, 24(7):738 – 743, 2014.

[23] W. E. Rinn. The neuropsychology of facial expression: A review of the neurological and psychological mechanisms for producing facial expressions. *Psychological Bulletin*, 95(1):52–77, 1984.

[24] R. Reisenzein, M. Studtmann, and G. Horstmann. Coherence between emotion and facial expression: Evidence from laboratory experiments. *Emotion Review*, 5(1):16–23, 2013.

[25] G. A. Bonanno, D. Keltner, J. G. Noll, F. W. Putnam, P. K. Trickett, J. LeJeune, and C. Anderson. When the face reveals what words do not: facial expressions of emotion, smiling, and the willingness to disclose childhood sexual abuse. *Journal of personality and social psychology*, page 94, 2002.

[26] M. Cristani, R. Raghavendra, A. D. Bue, and V. Murino. Human behavior analysis in video surveillance: A social signal processing perspective. *Neurocomputing*, 100:86 – 97, 2013. Special issue: Behaviours in video.

[27] D. Valiune and A. Perminas. Differences in anger, aggression, bullying among adolescents in different self-esteem groups. 6:76, 03 2017.

[28] K. Bosworth, D. L. Espelage, and T. R. Simon. Factors associated with bullying behavior in middle school students. *The journal of early adolescence*, 19(3):341–362, 1999.

[29] X. Wang, L. Yang, J. Yang, P. Wang, and L. Lei. Trait anger and cyberbullying among young adults: A moderated mediation model of moral disengagement and moral identity. *Computers in Human Behavior*, 73:519 – 526, 2017.

[30] A. Hussian and S. Sharma. Anger expression and mental health of bully perpetrators. *FWU Journal of Social Sciences*, 8(1), 2014.

[31] A new approach to school bullying: Eliminate their anger, 2016. Last accessed on 10 January, 2019 https://www.psychologytoday.com/us/blog/the-forgiving-life/201612/new-approach-school-bullying-eliminate-their-anger.

[32] A. P. Gosavi and S. R. Khot. Emotion recognition using principal component analysis with singular value decomposition. In *2014 International Conference on Electronics and Communication Systems (ICECS)*, pages 1–5, Feb 2014.

[33] X. Li, H. Xiaopeng, A. Moilanen, X. Huang, T. Pfister, G. Zhao, and M. Pietikainen. Towards reading hidden emotions: A comparative study of spontaneous micro-expression spotting and recognition methods. *IEEE Transactions on Affective Computing*, 2017.

[34] M. Gheiratmand, I. Rish, G. A. Cecchi, M. R. G. Brown, R. Greiner, P. I. Polosecki, P. Bashivan, A. J. Greenshaw, R. Ramasubbu, and S. M. Dursun. Learning stable and predictive network-based patterns of schizophrenia and its clinical symptoms. *npj Schizophrenia*, 3(1):s41537–017–0022–8, 5 2017.

[35] F. Monrose, R. Kowalski, and A. D. Rubin. Keystroke dynamics as a biometric for authentication. *âĂĆFuture Generation Computer Systems*, 16(4):351–359, 2000.

[36] Face detection using haar cascades., 2001. Last accessed on 10 January, 2019 http://docs.opencv.org/3.1.0/d7/d8b/tutorial_py_face_detection.html/.

[37] Changing the contrast and brightness of an image!, 1999. Last accessed on 10 January, 2019 http://docs.opencv.org/2.4/doc/tutorials/core/basic_linear_transform/basic_linear_transform.html.

[38] T. Tasneem, A. Shome, and S. K. A. Hossain. A gaming approach in physical therapy for facial nerve paralysis patient. In *Computer and Information Technology (ICCIT), 2013 16th International Conference on*, pages 345–349, March 2014.

[39] N. Vo, Q. Tran, T. B. Dinh, T. B. Dinh, and Q. M. Nguyen. An efficient human-computer interaction framework using skin color tracking and gesture recognition. In *2010 IEEE RIVF International Conference on Computing Communication Technologies, Research, Innovation, and Vision for the Future (RIVF)*, pages 1–6, Nov 2010.

[40] P. Michel and R. El-Kaliouby. Real time facial expression recognition in video using support vector machines. In *Proceedings of the 5th international conference on Multimodal interfaces*, pages 258–264. ACM, 2003.

[41] M. Pantic and I. Patras. Dynamics of facial expression: recognition of facial actions and their temporal segments from face profile image sequences. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 36(2):433–449, 2006.

[42] A. Singh, D. Patii, G. M. Reddy, and S. Omkar. Disguised face identification (dfi) with facial keypoints using spatial fusion convolutional network. In *2017 IEEE International Conference on Computer Vision Workshop (ICCVW)*, pages 1648–1655. IEEE, 2017.

[43] A. Lanitis, C. J. Taylor, and T. F. Cootes. Automatic interpretation and coding of face images using flexible models. *IEEE Transactions on Pattern Analysis and machine intelligence*, 19(7):743–756, 1997.

[44] D. Ghimire and J. Lee. Geometric feature-based facial expression recognition in image sequences using multi-class adaboost and support vector machines. *Sensors*, 13(6):7714–7734, 2013.

[45] A. S. Dhavalikar and R. K. Kulkarni. Facial expression recognition using euclidean distance method. 2, 03 2014.

[46] A. De, A. Saha, and M. C. Pal. A human facial expression recognition model based on eigen face approach. *Procedia Computer Science*, 45:282 – 289, 2015. International Conference on Advanced Computing Technologies and Applications (ICACTA).

[47] J. Kalita and K. Das. Recognition of facial expression using eigenvector based distributed features and euclidean distance based decision making technique. 4, 03 2013.

[48] Wiener M. Liaw, A. Classification and regression by randomforest. *R news*, pages 18–22, 2002.

[49] L. Mazerolle. Social behavior in public space: An analysis of behavioral adaptations to cctv. *Security Journal*, 15(1), 7 2002.

[50] Weka 3: Data mining software in java, 2016. Last accessed on 10 January, 2019 http://www.cs.waikato.ac.nz/ml/weka/index.html.

[51] ml. machine learning., 1999. Last accessed on 10 January, 2019 http://docs.opencv.org/2.4/modules/ml/doc/ml.html.

[52] Snapchat, 2017. Last accessed on 10 January, 2019 https://www.snapchat.com/.

[53] C. J. Cohen, K. A. Scott, M. J. Huber, S. C. Rowe, and F. Morelli. Behavior recognition architecture for surveillance applications. In *2008 37th IEEE Applied Imagery Pattern Recognition Workshop*, pages 1–8, Oct 2008.

[54] Facebook privacy policy, 2017. Last accessed on 10 January, 2019 https://www.facebook.com/about/privacy/.

[55] Seung H. S. Shamir E. Tishby N. Freund, Y. Selective sampling using the query by committee algorithm. *Machine learning*, pages 133–168, 1997.

[56] D. Aldous. The continuum random tree - i. *The Annals of Probability*, pages 1–28, 1991.

[57] Dorling S. R. Gardner, M. W. Artificial neural networks (the multilayer perceptron) a review of applications in the atmospheric sciences. *Atmospheric environment*, pages 2627–2636, 1998.

[58] R. Kohavi. Scaling up the accuracy of naive-bayes classifiers: A decision-tree hybrid. *KDD*, pages 202–207, 1996.

[59] N. C. Oza. Online bagging and boosting. *n Systems, man and cybernetics, 2005 IEEE international conference on*, pages 2340–2345, 2005.