Detection of Possible Illicit Messages UsingNatural Language Processing and ComputerVision on Twitter and Linked Websites

Vaibhav Kumar Singh
Department Of Computer Science And Engineering
Apex Institute Of Technology Chandigarh University
Mohali -140413, Punjab
20BCS3842@cuchd.in

ABSTRACT

Millions of victims of human trafficking suffer from this widespread issue, which robs them of their dignity. Social networks are currently utilised to disseminate this crime across the internet by disseminating hidden messages that advertise these illicit services. It is crucial in this situation to automatically discover messages that may be connected to this crime and could possibly act as clues because law enforcement resources are limited. In this study, we use natural language processing to find tweets that might advertise these illicit services and take advantage of children. It is feasible to find pictures of children under 14 years old since the images and URLs obtained in suspicious messages were processed and categorised by gender and age. The approach we took was as follows. First, real-time hashtag mining is done on tweets referencing minors. These tweets are preprocessed to remove background noise and typos before being categorised as suspicious or not. Additionally, Haar models are used to choose the geometrical characteristics of the face and body. We can identify a person's gender and age group by using support vector machines (SVM) and convolutional neural networks (CNN), even when the details of the face are blurry or the torso is out of proportion to the head. As a result, the SVM model with torso-only features performs better than CNN.

INDEX TERMS CNN, features detection, image classification, natural language processing, SVM.

I. INTRODUCTION

Since the user could not actually interact with the web at first, the websites were segregated and just intended for reading. The user ceased being a passive observer and started participating actively in social networks like Facebook, Twitter, and Instagram, among others, as a result of the invention and arrival of web 2.0.

Unfortunately, this has also made it easier for criminal enterprises like human trafficking to flourish. Latin American nations, for example, have the greatest rates of persons being smuggled, particularly young children and adolescents under the age of 14. The average age of consent in Latin American nations is 14, therefore if minors are employed for illegal services, it is necessary to emphasise that directly regarded as human trafficking victims. Currently, it is easy to find escort or similar service websites on Twitter, where young girls are advertised for the consumption of "customers." Typically, these girls are subjected to physical, psychological, and sexual abuse.

Many criminal organisations have been utilising social media to sell these "wrong services," concealing their illicit activity by using seemingly innocent words like "chicken soup" to refer to child pornography. Websites and social media platforms are utilised to bring this crime into the online world, where covert messaging and advertising are used to advertise illegal services and take advantage of the victims of this crime, who are primarily kids.

Although there have been prior attempts to identify unlawful messages using picture classification and tweet filtering, most of these systems combine computer vision and natural language processing. It does, however, treat text and graphics differently. This essay's authors emphasise their work analysing online advertisements for the automatic detection of ambiguous messages. For this task, 10,000 manually annotated ads are used. This work categorises text- and image-based advertising, and the analysis integrates both forms of data. With the use of a deep multi-modal model they developed dubbed the Human Trafficking Deep Network, they were able to get an F1 value of 75.3% and a recall of 70.9%.

However, the existing picture classification methods exclusively take into account facial information, ignoring the fact that the majority of photographs feature obscured faces. In, the authors estimate that age can be predicted with an accuracy of 86.64% using computer vision techniques. In, the gender of an individual is determined using the SVM and CNN classification algorithms. To the best of our knowledge, there are no works that take the upper body (upper torso) features in the photos into account when categorising age groups.

The current project comprises two stages. First, postings on Twitter that advertise illegal services offered by minors are found using natural language processing algorithms. Images are retrieved in the second phase from websites flagged as suspects in order to perform image processing and gender recognition for two age groups: those above 14 and those under or equal to 14. The traits of the body as well as the face features were utilised for this identification. It is important to note that many photos have blurry or pixelated edges.

The present introduction is the first of eight sections that make up this text. Related works are shown in Section 2. Then, Section 3 presents our system concept to identify potential people trafficking using a torso analysis. The initial stage of our effort, which involved extracting and analysing tweets that might be relevant to human trafficking, is described in Section 4. Section 5 provides information on image extraction, processing, and age and gender classification. The SVM and CNN machine learning techniques utilised are discussed in Section 6 along with how they operate. The experimental findings are then discussed in Section 7. Section 8 presents the conclusions and suggested next research.

II. LITERATURE REVIEW

The use of social networks for deception has been the subject of study studies. In, the authors examine the use of cheating methods in blogs, group projects, microblogging, news sites, social networks, contest communities, virtual social worlds, and virtual games. They also examine how these methods are used to manipulate contest, falsify information, handle images, and manipulate videos. Depending on the attacker's expertise, these deception techniques may have a high possibility of success.

The results of some analysis utilising natural language processing are helpful in the fight against this crime. For instance, a study is described that examines suspicious tweets to find illegal adverts. In order to detect linked adverts, the authors of utilise a semi-supervised learning approach to identify suspected human trafficking tendencies. They also employ non-parametric learning

techniques for using text analysis. Some computer vision works categorise photos by age categories such kids, teens, and adults while concentrating just on the person's face.

Similar to this, in since Facebook does not use security filters, the flaws of social networks, notably Facebook, are examined. Similar text manipulation occurs when scam surveys and product reviews are delivered to victims as spam in an effort to trick unwary users, particularly those who are looking for employment or educational possibilities.

In, it is discovered that hackers employ advanced tools and methods, such as encrypted communication or highly secure web servers, to evade detection and maintain an ambiguous status. Some experts advise examining the following factors to look for irregularities in the profiles of supposed followers who are online attackers: age inconsistency, alias variation, frequency of content, shared management, race, nationality, and third-party publications. There have been certain social network scam warning signs found. The quality of the information processed still has some limitations, though.

According to, there are often no accurate statistics on migrants because of the large number of undocumented and illegal immigrants, as well as the fact that many children are not registered or enrolled in schools. Additionally, there have been no claims of inefficiency or disregard on the part of the authorities from the victim's family. On the other hand, some dysfunctional families use their children as prostitutes to raise money.

Since there are numerous obstacles to overcome in this field, improved methodologies must be created to aid in the identification of human trafficking signs on social networks and associated websites.

SYSTEM PROPOSAL

Two parts make up our suggested method for identifying suspicious websites: (i) the treatment, analysis, and classification of tweets using natural language processing, and (ii)

ii) Classifying and processing photos hosted on websites that have been deemed questionable. Some search criteria, particularly with regard to young girls, were used in the first phase of the investigation to look for probable human trafficking.

The entire process is depicted in Figure 1 starting with the initial search for information about human trafficking or human slavery, followed by the downloading and processing of this information up until the extraction of features and their classification. The major goal of this stage is to compile a blacklist of questionable websites that are connected to tweets. The classification of photographs downloaded from the blacklist is the subject of the second step. The classification of images is carried out using prediction models, such as Vector Support Machine (SVM) and Convolutional Neural Networks (CNN), in two stages: training and testing.

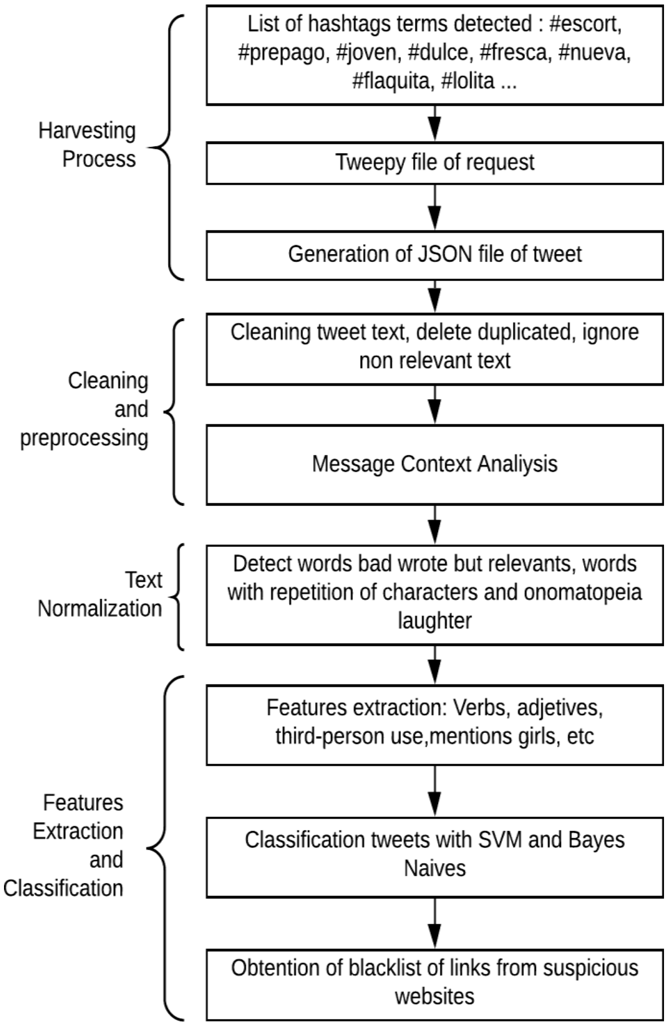
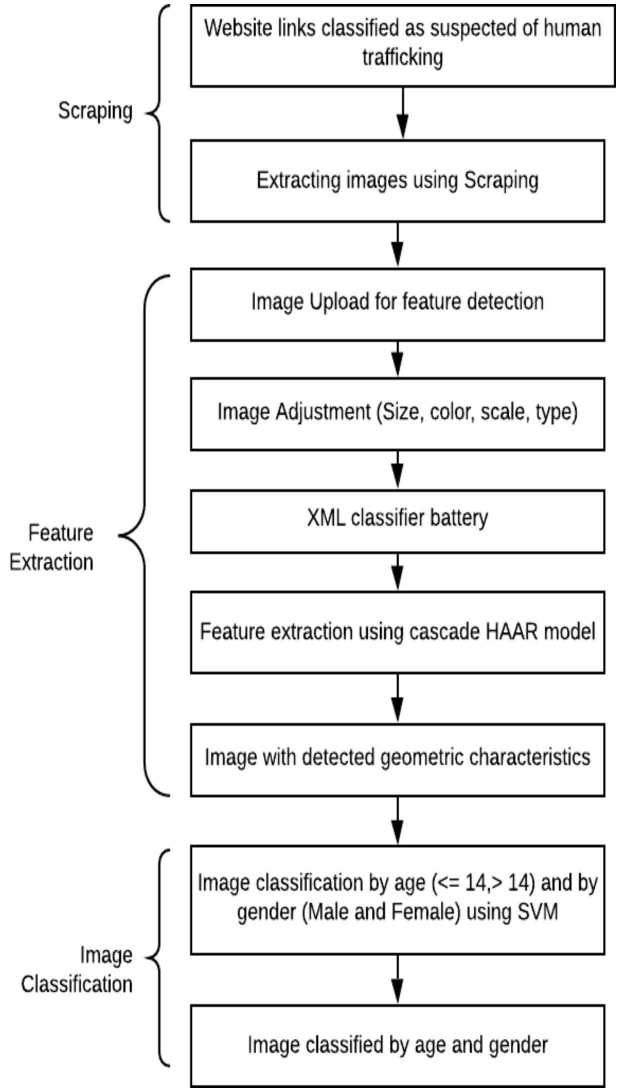


FIGURE 1. Tweet classification based on natural language processing(REFERENCE)



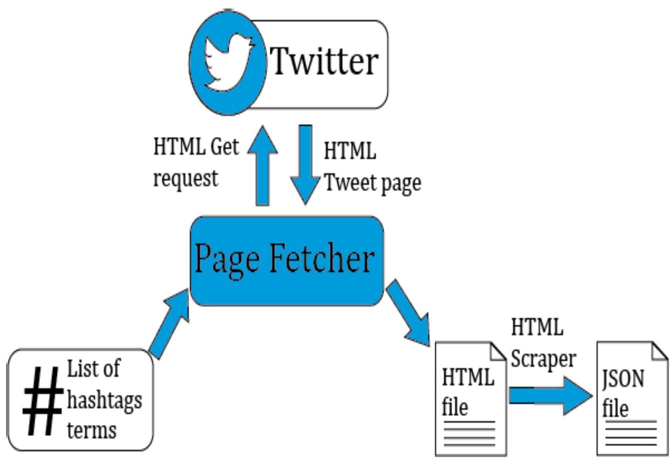


FIGURE 3. Tweets crawling process. (REFERENCE)

III. TWEET EXTRACTION AND PROCESSING

This section describes the process of tweets extraction, processing, and classification to determine if there are signs of human trafficking.

A. HARVESTING PROCESS

Each captured day's data is initially saved on a JSON file that contains details about the tweet post. The tweet's text, user information, user mentions, connected URLs, and posting time are the information that is most important. Figure 3 depicts the process of capture.

The following hashtag searches are used to gather information in Spanish: #escort, #prepago, #joven, #Dulce, #Fresca, #nueva, #lolita, and #flaquita. Hashtags were chosen as signposts for the underage population. The following criteria were used to filter tweets: mentions of individuals from foreign nations, tweets written in the third person that indicate the Twitter user is promoting the services of another individual, or tweets from the same user promoting the services of many individuals. Terms that denote victims who are underage were used, such as references to slender young people or phrases from the paedophilia vocabulary, to determine age. A preliminary investigation was conducted into tweets and Facebook posts that were accused of being involved in sex trafficking. The Spanish adjectives joven (young), dulce (sweet), fresca (fresh), nueva (new), Lolita, and flaquita (thin) are widely employed

B. CLEANING AND PREPROCESSING

The JSON format is changed into data frames corresponding to the features after the data has been loaded and cleaned. Data can then be filtered using specific criteria. To carry out an analysis more effectively, some columns were additionally established. The tweets were then automatically tokenized, or all words of a text were broken up into separate words and changed to lower case, using a Python programme.

C. FEATURES EXTRACTION AND CLASSIFICATION

Some criteria relating to deception and cybercrime are used to describe features. These standards take a victim's young age into account when looking for them. Table 3 explains each of the input qualities and why it was taken into consideration.

Syntax analysis is carried by using a Python programme to assess this.

At least one of the hashtags listed in Table 1 is always present in any downloaded messages, and these messages are kept locally in a JSON file. Using a Python application, the data is cleaned and analysed, and tweets that meet the following requirements are removed:

Hashtag	Number of occurrences
#escort	45604
#prepago	15890
#joven	3456
#dulce	1256
#fresca	1456
#nueva	5743
#flaquita	6580
#lolita	867
#penguin	23980
#caldodepollo	45990
#cp	34562
Twits with a URL link	1765

TABLE 1. Total tweets post by hashtag.

To categorise the tweets as "suspect" or "not suspicious" of being connected to sex trafficking, a semi-supervised learning technique utilising Naive Bayes and SVM algorithms was utilised. Based on average Precision (P), average Recall (R), and average F-Measure (F), each classifier's performance was assessed. SVM and Nave Bayes were selected to categorise the data using a semi-supervised technique after several algorithms were evaluated and found to perform well and process data quickly.

TABLE 3. Considered features.

Features	*Reasons to considered each characteristic*
Quantity of words	Deceptive messages have more words to make them forgettable
URL links that are related to night club websites or massage therapy sites	It serves to detect the use of Twitter to publicize these sites
Third-person use	Deceptive messages have few self-references to avoid accountability. On the other hand, other people advertise the victims' services.
Same Twitter user talking about more than one victim	Covert publicity of illicit activities
The number of hashtags considered to harvest the data	Confirmation of Twitter relevance
A user account from one country that mentions girls from another country	it is a well-known signal of sex trafficking
The number of adjectives and verbs is an indicator of a possible deceptive message	This number is higher than a standard message because deceptive messages are usually very expressive
Similar advertising from the same account promoting different women	It is a well-known sign of sex trade
Weight of women	Less than 100 pounds for very young girls
One account is promoting more than two different women	it is a well-known sign of sex trade

Due to the shortness of the texts and the fact that Twitter messages are typically made using mobile devices, there is typically a lot of noise in them. In addition, a lot of tweets contain incomplete, misspelt, or twisted words, which reduces the effectiveness of natural language processing. As a result, it is important to use twitter standardisation techniques that are written in Spanish during the preprocessing. In order to identify words that were OOV (Out of Vocabulary), the text of Tweet messages was processed with a Spanish spell-checker for lexical normalisation algorithm utilising the following criteria The performance result was the average of all tests. As it is mentioned above, Precision, Recall, and F-Measure were used in order to evaluate classifiers' per- formance against evaluation using the ground truth established from the previous annotation. To categorise the tweets as "suspect" or "not suspicious" of being connected to sex trafficking, a semi-supervised learning technique utilising Naive Bayes and SVM algorithms was utilised. Based on average Precision (P), average Recall (R), and average F-Measure (F), each classifier's performance was assessed.

which is important knowledge given that the false message is frequently very expressive. This data was recorded as a brand-new feature to be taken into account during classification. With the help of this programme, we were able to identify additional characteristics including the maximum amount of words, the identification of third-person speech, the usage of several victims by the same Twitter account, and the use of defined hashtags.

The quantity of adjectives and verbs, references of ladies from one country in messages from another origin country,

References to lightweight that might refer to extremely young girls, one account promoting multiple ladies, and identical advertising using the same language to promote other women.

From a report generated by a programme with a list of URLs, URL linkages are manually analysed to determine whether they are links to nightclub or massage therapy websites. The other properties are obtained by processing the corpus with certain Python programmes. In order to feed the classifier, the data is then imported into the input feature file. Furthermore, the hashtags listed in Table 1 are used to collect the corpus. According to the presence of more than one defined hashtag, the corpus was filtered to acquire the tweets that were the most pertinent. To demonstrate the validity of the characteristics, this corpus was automatically divided into suspicious and non-suspect tweets.

The annotated corpus, which is regarded as the ground truth, was used to compare the automatic classification performance. As ground truth, about 10% of the annotated original corpus was used. This corpus segment was randomly chosen and labelled. Analysis was done on the twitter account's attributes and the content of their messages. The tweets were also marked as "suspicious" when SVM performs better than Naive Bayes, which is to be expected. Compared to Naive Bayes, SVM has a little bit greater Precision, Recall, and F-Measure. Additionally, both classifiers have a reliable performance metric.The URLs from websites that were associated with suspected tweets were gathered and saved in an individual file called "Black-List." Once this blacklist of suspected websites was created, the following actions were taken to clean the data:

- Discard results that are duplicates.
- Remove broken or inaccessible links and websites.
- Remove any links that lead to abandoned or inactive websites.
- This improved Blacklist is used as the input data for the extraction and processing phase.

IV. IMAGEN EXTRACTION AND PROCESSING

This section describes the image extraction and processing to determine if the image is a person under 14 years old or not.

A. SCRAPING

There are websites that intentionally link tweets that mention sexual services, resulting in a tremendous influx of social network members. Infoscort, Punterking, and Shinagawaesthe are a few of the websites that are frequently mentioned in tweets containing a suspected blacklist. To find links to the images that are displayed on the websites' domains, the websites were scraped. Two steps are taken in the analysis of these suspected web sites' HTML and CSS code for this purpose:

Massive download: Because downloading suspicious photographs by hand requires that we do it one at a time, this operation is carried out automatically using scraping techniques. A website's link or URL is required as an input in order to scrape it. As a result, all of the connections to the website images were preserved and acquired as a plain text file. The photos were then downloaded automatically using the URLs in this file. It is important to be aware that if the websites demand a subscription, downloading the internal images is not an option because the subscription limits access to the subsequent pages without restriction. Scraping is thus restricted to open access pages. Furthermore, it's critical to emphasise that even while security methods have advanced, using a digital certificate [23] does not ensure that some websites' actions are lawful. This type of fraudulent enterprises exchange digital content (pictures and movies) using seemingly innocent terms like "club penguin" or "caldo de pollo." For instance, the use of these phrases is revealed in a threat posted on Twitter in Spanish
Image preprocessing: The following steps are used to clean up the data after the set of potential photos has been downloaded:

- Remove any images that aren't necessary, such icons.
- labels are one example.
- Throw away duplicate images: It's essential to keep duplicate data to a minimum for better analysis.
- Greyscale photographs should be discarded because the categorization models require that the images have an RGB colour composition (red, green, and blue).

Images of unknown formats should be discarded because the categorization model in the current project employs the.jpg format. If an image in a different format contains information that is significant, it is converted to a.jpg file.
Resize the images to ensure that they all have the same dimensions for the categorization models. The image size was set to 150150 pixels as a result.
The image bank was created using all of the photographs from dubious websites to test data. On the other hand, we utilised the following available datasets for training data:
www.flickr.com is the URL for Flickr.
UKBench is available at http://archive.org/details/ukbench.
(http://deeplearning.net/datasets/) Deep learning.
.

B. GEOMETRIC FEATURE EXTRACTION

Once the data is ready, the feature extraction process is started, as is detailed in Figure 4.

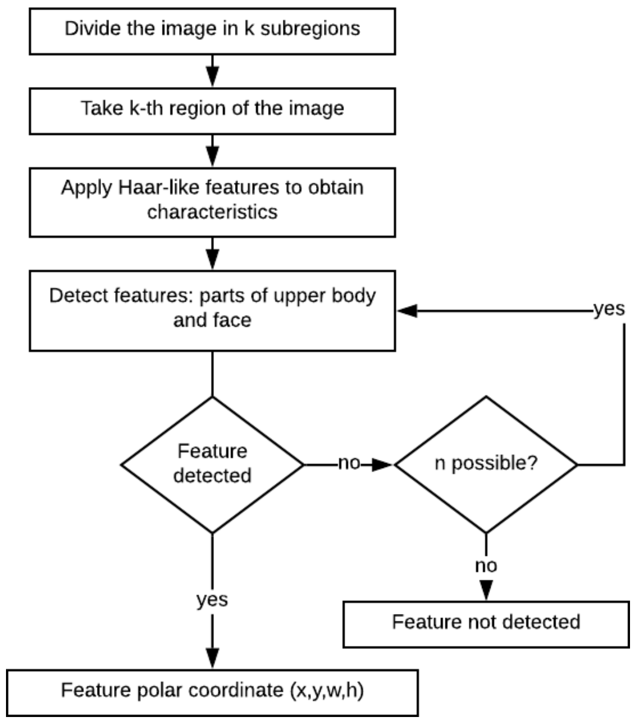


FIGURE 4. Feature extraction process. REFERENCE

The loaded image is first separated into k distinct local regions, and each one has a HAAR cascade classifier applied to it. It makes use of the Euclidian distance between the pixels and other geometric features to identify some visual patterns using the Viola-Jones technique. Additionally, these patterns are treated as distinct physical features of the upper torso, face, and eyes. In this study, the faces and upper bodies of the photos gathered from dubious websites were solely detected using geometric attributes.

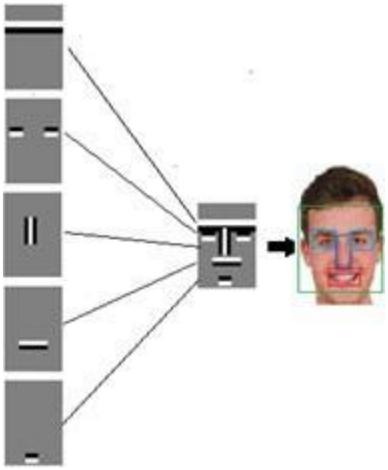


FIGURE 5. Face detection using Haar model(REFERENCE)

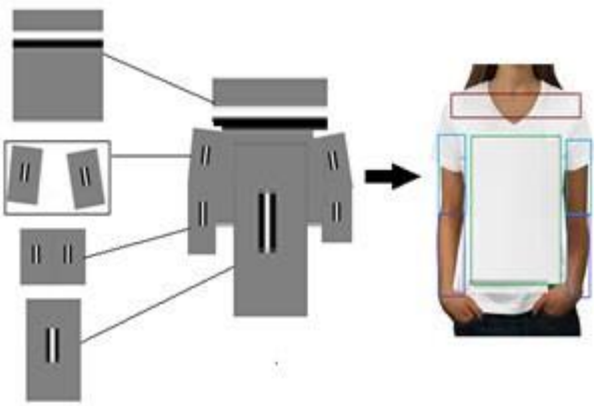


FIGURE 6. Upper body detection. (REFERENCE)

V. IMAGE CLASSIFICATION

In order to assess and contrast the image classification process, two algorithms—i) SVM utilising Haar features and ii) CNN—were used. The classification of each algorithm was then done for each age group (under fourteen years old) or for each gender (women or men).

A. SUPPORT VECTOR MACHINE - SVM

SVM stands for Support Machine Algorithm and it is one of most important algorithms of machine learning there are. Available in the market that gives the required output that is the particular result.

We used a linear kernel in this work because our variables are binary. The SVM prediction process is depicted .

A SVM algorithm took the extracted features as input. The boundary function f (x), defined by (1), where b is the bias value and yi, i is the Lagrange optimisation, is constructed by the SVM classifier using a linear kernel

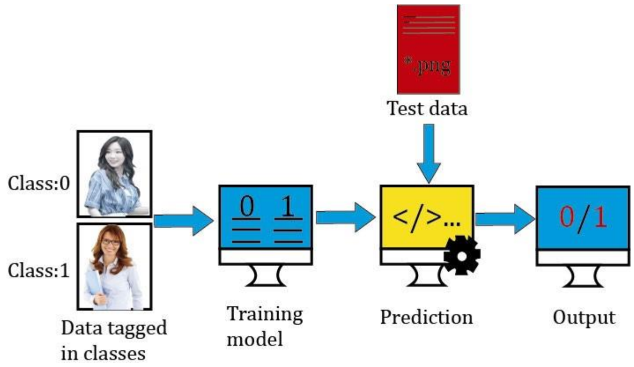


FIGURE 7. SVM model architecture. (REFERENCE)

parameters.

$$f(x) = \sum_{i=1}^{m} y_i \alpha_i K(x, x_i) \quad (1)$$

Furthermore, Classs 0 is for those who are younger than 14 years old and class 1 is for those who are older than 15 same as like that on the other hand it also does the same work like that with the elder additionaly a 16 layer CNN layer was the final choice that was determined by the final voting which was done inside the clas within a group of particular students so it was employed to compare the classification performance of these two algorithms I.e. the CNN and SVM which are used in this and hence the task that was required was completed successfully without any major problems that usually come.

B. CONVOLUTIONAL NEURAL NETWORK - CNN

A large image dataset is necessary for CNN, a supervised machine learning model, to create a classification model after several iterations. The main drawback of CNN is that it uses a lot of computational resources and takes a long time to train because it requires a lot of images and iterations to get the right prediction. In this work, an i7-3770 computer with 8 GB of RAM was used for testing purposes. It took 55 minutes to complete each iteration, so ten iterations of binary classification would take 9 to 10 hours. Using Haar-like features, the effectiveness of CNN and SVM was compared. Two leading indicators are considered for this purpose: (1) Accuracy, which counts the number of accurate predictions, and (2) Mean Square Error (MSE).

VI. TESTING RESULTS

In this section, the testing results for both SVM and CNN were described. Two of the algorithm that were used here were described and hence the results that came out were displayed which lead to the summary of the research that gave us these positive results.

Between these experiments, a comparison was made. It is significant to notice that each test categorises participants by age group and work activity.

A. SVM RESULTS

The upper-body characteristics can be used to determine age using the SVM algorithm, which can use Haar-like features. Because we frequently only have information on the upper-body and not the face, this combination enables classification based on the upper-body. Additionally, the training process must correctly classify and label the data into two different classes before the mean square error and accuracy indicator are used to gauge how well it performed. A labelled dataset with 4096 images from open repositories was made as training data. The majority of them were frontal or slightly profile, but there were also some with side views or others with faces, hats, caps, or glasses. 820 photos that were collected through the scrape of questionable websites were added to the testing data. In order to accomplish this, segments containing the face and upper body were extracted and categorised based on their Haar-like features.

First, a set of images where the face can be recognised was used in the classification, along with Haar filters. Table 7 displays the gender classification confusion matrix.

TABLE 7. Gender Difference.

	Man	Woman
Man	160	20
Woman	40	100

Following that, the accuracy and MSE performance indicators were determined by analysing the confusion matrix. The accuracy and MSE are 83% and 4.7%, respectively, as illustrated in Figure 8. Following that, Table 8 shows the classification by age group's confusion matrix. The results demonstrate that our model has a Mean Square Error of 3.8% and a classification accuracy of 83%. The results shown in Tables 7 and 8 demonstrate that experiment 1 classifies individuals based on their age group as well as their gender.The outcome that HAAR filters identified as upper body features is taken into consideration in the second experiment. First, the data were divided into gender categories (men and women).

The accuracy and MSE performance indicators were computed using the confusion matrix for gender classification. The values for accuracy and SME are 85.83% and 3.57%, respectively. The pictures were then categorised by age group.

FIGURE 8. People Job Categories that are taken as sample .

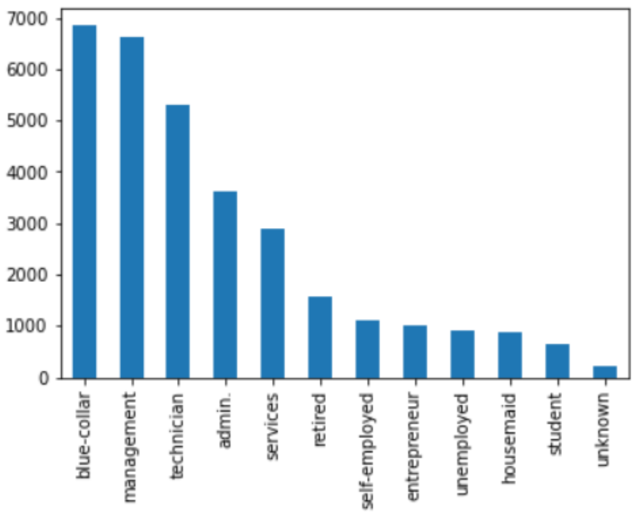


TABLE 8. Age group classification.

	Over14	Under14
Over14	160	20
Under14	40	100

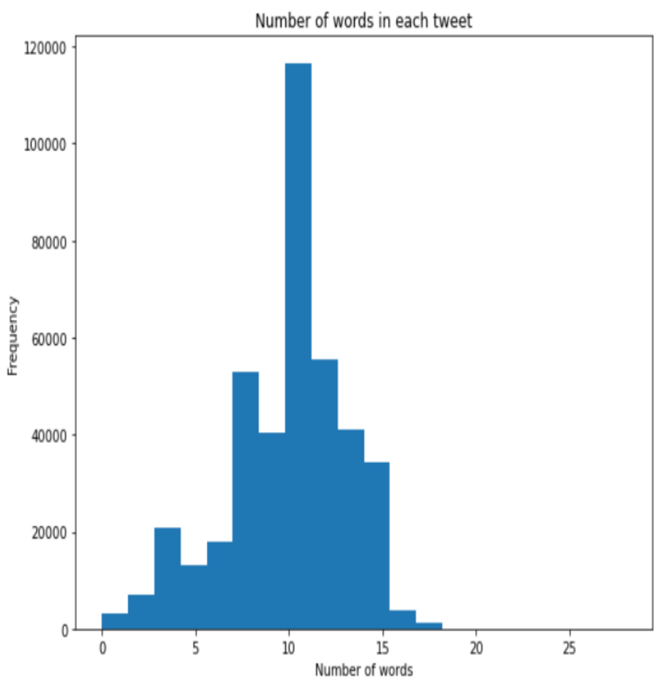


TABLE 9. gender classification (upper body).

	Man	Woman14
Man14	220	40
Woman	50	180

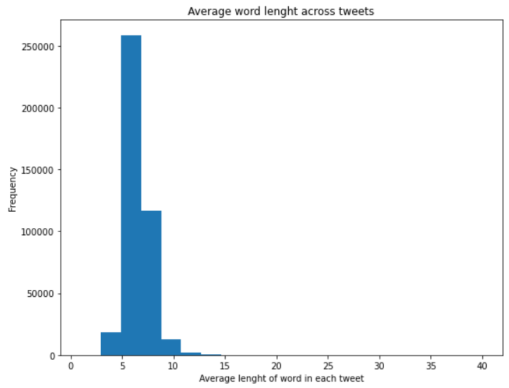


FIGURE 10. Tweets word classification (upper body).

TABLE 10. Age group classification.

	Over14	Under14
Over14	110	30
Under14	20	120

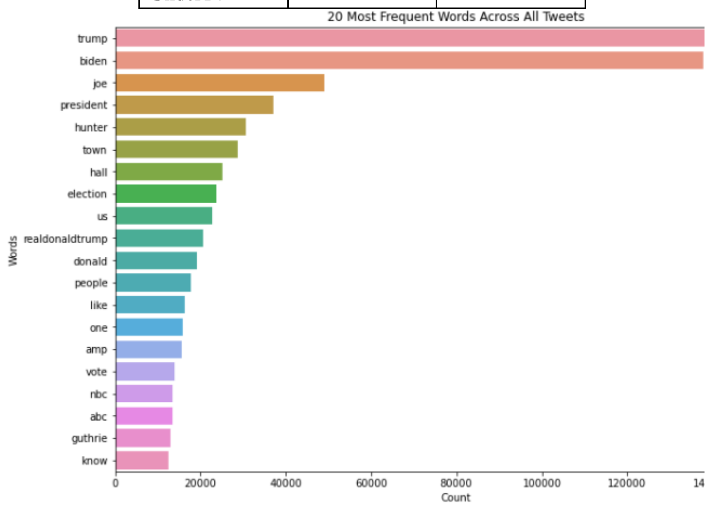


FIGURE 11. SVM age group classification (upper body).

1) SVM COMPARISON

Figure 12 displays the accuracy comparison between the results for the classification of the upper body and the face.

Figure 12 displays the upper body experiment's accuracy values, with age group accuracy at 82.1% and gender classification at 81.6%. The gender and age group accuracy values for face classification are 81,2% and 80,6%, respectively. These results evidence that the classification accuracy using the upper body is higher than the accuracy using faces. Tables 11 and 12 present the obtained indicators Accuracy, Precision, Recall, and F-measures to illustrate the effectiveness of the SVM model.

.

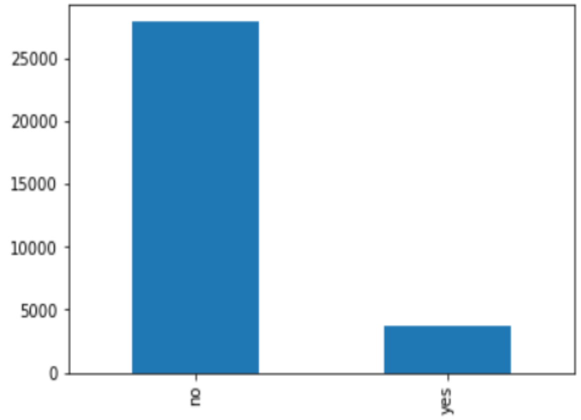


FIGURE 12. Spammer tweets.

TABLE 11. Performance SVM front From sample.

Face	Accuracy	Mean Square Error	Precision	Recall	F-measures
Gender	81,3%	3,5%	83,3%	71,4%	76,9%
Age Group	80,0%	3,75%	80,9%	66,6%	73,1%

TABLE 12. Performance SVM whole(from sample).

Upper Body	Accuracy	Mean Square Error	Precision	Recall	F-measures
Gender	81,6%	18,4%	81,82%	78,26%	80,0%
Age Group	80,0%	17,9%	84,6%	85,71%	85,16%

B. CNN RESULTS

This model involves a little human intervention, and it is essential to set a suitable number of iterations to minimise overfitting in classification results. The image dataset was divided into two groups: i) images with faces and ii) pictures with the upper body.

EXPERIMENT 1: FACES-BASED IMAGE CLASSIFICATION

The classification is Completed based on the face features for both Groups by male-female and by young-old group. The Figure below shows us the results of male-female classification.

As is shown in Figure below, the accuracy values of the first and tenth iteration are 86.5% and 99.2% respectively, and the last iteration has an SME value of 1,2%. The accuracy value has grown progressively during all iterations. After that, the results of the classification by age group are displayed in Figure next to the below figure.

The accuracy values are 82% and 94% in the first and last iteration, respectively, and the SME value of the last iteration is 96.3%. These results evidence that the classification accuracy using faces has a good result for both, gender and group of age.

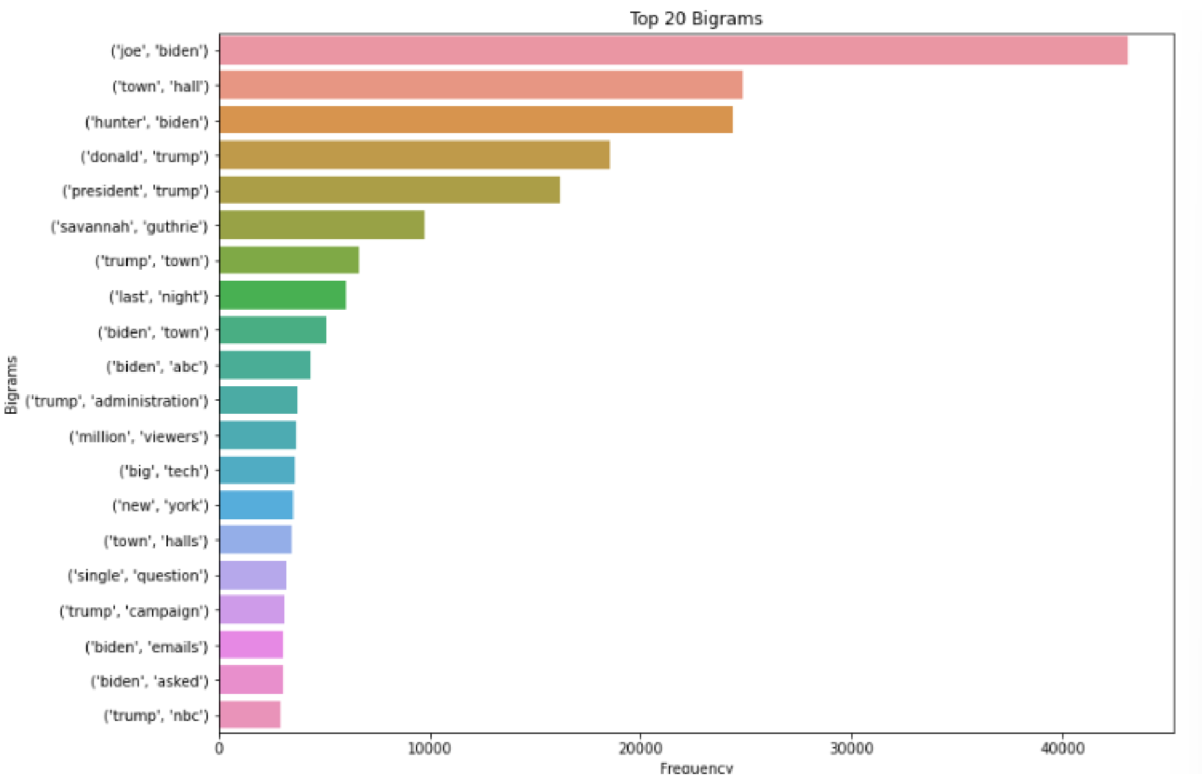
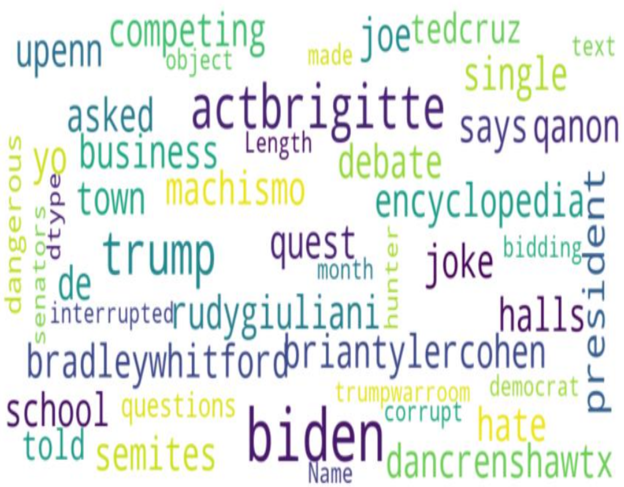


FIGURE 13. CNN age group classification (face).



EXPERIMENT 2 TWEET WORDS-BASED CLASSIFICATION

In this experiment, the tweet words was employed in order to classify photographs where the words are obscured, covered, or it is not displaying. It occurs when minors' faces are pixelated in order to conceal their identities. The result of this experiment for male-female classification is done.

In the gender classification, the accuracy values for the first and the tenth iterations are 52.7% and 66.5%, respectively. Moreover, the MSE value of the last iteration is 4.68%. In the same way, the age group classification using Tweet words characteristics is depicted in Figure 13.

The accuracy value in the first iteration is 53.4%. This value remains constant during the tenth iteration for this kind

FIGURE 15. CNN age group classification (upper body).

of classification. In reality, the accuracy and SME values for the last iteration are 51,4% and 48,6%, respectively, indicating this experiment has poor performance because the test identifies correctly one out of 2 examples. This result is equivalent to select a random image by the toss of a coin. The accuracy val- ues obtained in the last iteration for both gender and age group were 64,2% and 51,4%, respectively (Figure 15 and Figure 16). This experiment takes only upper body features into account, so this model has a poor performance. This outcome emerges because CNN has an acceptable prediction rate categorising faces but not the upper torso, like the SVM algorithm.

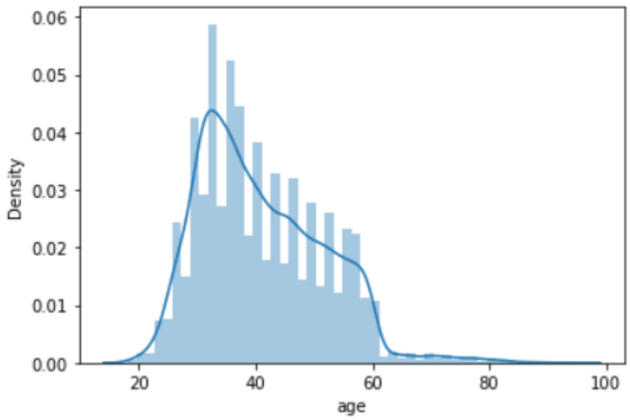


FIGURE 16. CNN Face and UpperBody.

1) CNN COMPARISON

This image illustrates that under any classifi- cation criterion (gender or age group), the CNN algorithm has a better performance when it evaluates the people to separate (facial features). However, CNN is not recommended when images only have upper-body features.For this experiment, the accuracy values for gender and age group are 64,2% and 51,3%, respectively. Consequently, if the images are analyzed, taking into account only upper body features, the SVM is a better option than the CNN model.

TABLE 13. CNN performance (face).

FACE	Accuracy	Mean Square Error
Gender	98,5%	1,2%
Age Group	97,3%	2,8%

TABLE 14. CNN performance (upper body).

UPPER BODY	Accuracy	Mean Square Error
Gender	64,2%	35,8%
Age Group	51,4%	48,6%

It is important to note that the main problem using only face features is that the images can be blurred, or they do not exist, especially in web sites that promote minor trafficking. As a result, the classification using CNN, under this con- dition, does not provide sufficient accuracy. A summary of accuracy and SME values for CNN experiments are shown in Tables 13 and 14.

C. COMPARATIVE ANALYSIS

In this section, a comparison between SVM and CNN per- formance results, taking into account accuracy value, is pre- sented. Firstly, the results obtained when face features can be detected from an image are analyzed. The accuracy values for both algorithms are displayed in Figure 18.

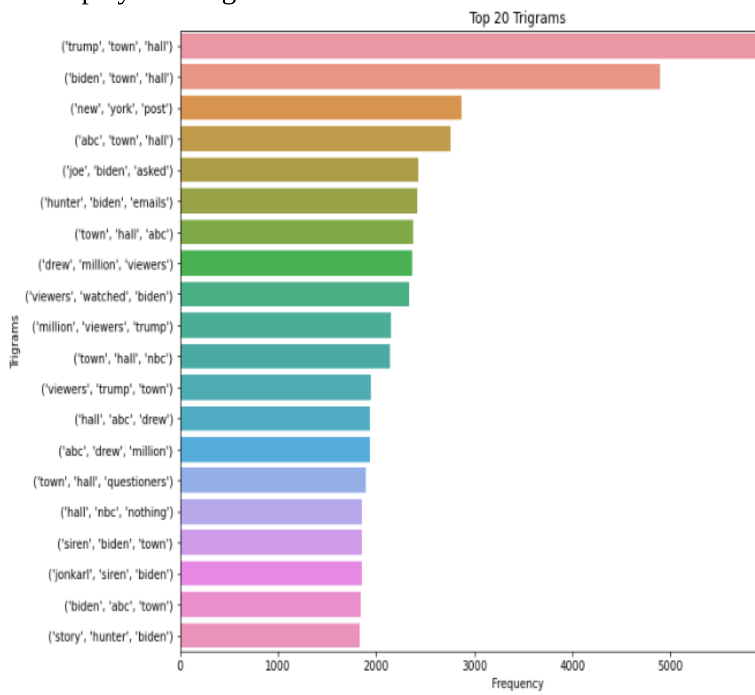
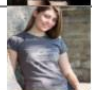


FIGURE 17. Top 20 Trigams.

The results reported (Figure 18) shows that the accuracy of CNN is higher than the accuracy of the SVM model for both gender and age group categorization. On the one hand,

On the other one hand, models have good accuracy values, so both can be used to classify images with facial features. Then, the findings of experiment 2 were compared, that it comprises of the image classification taking into consideration upper body attributes.

This image illustrates that under any classifi- cation criterion (gender or age group), the CNN algorithm has a better performance when it evaluates the people to separate (facial features). However, CNN is not recommended when images only have upper-body features that is the tweet words

Image Number	Gender	Group Age	Image
1	Man	Under 14 years old	
2	Woman	Over 14 years old	
3	Man	Over 14 years old	
4	Woman	Under 14 years old	
5	Woman	Over 14 years old	
6	Woman	Under 14 years old	
7	Woman	Under 14 years old	
8	Woman	Over 14 years old	
9	Man	Under 14 years old	
10	Man	Over 14 years old	

models have good accuracy values, so both can be used to classify images with facial features. Then, the findings of experiment 2 were compared, that it comprises of the image classification taking into consideration upper body attributes.

The results reported in demonstrate that the classi- fication accuracy of SVM is higher than the accuracy of the CNN model when just upper body features in the images are evaluated. On the one hand, the accuracy values of SVM for gender and age group classification are 79%. On the other hand, the accuracy values of CNN for gender and age group classification are 78% .

These results demonstrate that the SVM accuracy in experi- ment 1 (face) is similar to experiment 2 (upper body) not only for gender classification but also for age group classification. Moreover, the CNN performance is worse than not only SVM performance but also CNN results produced in experiment 1. Therefore, the best option to detect a possible case of human trafficking of minors is using the SVM algorithm. As men- tioned above, trafficking web sites usually use blurred or pixelated images, or there are no facial features in the image. Figure 20 shows some images classified in this research.

The images with blurred or pixelated faces were classified using the SVM algorithm. Moreover, CNN is commonly used in face detection.

VII. CONCLUSIONS AND FUTURE WORK

In recent years, advances have been made in machine learning models and face recognition algorithms. For example, in the ILSVRC competition, an accuracy value of 90%was obtained. In these circumstances, machine learning recognition may be comparable to human visual object identification. Many factors have a direct impact on image recognition, such as size, color, opacity, resolution, kind of image format, among others. Therefore, the results of image recognition and classification depend on the dataset quality.

In this study, we investigated whether successful performance could be achieved with only geometric features of the torso, rather than just facial features. For this paper, Haar filters combined with an SVM classifier were utilised for the extraction process of features, and then we categorised the age group and gender with an SVM classifier. The resulting findings were contrasted with the outcomes of a CNN algorithm.

SVM is a frequently used model, and in this study, we were able to accurately classify both gender and age groups for both tests with an accuracy rate of more than 80%. Our primary contribution to this paper is the image classification using the upper body to identify the age range and identify human trafficking.

To the best of our knowledge, this work is the first approach related to image classification without facial features but just the upper-body geometric characteristics. Currently, there is no analogous research that takes into account simply the upper body features of minors. Consequently, this paper's findings can

be applicable to human trafficking, disappearance, kidnapping, among others. The police or other security institutions may also use the information that was gathered.

The study of some characteristics related to ethnic and racial features, the extension of the proposal to extract geometric features from the entire body, other types of images, or inclusive videos in various formats, the detection of medical issues through the analysis of features extracted from torso images, legs, and back, among other characteristics, and the use of other algorithms or the applicability in other networks like Instagram are all examples of future work.

ACKNOWLEDGMENT

The authors would like to thank the Escuela Politécnica Nacional for creating Seed Research Project PIS-17-10 ‘‘Data Mining, Feature Vector Extraction, and Modeling using Pattern Recognition and Machine Learning to Detect Scenarios linked to the Crime of Human Trafficking’’.

REFERENCES

[1] F. Laczko, ‘‘Data and research on human trafficking,’’ *Int. Migration*, vol. 43, nos. 1–2, pp. 5–16, Jan. 2005.

[2] M. Lee, ‘‘Human trafficking and border control in the global south,’’ in *The Borders of Punishment: Migration, Citizenship, and Social Exclusion*. Oxford, U.K.: Oxford Univ. Press, 2013, pp. 128–149.

[3] E. Cockbain and E. R. Kleemans, ‘‘Innovations in empirical research into human trafficking: Introduction to the special edition,’’ *Crime, Law Social Change*, vol. 72, no. 1, pp. 1–7, Jul. 2019.

[4] R. Weitzer, ‘‘Human trafficking and contemporary slavery,’’ *Annu. Rev. Sociol.*, vol. 41, pp. 223–242, Aug. 2015.

[5] J. V. D. Wolfshaar, M. F. Karaaba, and M. A. Wiering, ‘‘Deep convolutional neural networks and support vector machines for gender recognition,’’ in *Proc. IEEE Symp. Ser. Comput. Intell.*, Dec. 2015, pp. 188–195.

[6] M. Hernandez-Alvarez,‘‘Detection of possible human trafficking in Twitter,’’ in *Proc. Int. Conf. Inf. Syst. Softw. Technol. (ICIST)*, Nov. 2019, pp. 187–191.

[7] H. Alvari, P. Shakarian, and J. E. K. Snyder, ‘‘A non-parametric learning approach to identify online human trafficking,’’ in *Proc. IEEE Conf. Intell. Secur. Informat. (ISI)*, Sep. 2016, pp. 133–138.

[8] M. M. Dehshibi and A. Bastanfard, ‘‘A new algorithm for age recogni- tion from facial images,’’ *Signal Process.*, vol. 90, no. 8, pp. 2431–2444, Aug. 2010.