On Detecting Domestic Abuse via Faces

Puspita Majumdar, Saheb Chhabra, Richa Singh, and Mayank Vatsa
IIIT Delhi, India
{pushpitam, sahebc, rsingh, mayank}@iiitd.ac.in

Abstract

Domestic violence is considered a major social problem worldwide. Different countries have enacted the law to contain and protect the victims of domestic violence. In order to understand the nature of domestic violence, medical professionals and researchers have performed manual analysis of facial injuries. The aim of these studies is to find commonly affected facial regions, to determine the types of maxillofacial trauma associated with domestic violence, and to distinguish the injuries of domestic violence from accidents. Analysis of these injuries assist the service providers in providing proper treatment to the victims as well as facilitate law enforcement investigation. This paper automates the process of analyzing the facial injuries to distinguish the victims of domestic abuse from others. For this purpose, Domestic Violence Face database of 450 subjects with two classes namely, Domestic Violence and Non-Domestic Violence, is prepared. The paper also presents a novel framework using activation maps of deep learning features for determining whether an image belongs to domestic violence class or not. The results on the proposed database show that deep learning based framework is effective in detecting domestic injuries.

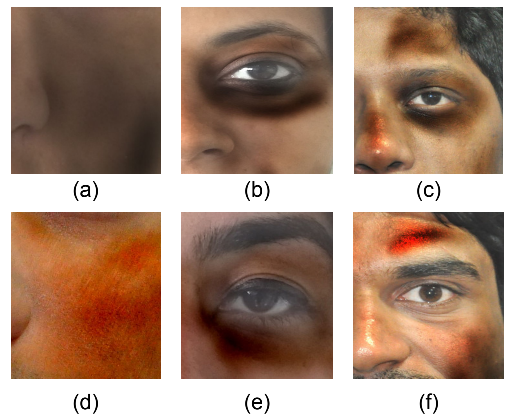


Figure 1: Sample images showing visual similarities in the injuries of domestic violence and non-domestic violence cases: (a), (b) and (c) are the images of the facial injuries of domestic violence, (d), (e) and (f) show the images of non-domestic violence facial injuries, (a) and (d) show similarities in the cheek region, (b) and (e) show similarities in the eye region, and (c) and (f) show similarities in the forehead region.

1. Introduction

Domestic violence is an aggressive behavior used by one person to dominate or to gain control over another in a domestic setting. It affects people of all age, gender, nationality, socio-economic status, religion, and culture [2]. It is a global problem and every year such cases are increasing at a rapid pace. According to National Coalition Against Domestic Violence (NCADV), every minute an average of 20 people become a victim of domestic violence, which in turn result in 10 million victims annually [3]. According to the World Health Organization (WHO), 35% of women across the world experience domestic violence in their lifetime [5].

Domestic violence is an offense and statements of the victim are considered sufficient evidence to charge the accused person. However, in majority of the cases, victims change their statement during the proceedings of the trial. In such scenarios, many times the accused is set free in absence of any other supporting evidence [1, 4]. A crime goes unpunished which in turn increases the risk of future domestic violence on the victim. Moreover, the cases of false allegations of domestic violence are also becoming common. Every year, the number of such cases are increasing rapidly [1]. False cases lead to serious legal trouble for the accused. These challenges can only be addressed if there is a means to identify the real victim of domestic violence.

Due to intraclass variations in domestic violence injuries and interclass similarities with other injuries, identifying the victims is a difficult problem. Specifically, many times the injuries of domestic violence look visually similar to the injuries of non-domestic violence. Figure 1[1] illustrates the

[1]The figure contains synthetically created injuries for illustrative purpose only.

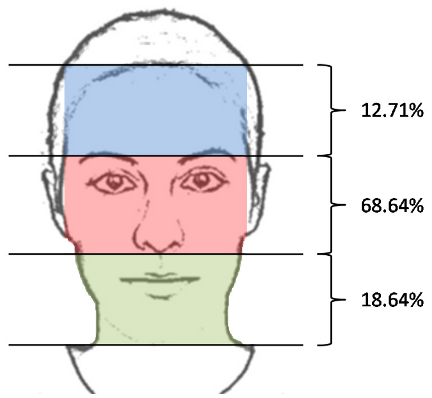


Figure 2: Distribution of injuries in facial images during domestic violence [11]. Red region represents the most affected region. Green and yellow regions are less susceptible than the red region.

similarities in the injuries of domestic violence and non-domestic violence. In medical literature, research has been conducted on analyzing the facial injuries of the victims of domestic violence [8, 9, 12, 15, 19]. As illustrated in Figure 2, it is found that the middle third of the face is mostly affected (68.64%), followed by lower third (18.64%) and upper third (12.71%) [11]. All these studies are observational (manual) and require an expert's help. The manual process is time-consuming, require physical labor, and non-scalable. Therefore, an automated system to analyze and classify injuries (into domestic and non-domestic classes) is essential in helping service providers for treatment and law enforcement agencies in investigation.

With the advancement of deep learning and machine learning techniques, the area of biometrics has reached a new level of success, especially in facial analysis and recognition [20]. These advanced techniques can be used to detect the cases of domestic violence by analyzing face images. Currently, there is no automated system to detect the victims of domestic violence. Discriminative features in battered face images of the victims can be used by the classification algorithms to distinguish them from other faces. However, a solution is always encompassed with multiple challenges. Following are some of the scenarios that make the problem difficult:

1. During a violence, facial features get heavily affected by the injuries. Automatic face detection becomes difficult in the presence of injured facial features.

2. Injuries during accidents and sports like boxing are similar to the injuries of domestic violence (i.e., interclass similarities).

3. Identifying discriminative features of domestic violence may have large intraclass variations.

1.1. Related Work

In the literature, researchers have studied and analyzed the injuries caused by domestic violence. These studies are primarily observational (manual) and the aim of these studies is to find the commonly affected parts of the face by domestic violence. The studies also focus on the analysis of the type of facial fractures associated with domestic violence. For instance, Goulart et al. [9] have studied the abnormalities of maxillofacial injuries caused by interpersonal violence and noted that interpersonal violence results in a higher rate of facial fractures. Therefore, domestic violence is called as interpersonal violence as well. The pattern of oral maxillofacial trauma caused by domestic violence is studied by Ferreira et al. [8]. According to the study, domestic violence is identified as a factor for the fracture of jaw bone, bruises on face, and injury of tooth-supporting structure. Lee [12] has observed that jawbone and cheekbone are commonly fractured by domestic violence. In some studies, the association between alcohol and domestic violence is established. For this purpose, OMeara et al. [15] have studied the increase in the severity of facial fractures with alcohol and interpersonal violence. Apart from these studies, a recent study focuses on differentiating facial injuries of domestic violence and accidents [19]. The authors observed that the injuries of ear, nose, upper and lower jaw point indicate a criminal violence.

1.2. Research Contributions

To the best of our knowledge, there is no automated system to detect or to classify the victims of domestic violence. This paper attempts to automate the process of analyzing facial injuries for identifying the victims of domestic violence. For this purpose, Detection of Domestic Violence (DDV) framework is proposed to learn the discriminative features by paying attention to the injured regions of the facial image. Key contributions of this research are:

1. Proposing a novel Domestic Violence Face (DVF) database of 450 subjects with two classes namely, Domestic Violence (DV) and Non-Domestic Violence (NDV).

2. Proposing DDV framework to distinguish faces of the victims of domestic violence from others.

2. Domestic Violence Face (DVF) Database

The main challenge of this research is the creation of a database to address the problem of identifying the victims of domestic violence. Majority of the times the face images are not available, thereby making it difficult to find the images of the victims. To prepare the DVF database we collect the images from Internet, specifically from news articles, and

Table 1: Description of the proposed DVF database.

		No. of Images
Domestic Violence	**Injured Faces**	150
Non-Domestic Violence	**Injured Faces**	85
	Normal	130
	Others	85
Total		450

social media[2]. The proposed DVF database consists of 450 images with one image per subject. The database contains two classes namely, 1) Domestic Violence (DV) with 150 images and 2) Non-domestic violence with 300 images. The details of both the classes are discussed below and Table 1 presents the summary of the database.

2.1. Domestic Violence (DV)

DV class contains the images of real victims of domestic violence. Images of the victims published in news or in social media are taken to create the database. It also includes images of the celebrities with domestic violence makeup to increase awareness about the offense. Majority of the images are frontal with few non-frontal images. While in majority of cases, eye regions are badly affected by domestic violence, bruises and swellings are also common in the injuries of domestic violence. According to Sheridan et al. [17], the most commonly wounded part is the middle third part of the face. As 90% of the population is right-handed, therefore mostly the left side of the face is more injured as compared to the right side. Victims often turn right for avoiding hits [6] that result in hurting left part of the face in most of the cases. This phenomena is also observed in the proposed database.

2.2. Non-Domestic Violence (NDV)

As shown in Figure 3, bruises and swellings are common in sports such as boxing and wrestling. These injuries look visually similar to the injuries of domestic violence. Some diseases such as chicken pox, measles, and vitiligo change the appearance of the face. Use of illicit drug also affect the facial features to a great extent and long-term usage cause skin sores and scars [22]. Further, scars, marks and tattoos also change the facial appearance. For correct classification of the victims, the algorithms must be robust to handle such changes in facial features. For this reason, NDV class is created to incorporate multiple variations in faces. This class contains faces with non-domestic violence injuries, faces with diseases, tattoos, face packs, dirty face images, and drug abuse faces. Some clean face images in a controlled and uncontrolled environment are also included.

[2]The URLs to these images will be released to the research community as a part of the research efforts.

Figure 3: Sample images from the DVF database: (a) samples of domestic violence injuries and (b) samples of non-domestic violence injuries. Facial injuries (eye, cheek and forehead) show visible similarities in domestic and non-domestic violence.

To create NDV class, images are download from the Internet and taken from CMU-MultiPIE [10] and CelebA [13] databases. These variations can help to evaluate the robustness of the algorithms in different challenging scenarios.

3. Framework for Detection of Domestic Violence (DDV)

For distinguishing domestic and non-domestic violence victims, facial features are required to be learned from the injured regions. For this purpose, a deep-learning based framework is proposed to learn the discriminative features from the injured regions of the face image and identify the victims of domestic violence. As shown in Figure 4, the DDV framework uses class activation maps that provide weights to the features learned from the injured regions. In the proposed DDV framework, VGGFace model [16] is fine tuned over the DVF dataset and corresponding class activation maps are generated.

Feature Extraction and Class Activation Map: Let $\mathbf{F}$ be the set of feature map of last convolutional layer and $F_k(x, y)$ represent the value of the k^{th} feature map at spatial location (x, y). Let $\mathbf{C}$ be the class set with two classes namely $\{D, N\}$. The aim is to generate the class activation map $\mathbf{M}$ using feature map $\mathbf{F}$. Mathematically it is represented as:

$$\mathbf{M} = f(\mathbf{F}) \tag{1}$$

where, $f(.)$ is a function to generate class activation map $\mathbf{M}$ from feature map $\mathbf{F}$. To generate $\mathbf{M}$, global average pooling is first applied on the last convolutional layer representing feature map set $\mathbf{F}$. The output of the global average pooling is the scalar value P_k corresponding to each feature map $\mathbf{F_k}$. Mathematically, it is written as:

$$P_k = \frac{1}{xy} \sum_{x,y} F_k(x, y) \tag{2}$$

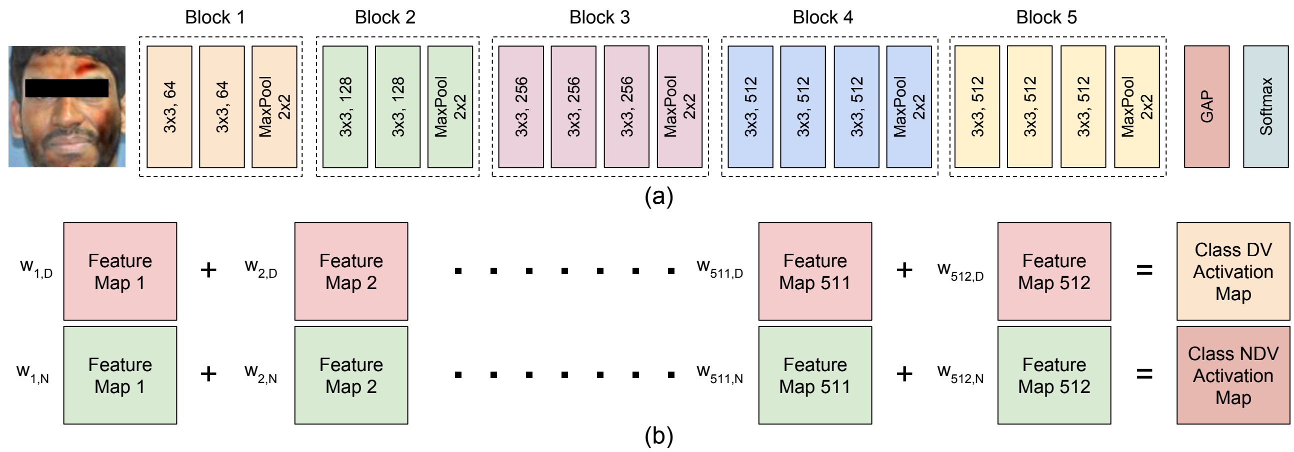


Figure 4: (a-b) illustrates the block diagram of the proposed framework. (a) shows the architecture of VGGFace model. The model is fine-tuned over DVF database and weights are learned for classification. (b) illustrates the generation of class activation maps corresponding to each class.

Once the output from the global average pooling is obtained, the weights $\mathbf{W}$ of the fully connected Softmax layer learned during training is used. As mentioned in [23], the best representation for generating class activation map is the weighted sum of the individual feature map. Therefore, class activation map (M) corresponding to image of class $\{D, N\}$ is represented as:

$$M_D = \sum_k w_{k,D} \mathbf{F_k} \quad \forall k \tag{3}$$

$$M_N = \sum_k w_{k,N} \mathbf{F_k} \quad \forall k \tag{4}$$

where, the $w_{k,D}$ and $w_{k,N}$ represent the weights learned for k^{th} feature map corresponding to class D and N respectively.

Classification: In order to classify the learned features into class $\{D, N\}$, the output of global pooling layer P_k is given as input to the Softmax layer.

$$S_D = \sum_k w_{k,D}, P_k \tag{5}$$

$$S_N = \sum_k w_{k,N}, P_k \tag{6}$$

The output of the Softmax for class C is represented as:

$$S_C = \frac{e^X}{\sum_{C \in \{D,N\}} e^{X_C}} \tag{7}$$

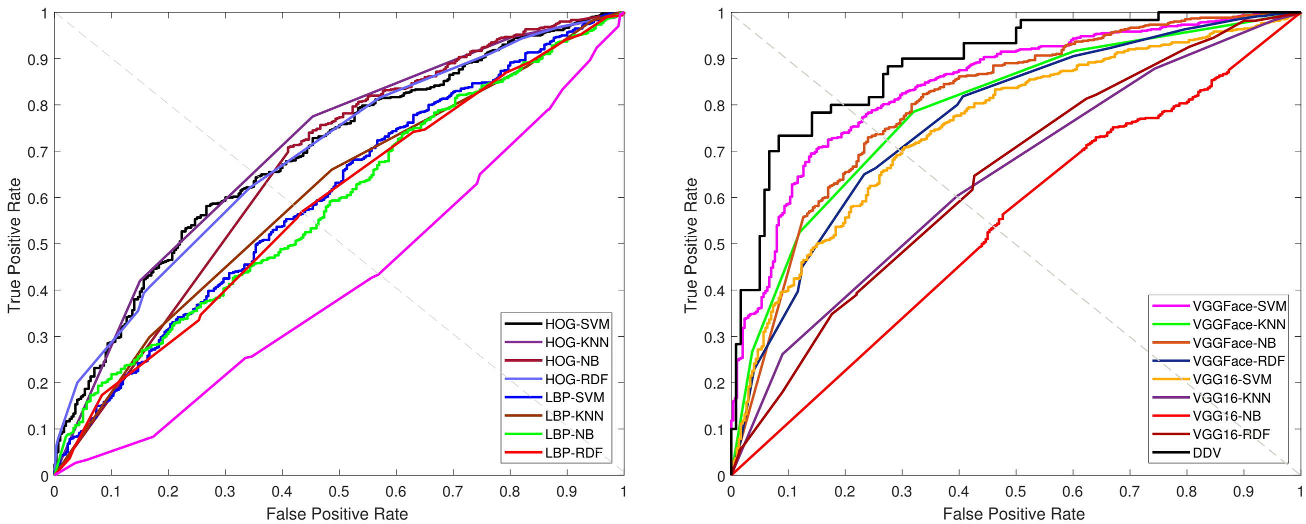
where, $X \in \{S_D, S_N\}$ and S_C represents the output of the Softmax for class C.

4. Experiments and Results

Viola-Jones face detector [21] is applied to all the images to segment the facial region. The detector failed to detect 91 images which are cropped manually for further processing. For experimental evaluation, the DVF database is partitioned into training and testing sets with 60% images in training and 40% images in the testing set corresponding to each class. This results in a total of 270 images in training set and 180 images in testing set. Experiments are performed with five times repeated random sub-sampling for training and testing partitioning. Further, experiments are performed on Tensorflow with Nvidia GTX 1080Ti GPU and Intel Core i7-8700 CPU with 64 GB of RAM.

4.1. Baseline Performance

To evaluate the baseline performance on the DVF database, two deep learning models, VGGFace [16], VGG16 [18], and two handcrafted feature extraction algorithms, Local Binary Pattern (LBP) [14] and Histogram of Oriented Gradient (HOG) [7] are used. Features extracted using LBP, HOG, pre-trained VGGFace, and pre-trained VGG16 are used to train four different classifiers: Support Vector Machine (SVM), K-Nearest Neighbor (K-NN), Naive Bayes (NB), and Random Decision Forest (RDF). Results and confusion matrix of the baseline algorithms are shown in Table 2 and Table 3, respectively. Table 2 shows that VGGFace descriptor with SVM outperforms existing baseline algorithms. Receiver Operating Characteristic (ROC) curves of the handcrafted feature extractors corresponding to all the classifiers are shown in Figure 5a. Figure 5b shows the ROC curves of the pre-trained deep CNN models and the proposed framework with class activation maps. Confusion matrix shown in Table 3 indicates that



(a) LBP and HOG with four different classifiers.

(b) Pretrained VGG16, VGGFace and proposed framework.

Figure 5: ROC curves showing the performance of the baseline algorithms and proposed DDV framework on the DVF database.

Table 2: Classification accuracy (%) of the baseline and proposed framework

Feature	Classifier	Accuracy @ EER
LBP	SVM	56.42
	NB	54.00
	RDF	56.75
	KNN	58.67
HOG	SVM	64.33
	NB	64.42
	RDF	63.75
	KNN	66.08
VGG16	SVM	70.00
	NB	53.92
	RDF	59.67
	KNN	60.33
VGGFace	SVM	77.00
	NB	73.67
	RDF	70.50
	KNN	73.25
	DDV Framework	**80.00**

False Reject Rate (FRR) is high for all the cases. High FRR is an indication of the misclassification of large number of samples of the DV class as NDV class. For instance, 80% samples pertaining to domestic violence class are misclassified as non-domestic violence using LBP-RDF. The reason behind the misclassification is the inability of the algorithms in distinguishing the discriminative features among the images of the two classes.

4.2. Performance Evaluation of the Proposed Framework

In order to evaluate the performance of the proposed framework on the DVF database, VGGFace model is fine-tuned. ROC curves of the proposed framework, pre-trained VGGFace, and VGG16 model are shown in Figure 5b. The accuracies of the baseline algorithms and the proposed framework at Equal Error Rate (EER) are shown in Table 2. The proposed framework shows 3% improvement over the best performing baseline algorithm, i.e. pre-trained VGGFace-SVM. Further, confusion matrix of VGGFace is compared with the proposed framework in Table 4. The proposed framework shows lower FRR compared to the VGGFace-SVM. The improvement in overall performance of the proposed algorithm can be attributed to the class activation maps computed using deep learning models which helps in distinguishing DV and NDV classes.

5. Conclusion

Domestic violence is a social problem and automatically classifying such injuries from non-domestic violence injuries help in specialized treatments and law enforcement investigations. With the goal of “Artificial Intelligence for Social Cause”, this paper introduces Domestic Violence Face database and a deep learning framework for detecting such injuries. Experiments are performed with multiple feature descriptors and classifiers, and the results show the ad-

Table 3: Confusion matrix (%) of baseline algorithms on the DVF database.

	Algorithm		SVM		Naive Bayes		RDF		KNN	
			DV	NDV	DV	NDV	DV	NDV	DV	NDV
Ground Truth	LBP	DV	45	55	48.3	51.7	20	80	15	85
		NDV	27.5	72.5	34.17	65.83	10	90	8.3	91.7
	HOG	DV	51.7	48.3	55	45	31.7	68.3	23.3	76.7
		NDV	19.1	80.9	21.6	78.4	7.5	92.5	2.5	97.5
	VGGFace	DV	68.3	31.7	76.7	23.3	38.3	61.7	80	20
		NDV	12.5	87.5	26.7	73.3	5.8	94.2	25	75
	VGG16	DV	51.7	48.3	43.3	56.7	25	75	36.7	63.3
		NDV	19.2	80.8	20	80	5.8	94.2	13.3	86.7

Table 4: Confusion matrix (%) of VGGFace-SVM and the proposed framework.

	Algorithm	Predicted		
			DV	NDV
Ground Truth	VGGFace-SVM	DV	68.3	31.37
		NDV	12.5	87.5
	DDV Framework	DV	73.33	26.67
		NDV	9.17	90.83

vantages of the proposed algorithm, specifically in reducing false reject cases. In future, we plan to extend the database and improve the domestic violence classification algorithm.

6. Acknowledgement

This research is partly supported by the Infosys Center of Artificial Intelligence, IIIT Delhi, India.

References

[1] Being Falsely Accused of Domestic Violence. http://www.separateddads.co.uk/falsely-accused-domestic-violence.html. 1

[2] Domestic violence issues. https://tinyurl.com/ycgxpnnk. 1

[3] Statistics. https://ncadv.org/statistics. 1

[4] The Recanting Victim and Domestic Violence. https://tinyurl.com/yaj8jdqo. 1

[5] Violence against women. http://www.who.int/mediacentre/factsheets/fs239/en/. 1

[6] C. Anyanechi. Mandibular fractures associated with domestic violence in calabar, nigéria. *Ghana Medical Journal*, 44(4), 2010. 3

[7] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 886–893, 2005. 4

[8] M. C. Ferreira, A. M. Batista, F. d. O. Ferreira, M. L. Ramos-Jorge, and L. S. Marques. Pattern of oral–maxillofacial trauma stemming from interpersonal physical violence and determinant factors. *Dental Traumatology*, 30(1):15–21, 2014. 2

[9] D. R. Goulart, L. do Amaral Colombo, M. De Moraes, and L. Asprino. What is expected from a facial trauma caused by violence? *Journal of Oral & Maxillofacial Research*, 5(4), 2014. 2

[10] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-pie. *Image and Vision Computing*, 28(5):807–813, 2010. 3

[11] B. T. Le, E. J. Dierks, B. A. Ueeck, L. D. Homer, and B. F. Potter. Maxillofacial injuries associated with domestic violence. *Journal of Oral and Maxillofacial Surgery*, 59(11):1277–1283, 2001. 2

[12] K. H. Lee. Interpersonal violence and facial fractures. *Journal of Oral and Maxillofacial Surgery*, 67(9):1878–1883, 2009. 2

[13] Z. Liu, P. Luo, X. Wang, and X. Tang. Deep learning face attributes in the wild. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3730–3738, 2015. 3

[14] T. Ojala, M. Pietikäinen, and D. Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29(1):51–59, 1996. 4

[15] C. OMeara, R. Witherspoon, N. Hapangama, and D. M. Hyam. Alcohol and interpersonal violence may increase the severity of facial fracture. *British Journal of Oral and Maxillofacial Surgery*, 50(1):36–40, 2012. 2

[16] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. In *British Machine Vision Conference*, 2015. 3, 4

[17] D. J. Sheridan and K. R. Nash. Acute injury patterns of intimate partner violence victims. *Trauma, Violence, & Abuse*, 8(3):281–289, 2007. 3

[18] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 4

[19] V. Sterzik, D. Duckwitz, and M. Bohnert. Accident or crime? About the meaning of face injuries inflicted by blunt force. *Forensic Sciences Research*, 1(1):14–21, 2016. 2

[20] M. Vatsa, R. Singh, and A. Majumdar. *Deep Learning in Biometrics*. CRC Press, 2018. 2

[21] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *IEEE Computer Society Con-*

ference on Proceedings of the Computer Vision and Pattern Recognition,, volume 1, pages I–I, 2001. 4

[22] D. Yadav, N. Kohli, P. Pandey, R. Singh, M. Vatsa, and A. Noore. Effect of illicit drug abuse on face recognition. In *IEEE Winter Conference on Applications of Computer Vision*, pages 1–7, 2016. 3

[23] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba. Learning deep features for discriminative localization. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2921–2929, 2016. 4