# ON THE ROBUSTNESS OF DRUG ABUSE FACE CLASSIFICATION

**Hruturaj M. Dhake, Akshay Agarwal**
Department of Data Science and Engineering
Indian Institute of Science Education and Research, Bhopal

{hruturaj19,akagarwal}@iiserb.ac.in

ABSTRACT

Face recognition is one of the secure mediums to access various security-restricted areas such as border control and mobile unlocking. However, face recognition can be severely impacted due to several factors including illicit drug abuse on the facial regions. These abuses drastically alter the appearance of the faces and hence lead to the poor performance of the face recognition algorithms. However, due to the limited availability of the datasets, the research in this field is still in the novice stage. To advance the research, in this research, 'we have collected drug abuse face images and proposed a benchmark study to identify whether the face in question is clean or drug abused'. *Further, we have performed the robustness study of detection networks by altering the images by adopting several enhancement filters popular to use before uploading the face images on social media platforms.*

## 1 INTRODUCTION

An estimated 246 million people between the ages of 15 and 64 have used illegal drugs, primarily stimulants similar to amphetamines, cocaine, or cannabis (Yadav et al., 2016). Further, according to the World Drug Report, illicit drug abuse is a global issue, and more than $750,000$ drug-related deaths were recorded in 2019 (Li et al., 2021). The concern of drug abuse is not only limited to health but reflects the vulnerabilities of current facial recognition algorithms including commercial systems (Harastani et al., 2020; Ross et al., 2022; Raghavendra et al., 2016; Pandey et al., 2016). (Ross et al., 2022) have studied the impact of various factors including illicit drug abuse in the identification of health cues such as body mass index (BMI) and stress from face images. (Harastani et al., 2020) have conducted a study on the identification of the impact of illicit drugs on face symmetry. Moreover, due to the lack of publicly available datasets, the research field has not seen significant progress. A limited amount of work has been done to identify drug abuse faces and improve face recognition performance. To tackle this limitation, in this research, we have proposed a novel drug face abuse dataset and presented a benchmark study by evaluating the effectiveness of state-of-the-art (SOTA) convolutional neural networks (CNNs) for drug abuse face detection.

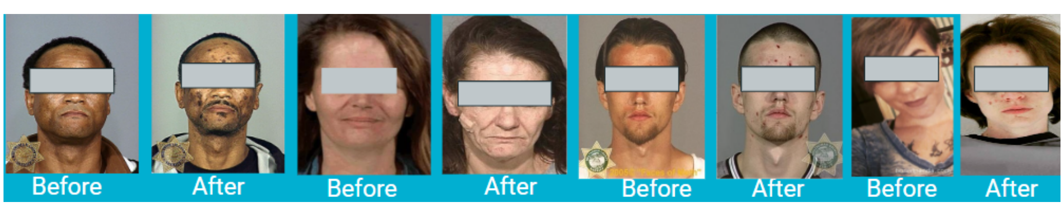


Figure 1: A few drug abuse samples of the proposed dataset reflect the sharp changes in the facial features including texture and symmetry.

## 2 PROPOSED DRUG ABUSE FACE DETECTION

In this section, we present a brief analysis of the proposed dataset followed by a description of the benchmarking algorithms developed for illicit drug abuse face detection.

Table 1: Drug abuse face classification performance of the benchmark image classification networks.

| CNN Model | Precision | Recall | F1 Score | Accuracy | AUC-ROC |
|---|---|---|---|---|---|
| VGG 16 | 0.75 | 0.60 | 0.67 | 0.70 | 0.79 |
| Xception | 0.70 | 0.64 | 0.67 | 0.68 | 0.70 |
| **DenseNet** | **0.81** | **0.84** | **0.82** | **0.82** | **0.88** |
| InceptionV3 | 0.79 | 0.76 | 0.78 | 0.78 | **0.88** |

**Dataset:** The proposed drug abuse dataset contains facial images of before and after drug abuse faces of 115 individuals. Since these images are collected from Internet sources, they reflect several unconstrained factors such as illumination, pose, and occlusion. Due to this unconstrained nature, classification of drug abuse faces is a challenging task. A few example images shown in Figure 1 showcase the alteration in the face features due to illicit drug abuse which in turn degrade the face recognition performance.

**Classification Networks:** Inspired by the literature (Agarwal et al., 2023; Ojaswee et al., 2023), we have used various pre-trained CNNs namely VGG16 (Simonyan & Zisserman, 2014), Xception (Chollet, 2017), DenseNet (Huang et al., 2017), and InceptionV3 (Szegedy et al., 2016) for illicit drug abuse face detection. A batch size of 32, an initial learning rate of 0.0001, and the Adam optimizer are used to train networks. The models are fine-tuned using the cross-entropy loss and trained with early stopping to avoid overfitting. These networks are fine-tuned on 180 clean and drug abuse images along with traditional data augmentation and the remaining 50 images are used for the evaluation.

## 3 Experimental Results and Analysis

The results reported in Table 1 showcase that the DenseNet yields state-of-the-art performance surpassing each network by a significant margin. For example, the DenseNet model yields an accuracy of **82**% in comparison to 70%, 68%, and 78% achieved by VGG, Xception, and Inception models, respectively. When we analyze the performance of individual networks on specific classes (Table 2), it is also observed that each classifier is effective in handling before drug abuse class (original/clean images) but is less effective in the after-drug abuse class except DenseNet.

Further, since face images are one of the most popular mediums of communication they have been extensively uploaded on social media platforms where enhancing the quality of face images is a popular practice. These filter-based enhancements drastically alter the spatial features (Figure 2) and impact the performance of classifiers and face recognition networks (Wu et al., 2020; Agarwal et al., 2021). Therefore, before assuring the effectiveness of DenseNet in detecting drug abuse faces, we have evaluated its effectiveness against social-media-based filter enhancement. *For that, we applied* 26 *popular Instagram filters to the test set of the proposed dataset and generated* 1300 *images in total*. The classifier trained on original images is used to evaluate these unseen enhanced images.

It is observed that the few filters such as 'Aden' and 'Hudson' which perform beautification do not leave any impact on the classification performance. However, a few filters such as 'Inkwell', 'Moon', and 'Xpro2' drastically reduce the detection performance of DenseNet (Table 3). However, interestingly, the filter does not impact the performance of the after-drug face class but affects the classification of the before-drug abuse class (Table 4). In other words, it increases false negative errors.

## 4 Conclusion

Despite the global impact of illicit drug abuse, limited datasets have hindered significant progress in this field. To advance research, we proposed a novel drug abuse face dataset and conducted a benchmark study using state-of-the-art convolutional neural networks (CNNs). *'Further, for the first time in the literature, we have studied the impact of Instagram filters on the detection of drug abuse faces'*. These enhancements can drastically impact our view on the social media profile of a person who might be addicted to a drug and can easily bypass the detectors through tap-based filter enhancements. We want to mention that the alteration in faces can be due to several other factors such as medical diseases, generative networks, or adversarial perturbation, we have to be vigilant before calling any image a drug abuse image. In the future, we aim to incorporate other factors that can degrade face recognition and how we can address them in a unified manner.

## URM Statement

The authors acknowledge that the key author of this work meets the URM criteria of the ICLR 2024 Tiny Papers Track.

## Ethics statement

While we acknowledge the image sources, specific details about the dataset origins are not provided to maintain privacy and ethical considerations. The links for image retrieval will be shared for research purposes only with the research community upon signing the license agreement. We encourage the users of the dataset to ethically use it for research purposes only by maintaining proper privacy and security concerns. Further, in the paper, the face images added are marked with strips to protect the privacy of individuals.

## References

Akshay Agarwal, Richa Singh, Mayank Vatsa, and Afzel Noore. Magnet: Detecting digital presentation attacks on face recognition. *Frontiers in Artificial Intelligence*, 4:643424, 2021.

Akshay Agarwal, Nalini Ratha, Richa Singh, and Mayank Vatsa. Robustness against gradient based attacks through cost effective network fine-tuning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 28–37, 2023.

Tadas Baltrušaitis, Peter Robinson, and Louis-Philippe Morency. Openface: an open source facial behavior analysis toolkit. In *IEEE Winter Conference on Applications of Computer Vision*, pp. 1–10, 2016.

François Chollet. Xception: Deep learning with depthwise separable convolutions. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1251–1258, 2017.

Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4690–4699, 2019.

Sixue Gong, Vishnu Naresh Boddeti, and Anil K Jain. On the intrinsic dimensionality of image representations. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3987–3996, 2019.

Mohamad Harastani, Amine Benterkia, Farnaz Majid Zadeh, and Amine Nait-Ali. Methamphetamine drug abuse and addiction: Effects on face asymmetry. *Computers in Biology and Medicine*, 116:103475, 2020.

Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4700–4708, 2017.

Yongjie Li, Xiangyu Yan, Bo Zhang, Zekun Wang, Hexuan Su, and Zhongwei Jia. A method for detecting and analyzing facial features of people with drug use disorders. *Diagnostics*, 11(9): 1562, 2021.

Ojaswee Ojaswee, Akshay Agarwal, and Nalini Ratha. Benchmarking image classifiers for physical out-of-distribution examples detection. In *IEEE/CVF International Conference on Computer Vision Workshops*, pp. 4427–4435, October 2023.

Wanli Ouyang, Xiaogang Wang, Xingyu Zeng, Shi Qiu, Ping Luo, Yonglong Tian, Hongsheng Li, Shuo Yang, Zhe Wang, Chen-Change Loy, et al. Deepid-net: Deformable deep convolutional neural networks for object detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2403–2412, 2015.

P Pandey, R Singh, and M Vatsa. Face recognition using scattering wavelet under illicit drug abuse variations. *Internation Conference on Biometrics*, 2016.

Table 2: Confusion matrix of drug abuse face classification using different CNNs. It shows that the DenseNet shows the effectiveness in detecting the after class along with the before (clean) face image class.

| Model→ | VGG16 | | Xception | | Inception | | DenseNet | |
|---|---|---|---|---|---|---|---|---|
| True → Predicted ↓ | Before | After | Before | After | Before | After | Before | After |
| Before | 20 | 10 | 18 | 9 | 20 | 6 | 20 | 4 |
| After | 5 | 15 | 7 | 16 | 5 | 19 | 5 | 21 |

Omkar Parkhi, Andrea Vedaldi, and Andrew Zisserman. Deep face recognition. In *British Machine Vision Conference*, 2015.

Ramachandra Raghavendra, Kiran B Raja, and Christoph Busch. Impact of drug abuse on face recognition systems: a preliminary study. In *International Conference on Security of Information and Networks*, pp. 24–27, 2016.

Arun Ross, Sudipta Banerjee, and Anurag Chowdhury. Deducing health cues from biometric data. *Computer Vision and Image Understanding*, 221:103438, 2022.

Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 815–823, 2015.

Sefik Ilkin Serengil and Alper Ozpinar. Lightface: A hybrid deep face recognition framework. In *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*, pp. 23–27. IEEE, 2020. doi: 10.1109/ASYU50717.2020.9259802. URL `https://doi.org/10.1109/ASYU50717.2020.9259802`.

Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2818–2826, 2016.

Zhe Wu, Zuxuan Wu, Bharat Singh, and Larry Davis. Recognizing instagram filtered images with feature de-stylization. In *AAAI Conference on Artificial Intelligence*, volume 34, pp. 12418–12425, 2020.

Daksha Yadav, Naman Kohli, Prateekshit Pandey, Richa Singh, Mayank Vatsa, and Afzel Noore. Effect of illicit drug abuse on face recognition. In *IEEE Winter Conference on Applications of Computer Vision*, pp. 1–7, 2016.

## A Proposed Drug Abuse Dataset

Drug Abuse Faces (DAF) dataset comprises two distinct classes: 'After,' representing post-drug abuse face images, and 'Before,' signifying pre-drug abuse face images. The before drug abuse faces can be seen as clean faces that do not have any impact on illicit drugs. However, once a person illegally consumes the illicit drugs in a large amount the impact of drugs seems visible on the facial features. In total 230 images of 115 subjects are meticulously compiled from various internet repositories.

Table 2 shows the performance of the different classifiers in handling individual (before and after) drug abuse classes. It can be seen that while each network shows similar performance on before drug abuse classes; the DenseNet model outperforms each network in detecting after drug abuse classes.

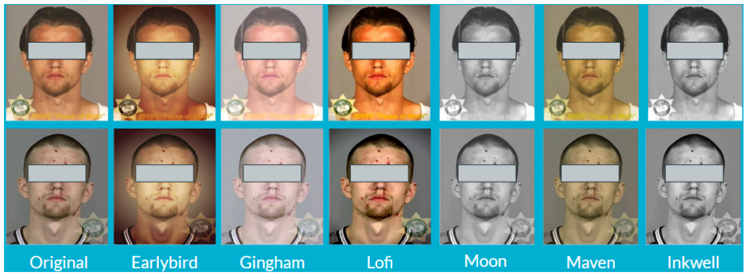


Figure 2: Showcasing the impact of Instagram filters on drug abuse faces.

Table 3: Drug abuse face classification performance of the DenseNet model when the test images went through filter-based image enhancements.

| Filter | Precision | Recall | F1 Score | Accuracy | AUC-ROC |
|---|---|---|---|---|---|
| None (Original) | 0.81 | 0.84 | 0.82 | 0.82 | 0.89 |
| _1977 | 0.87 | 0.80 | 0.83 | 0.84 | 0.88 |
| Aden | 0.79 | 0.88 | 0.83 | 0.82 | 0.89 |
| Brannan | 0.74 | 0.92 | 0.82 | 0.80 | 0.88 |
| Brooklyn | 0.80 | 0.80 | 0.80 | 0.80 | 0.90 |
| Clarendon | 0.80 | 0.80 | 0.80 | 0.80 | 0.87 |
| Earlybird | 0.85 | 0.68 | 0.76 | 0.78 | 0.83 |
| Gingham | 0.86 | 0.72 | 0.78 | 0.80 | 0.85 |
| Hudson | 0.81 | 0.84 | 0.82 | 0.82 | 0.87 |
| **Inkwell** | **0.69** | **0.88** | **0.77** | **0.74** | **0.85** |
| Kelvin | 0.81 | 0.84 | 0.82 | 0.82 | 0.89 |
| Lark | 0.84 | 0.84 | 0.84 | 0.84 | 0.90 |
| Lofi | 0.79 | 0.76 | 0.78 | 0.78 | 0.86 |
| maven | 0.74 | 0.92 | 0.82 | 0.80 | 0.89 |
| Mayfair | 0.83 | 0.80 | 0.82 | 0.82 | 0.87 |
| **Moon** | **0.71** | **0.88** | **0.79** | **0.76** | **0.84** |
| Nashville | 0.80 | 0.80 | 0.80 | 0.80 | 0.86 |
| Perpetua | 0.79 | 0.88 | 0.83 | 0.82 | 0.90 |
| Reyes | 0.84 | 0.84 | 0.84 | 0.84 | 0.89 |
| Slumber | 0.79 | 0.92 | 0.85 | 0.84 | 0.90 |
| Stinson | 0.81 | 0.84 | 0.82 | 0.82 | 0.89 |
| Toaster | 0.82 | 0.82 | 0.82 | 0.82 | 0.87 |
| Valencia | 0.84 | 0.84 | 0.84 | 0.84 | 0.90 |
| Walden | 0.80 | 0.80 | 0.80 | 0.80 | 0.86 |
| Willow | 0.75 | 0.88 | 0.79 | 0.78 | 0.84 |
| **Xpro2** | **0.71** | **0.80** | **0.75** | **0.74** | **0.81** |

## B ROBUSTNESS OF DRUG ABUSE FACE DETECTORS

It is observed when users upload their face images on various social media platforms, they apply various image processing techniques. Out of them, one of the popular image processing methods is filter-based enhancement. These tap-based enhancements are easy to perform and do not require any technical knowledge of the messaging platforms. We assert that these enhancements which drastically after the spatial properties can help in bypassing the drug abuse face detectors. Therefore, we have further investigated the impact of various Instagram filters on the drug abuse face classification model's performance. A few sample images of both before and after drug abuse classes that went through different filter-based enhancements are shown in Figure 2.

Table 3 in-depth analysis provides critical insights into the nuanced effects of specific Instagram filters on the robustness and accuracy of automated systems designed for drug abuse detection. Specifically, filters such as Xpro2, Inkwell, and Moon are observed to significantly reduce the classification model's accuracy to 74%. These filters induced a noticeable decline in the model's ability to identify drug abuse correctly faces up to 8%. Furthermore, filters Earlybird, Gingham, and Lofi tended to misclassify instances of drug abuse as non-drug abuse. Conversely, filters Xpro2, Moon, Maven, and Inkwell are found to detect non-drug abuse inaccurately as drug abuse. Notably, the misclassifications introduced by these filters underscore the model's susceptibility to distortions arising from image manipulations.

Table 4: Confusion Matrix showcasing the robustness of the best-performing model (DenseNet) in predicting drug abuse faces when the images are modified using the most effective Instagram filters.

| Filter → | Earlybird | | Gingham | | Lofi | | Moon | | Maven | | Inkwell | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Predicted → True ↓ | After | Before | After | Before | After | Before | After | Before | After | Before | After | Before |
| After | 17 | 8 | 18 | 7 | 19 | 6 | 22 | 3 | 23 | 2 | 22 | 3 |
| Before | 3 | 22 | 3 | 22 | 5 | 20 | 9 | 16 | 8 | 17 | 10 | 15 |


Figure 3: Sample drug abuse faces misclassified by the DenseNet.

The findings from Table 4 reveal notable patterns in the drug abuse face classification model's performance under specific Instagram filters. For instance, the Earlybird filter led to 17 instances predicted as "After drug abuse," with 8 misclassified as "Before drug abuse," and similar trends are observed with filters like Gingham, Lofi, Moon, Maven, and Inkwell. Figure 3 shows a few incorrectly classified samples of the DenseNet model.

## C Integration of Instagram Filters in Training Phase

Earlier, we showcased that the Instagram-filtered images affect the drug abuse face detection performance. However, one probable solution to improve the detection performance is to include the filtered images while training the detection network. For that, we have performed an experiment where we included the filtered images obtained from the three most effective filters namely Inkwell, Moon, and Xpro2 while training the detection network. However, the trained DenseNet model is evaluated on the clean and drug abuse faces obtained by unseen Instagram filters (not used during training). The prime reason is that it is difficult to know beforehand which filter is going to be used in the future and these social media filters drastically keep changing with time. Therefore, evaluating the resiliency of the classifier is important. It is observed that through the results reported in Table 5, incorporating Instagram filters as data augmentation during training slightly diminishes the accuracy of the best-performing model, i.e., DenseNet from 80% to 76%. While this introduces a trade-off between model robustness and overall accuracy, it underscores the importance of careful consideration when using filter-induced variations in training to ensure a balanced performance in drug abuse face classification tasks.

## D Comparative Analysis of Face Recognition Models on Pre and Post-Drug Abuse Face Images

In the literature, a few preliminary studies have been conducted to analyze the impact of illicit drugs on face recognition systems. For example, Raghavendra et al. (Raghavendra et al., 2016) have studied the impact of drug abuse on hand-crafted feature-based face recognition algorithms. The authors have used the Local Binary Patterns (LBP), and Binarised Statistical Image Features (BSIF) and the recognition has been performed using Sparse Representation Classifier (SRC). Similarly, Yadav et al. (Yadav et al., 2016) have performed a preliminary analysis on the effect of drug abuse on commercial and handcrafted feature-based face recognition algorithms. However, no study evaluated the sensitivity of deep face recognition networks.

To tackle this limitation, in this research, we have conducted a comprehensive study by evaluating the sensitivity of several pre-trained deep face recognition networks including VGG-Face (Parkhi et al., 2015), ArcFace (Deng et al., 2019), FaceNet (Schroff et al., 2015), OpenFace (Baltrušaitis et al., 2016), FaceNet512 (Gong et al., 2019), SFace, and DeepID (Ouyang et al., 2015) available with deepface library (Serengil & Ozpinar, 2020). Before engaging in drug abuse, the VGG-Face model demonstrated the highest clean face recognition accuracy of 91.30% surpassing each of the

Table 5: Effectiveness and robustness of drug abuse face detector with and without incorporation of Instagram filters during training.

| Instagram Training | Precision | Recall | F1 Score | Accuracy | AUC-ROC |
|---|---|---|---|---|---|
| Without | 0.78 | 0.82 | 0.80 | 0.80 | 0.86 |
| With | 0.77 | 0.75 | 0.76 | 0.76 | 0.82 |

Table 6: Comparative evaluation of face recognition accuracy (%) of deep models using pre and post-drug abuse images.

| Model | VGG-Face | ArcFace | FaceNet | OpenFace | FaceNet512 | SFace | DeepID |
|---|---|---|---|---|---|---|---|
| Before | 91.30 | 89.57 | 84.35 | 30.43 | 66.96 | 84.35 | 81.74 |
| After | 73.91 | 58.26 | 33.04 | 0 | 13.04 | 39.13 | 16.52 |

deep face models used. While the other state-of-the-art (SOTA) face recognition models such as ArcFace and FaceNet models yield clean face accuracies of $89.57\%$ and $84.35\%$, respectively. However, a discernible decline in accuracy is observed across these models when analyzing post-drug abuse images. For example, VGG-Face which yields the best clean image face recognition accuracy suffers a drop of $17.39\%$. Similarly, FaceNet suffers a drop from $84.35\%$ to $33.04\%$. This decline in recognition performance underscores the potential impact of drug abuse on facial recognition systems, emphasizing the challenges in maintaining precision in the face of physiological changes induced by drug consumption. The results of this experiment are reported in Table 6.