Face Attributes and Detection of Drug Addicts

Sudarsini Tekkam Gnanasekar, Svetlana Yanushkevich
Department of Electrical and Computer Engineering
University of Calgary
Biometric Technologies Laboratory
Calgary, Canada, syanshk@ucalgary.ca

***Abstract*—This paper addresses the problem of detecting the prolonged drug abuse marks using "soft" face biometrics. We propose a two-stage approach that integrates both the machine learning approach and the probabilistic reasoning. We identified"drug affect facial attributes" for attribute classification, and then performed probabilistic inference for detecting the drug addicts using the five face attributes. The experiments were performed using pre-trained convolutional neural networks, GoogleNet, ResNet50 and VGG16, for face attribute classification. PCA was applied on the extracted features for dimensionality reduction along with Fisher's linear discriminant feature selection method, and then a classification of attributes was performed using a linear SVM. The average accuracy of the five "drug affect facial attribute" detection reached, in particular, 90% using the ResNet50 model. We then applied this statistics to create a Bayesian network which represented the causal model for the final decision-making to classify the subjects as drug addicts. This approach reached the accuracy of 84%.**

I. Introduction

Illicit drug abuse drastically affects the appearance and behavioral patterns of humans. Face attributes, body and gait are the most common biometrics used for detecting such effects. For instance, long-term abuse of methamphetamine and heroin results in the presence of multiple blisters/acne, fat tissue muscle loss which explains "prematurely aged" of the drug addicts, poor skin tone and, in some cases, hair loss.

The goal of this research is to investigate the use of facial marks that we identify as "drug affect facial attributes" for analysis of primary signs and symptoms of 0.5 – 3 years of usage of substances such as methamphetamine and heroin abuse. These attributes include blisters/acne, tissue muscle loss, hair loss, poor skin tone and gender. We propose to combine the machine learning techniques with probabilistic reasoning to extend and improve the performance of previous approaches to the problem.

II. Previous works

There has been quite a few studies on drug abuse detection primarily using face processing and recognition algorithms. Ramachandra et al.[10] used various face recognition algorithms to analyze the facial changes of person before and after substance abuse, with a Genuine match rate (GMR) of 12.90% at False match rate (FMR) of 0.01%. Yadav et al. [16] proposed the dictionary learning based drug abuse detection which involved a filtering step to distinguish between the normal and drug faces thereby, reducing computation time with an accuracy of 88.81%. Pandey et al. [8] used AutoScat and scattering wavelet network-based feature extraction for recognizing faces under different orientations and scales. They achieved an accuracy of 59.05% for ScatNet, and 50.48% for AutoScat, respectively. All the aforementioned methods are based on the "before" and "after" images of subjects. This paper proposes a novel approach of using facial marks as mainstream information and Bayesian network for risk assessment.

Fig. 1. Images of subjects from Faces of Meth dataset showing drug affect facial attributes

III. Proposed Approach

In this study, we focus on the use of the facial marks to detect the effects of drug abuse, given certain types of drugs such as methamphetamine and heroin. The database such as "Faces of Meth" [1], as well as additional image sets collected online, will be used in this research. Examples of images of subjects from Faces of Meth are shown in Figure 1.

This research shows that using facial marks is a viable option for drug addiction detection, considering that no prior record of a person exists in the database. For this purpose, we established a two stage Illicit Drug Abuse Detection framework. The architecture of the proposed approach is shown in Figure 2. In the first stage, the "drug affect facial attributes" are predicted through machine learning based on primary signs and symptoms of substance abuse. We used deep learning neural networks such as GoogleNet [13], ResNet50 [3] and VGG 16 [12]. The next stage is to pass the facial attributes to the Bayesian network for risk assessment and inference.

IV. Bayesian network for risk assessment

A *Bayesian network* (BN) is a probabilistic graphical model that encodes a joint probability distribution over a set

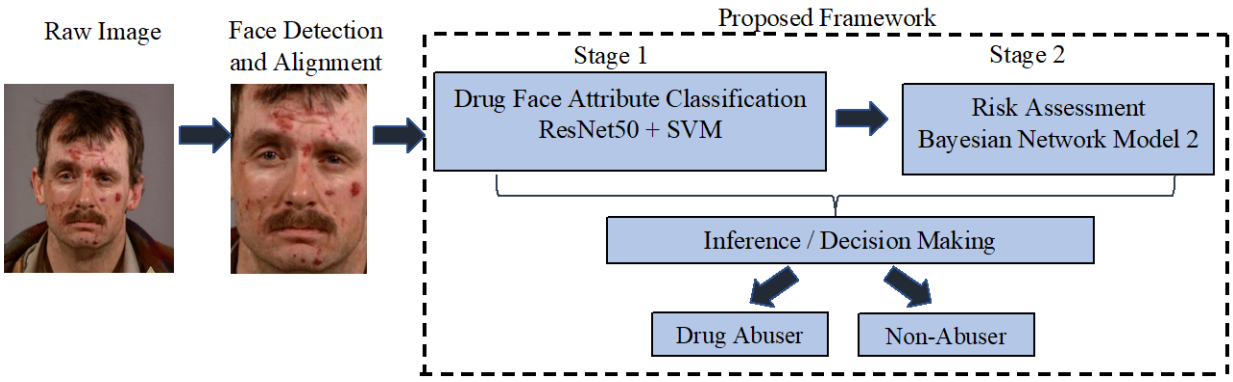

Fig. 2. Proposed Framework: The image of a subject is pre-processed using Multi-task Cascaded Convolutional Neural Networks (MTCNN) [17] which performs joint face detection and alignment. The processed image is passed into Stage 1 of the framework: Pre-trained CNN ResNet50 and SVM for "drug affect face attribute" classification, and the predicted face attributes are passed as evidence into Stage 2: Bayesian network for risk assessment. This framework classifies the subject as addict or non-addict

of random variables [9]. Each node in the BN represents a variable, and the arcs denote the relation between them. In addition, each node is augmented with a set of probabilities, referred to as conditional probability distribution.

The objective of this study is to infer the posterior probability of a person being a drug addict denoted in terms of risk level given drug affect facial attributes as symptoms. Let D_1, D_2 denote the "cause" for two possible events of drug abuse: "yes" or "no", then $P(D)$ is the prior probability of drug abuse for every possible event of cause D. With a set of "effects" E, obtained from face images, the posterior probability $P(D/E)$ can be estimated as follows:

$$P(D|E) = \frac{P(D,E)}{P(E)} \tag{1}$$

where $E = \{B, T, H, S, G\}$ are the effects of drug abuse and B, T, H, S, G are symptoms chosen in our study, such as blisters, tissue loss, hair loss, poor skin tone and gender. This model can also be referred to as the diagnostic modeling or bottom-up inference, where the cause is determined based on the risk level which is evaluated considering the effects of the symptoms.

V. Experiments

The experiments were performed on a i5 processor with 4 cores, 8GB of RAM and 4GB Nvidia graphics card along with CUDA 10.0 and cuDNN for GPU acceleration. The implementation was done in MatConvNet [15], and the pre-trained models from MatConvNet repository were used.

A. Data Preparation

The dataset used in this study was Faces-of-Meth consisting of 112 images with before and after drug abuse images of each subject. The Faces-of-Meth includes mug shots of individuals from the detention center of Multnomah County [1] which was originally created to promote awareness of the deleterious effects of drug abuse. We labeled each image with five attributes using binary labels. The images were pre-processed to perform joint face detection and alignment using MTCNN [17] in MatCaffe [5]. MTCNN is a face detector designed using deep learning techniques and works on both frontal and side-faced images. All the images used in this work were re-sized to 217×178 pixels.

B. Pre-trained Convolutional Neural Networks

GoogleNet, ResNet50 and VGG16 are the CNNs trained on the large-scale ImageNet dataset.

a) GoogleNet: *GoogleNet*, also referred to as the inception module, comprises of 22 layers and four million parameters respectively [13].

b) ResNet50: *ResNet50*, also known as the residual network, is 50 layers deep and was introduced to manage the degradation problem. It consists of residual connection which works on the principle of "skip connections" to improve gradient flow and training time [3].

c) VGG16: *VGG* is a very deep network designed by Visual Geometry group at Oxford. VGG 16 consists of 13 convolution layers and three fully connected layers. The default input size of 224x224x3 is used for smaller receptive field [12].

C. Experiment 1: Face Attribute Classification

In experiment 1, we performed the classification of the five selected "drug affect" facial attributes (Stage 1).

a) Feature Extraction and Classification of "drug affect" facial attributes.: Feature extraction is performed using the pre-trained CNNs, and linear Support Vector Machine (SVM) for classification. The features were extracted from the fully connected layer of each of the pre-trained CNN models. Features were extracted from $fc8-$VGG16, $pool5-$ResNet50 and $cls3\_pool-$ GoogleNet.

b) Principle component analysis.: Principle component analysis (PCA) was applied on the extracted drug abuse facial features for dimensionality reduction to project the original representations to a new feature space. It selects the components with maximum variance, specified by eigenvalue λ. The corresponding eigenvector c is represented as var$(X \times c)$=$c' \times S \times c$=$\lambda c' \times c$=λ, where c is a vector of

constants, S is the sample covariance matrix of the dataset and $'$ denotes transposition [6]. In our case, the features were reduced to 100 components.

c) Feature Selection.: Feature selection methods were employed to select the discriminative features which contained the important information without manipulating the original data using Feature Selection Library (FSLib) [11].

The features with reduced dimensions were further evaluated, and only the relevant features were selected using Fisher's Linear discriminant [7] and Lasso [14] feature selection methods. This improved the performance of the classifier to a great extent, 90%-ResNet50 as shown in Table I. For a parent set of N features, there are 2^N possible feature subsets which exponentially grows with N. This forms the basis of selecting the optimal subset of features in our experiments.

d) Fisher's Linear Discriminant.: In supervised learning and for two-class problem, the Fisher's discriminant finds a projection to maximize the separation between the classes. Let $A_1 = \{a_1^1, ..., a_{n_1}^1\}$ and $A_2 = \{a_1^2, ..., a_{n_2}^2\}$ be the data points of two classes, Fisher's discriminant [7] is represented by the vector **w** which maximizes the separation while minimizing the in-class variance

$$J(w) = \frac{w^T S_B w}{w_T S_W w}$$

and class mean

$$m_i = \frac{1}{n_i} \sum_{k=1}^{n_i} x_k^i$$

where $S_B = (m_2 - m_1)(m_2 - m_1)^T$ and $S_W = \sum_{i=1,2} \sum_{k=1}^{n_i} (x_k^i - m_i)(x_k^i - m_i)^T$ are the between- and within-class covariance matrices, respectively.

e) Lasso algorithm.: The Lasso feature selection eliminates the corresponding features during the learning process and is based on the l_1- norm of the coefficient w_k of **w**. Lasso feature generates an estimation of **w** with zero coefficients [14] using the notation of penalty:

$$penalty(w) = \sum_{k=1}^{n} |w_k|$$

f) Classification.: Classification was performed with linear SVM-one classifier per attribute, in order to leverage its generalization ability on a small dataset. SVM is a selective classifier that finds the best hyperplane which distinctly classifies the data points with the maximum margin between the classes. This approach also produces acceptable accuracy with less computational power.

Let (p_j, q_j), $j = 1, ..., n$ be a set of training samples and labels, where $p_j \in R^n$ and $q_j \in \{1, -1\}^n$, the SVM [2] is formulated by the following optimization problem [4]:

$$\min_{w,b,\xi} \ \frac{1}{2}[W^T W] + C \sum_{j=1}^{n} \xi_j$$

$$\text{subject to } \ q_j \left(w^T \Phi\left(p_j\right) + b\right) \geq 1 - \xi_j, \xi \geq 0$$

where p_j is the training vector which is transformed into a higher dimensional space by the function Φ. $K(p_j, p_k) \equiv \Phi(p_j)^T \Phi(p_k)$ is called the kernel function and $C > 0$ is the penalty parameter of the error term [4]. The kernel function was selected to be linear in this work as the data points are linearly separable.

g) Experiment 1: Discussion and Results: Tables I and II show the average accuracy and accuracy of each attribute for pre-trained CNN models with features extracted from the fully connected layer. PCA and feature selection methods such as Fisher's discriminant and Lasso were applied for various subsets of n features. Overall, ResNet50 performed best with an average accuracy of 90% based on the five drug affect face attributes using Fisher (with PCA) feature selection method for an optimal subset of 32 features.

TABLE I
AVERAGE ACCURACY OF EACH PRE-TRAINED CNN USING FISHER'S LINEAR DISCRIMINANT AND LASSO FEATURE SELECTION METHODS WITH AND WITHOUT PCA AND SVM AS THE CLASSIFIER. THE MAXIMUM ACCURACIES ARE SHOWN IN BOLD

Pre-trained CNN	Feature Selection Method	Average Accuracy(%)		
		n=16	n=32	n=64
GoogleNet	Fisher(without PCA)	77.50	78.75	74.46
GoogleNet	Fisher(with PCA)	80.36	82.50	82.86
GoogleNet	Lasso(without PCA)	69.46	80.54	82.32
GoogleNet	Lasso(with PCA)	61.79	64.29	57.86
ResNet50	Fisher(without PCA)	79.46	77.86	77.32
ResNet50	Fisher(with PCA)	**87.50**	**90.00**	**86.96**
ResNet50	Lasso(without PCA)	71.43	83.57	84.29
ResNet50	Lasso(with PCA)	74.46	69.82	60.71
VGG16	Fisher(without PCA)	79.64	76.07	76.79
VGG16	Fisher(with PCA)	83.75	84.46	82.86
VGG16	Lasso(without PCA)	76.96	83.21	83.21
VGG16	Lasso(with PCA)	63.39	59.64	41.43

TABLE II
DRUG AFFECT ATTRIBUTE ACCURACY OF RESNET50 MODEL USING FISHER (WITH PCA) FEATURE SELECTION METHOD

# of features (n)	Drug Affect Attribute Accuracy(%)				
	Blister	Muscle loss	Gender	Hair loss	Poor skin tone
16	87.50	84.82	95.54	85.71	83.93
32	86.61	88.39	93.75	91.96	89.29
64	83.04	84.82	94.64	81.25	91.07

D. Experiment 2: Bayesian Network Model for Decision Making and Risk Assessment

In experiment 2, we carried out Stage 2 of the classification of drug abuse cases, - decision making. It is performed using Bayesian networks, i.e., the posterior probability or risk level of a subject is evaluated based on the fusion of the drug affect face attributes and a decision is made, which classifies the subject as an abuser or a non-abuser. This includes passing the drug affect face attributes (present T (true) and absent F (false)) from stage 1 of the framework as evidence into the

Bayesian network for assessing posterior probability or risk level of a person. The inference engine used for performing analysis with the evidence and marginal nodes is junction tree algorithm.

Two types of Bayesian network models were designed.

The two models have the same number of nodes but differ in the way the probability values are set in the CPTs for each node.

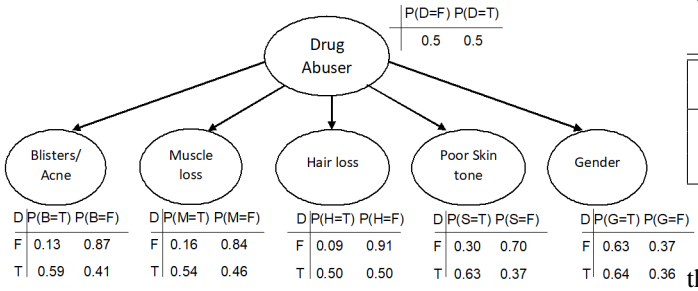

Fig. 3. Case 1: BN model 1 for Risk Assessment. The conditional probability tables (CPTs) for each attribute node are created based on the distribution of images in the dataset

a) Case 1: Bayesian network Model 1 for Decision Making: Figure 3 shows the Bayesian network model 1 with the conditional probability tables (CPTs) designed based on the distribution of images in the dataset in terms of positive and negative occurrences of a particular attribute. For instance, the total number of images (including drug and non-abuser classes, i.e., $D = T$ and $D = F$) that were positive for attribute blister was 40 out of 112 images. Similarly, for attributes tissue muscle loss and hair loss, the number of positive images were 39 and 33 out of 112 respectively. The attributes with the highest positive occurrences were gender (male) with 71 images and poor skin tone with 52 images out of 112. Since the dataset is imbalanced with uneven distribution of images, i.e., the number of positive images is less than the number of negative images for most of the attributes, we introduce a risk threshold (λ) for decision making. The experiments were performed for various values of λ as shown in Table III and the optimal threshold was determined to be 0.98. The risk level of a person is evaluated by comparing the calculated posterior probability of $P(D = T)$ to λ and is specified as

$$P(D/E(T)) < \lambda = RiskHigh(Abuser) \quad (2)$$

$$P(D/E(T)) > \lambda = RiskLow(Non\text{-}Abuser) \quad (3)$$

If the posterior probability of $P(D = T)$ is below the risk threshold (λ) then the person is classified as abuser or risk level is high, otherwise, the person is classified as non-abuser or risk is low as denoted in equations (3) and (4).

b) Case 1: Discussion and Results.: The Bayesian network model 1 was tested on unknown faces (not enrolled in gallery) of 44 drug abusers (collected from internet) and 56 non-abusers with the initial analysis of "drug affect facial attributes" from stage 1. Here, the classification of face attributes was performed using the best performing CNN model from stage 1, in this case, ResNet50. The classification accuracy for non-abuser, drug abuser and combined (drug and non-abuser) classes using ResNet50 and SVM for stage 1 and, Bayesian network model 1 for stage 2 are shown in Table III.

TABLE III
CLASSIFICATION ACCURACY OF ADDICT, NON-ADDICT AND COMBINED CLASSES: RESNET50 AND SVM FOR FACE ATTRIBUTE CLASSIFICATION, BN MODEL 1 FOR RISK ASSESSMENT.

Class	Classification Accuracy		
	$(\lambda = 0.95)$	$(\lambda = 0.98)$	$(\lambda = 0.99)$
Non-Addict	96.43%	87.50%	58.93%
Addict	43.18%	70.45%	81.82%
Combined (Addict & Non-Addict)	73.00%	80.00%	69.00%

These results imply that the Bayesian network model 1 with the CPTs created based on the number of images showed the ideal classification accuracy of 87.50% for non-abuser and 70.45% for drug abuser classes, respectively. This improved performance can be attributed to the appropriate selection of risk threshold (λ). A value of 0.98 was selected for λ such that there is even classification and mis-classification rates among the abuser and non-abuser classes with an average accuracy of 80% (both drug abuser and non-abuser classes combined).

c) Case 2: Bayesian network Model 2 for Decision Making.: Figure 4 shows the CPTs of each attribute node for BN model 2. In this research, we categorized the importance of each attribute and created the prior probabilities based on its impact on the final decision. In particular, the attributes such as blisters and tissue loss which are more prominent on faces of drug abusers are given greater weights, whereas attributes such as hair loss and poor skin tone are assigned lesser weights. Gender is given the lowest weight. It should be noted that the term weight in this context refers to the probability value assigned to each node.

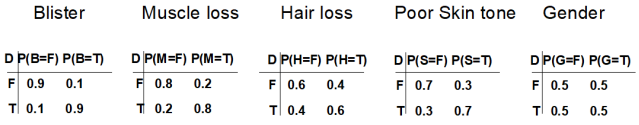

Fig. 4. Case 2: CPTs of each attribute node for BN model 2

A sample procedure for risk assessment is shown below. The posterior probability is estimated based on the fusion of the five "drug affect" facial attributes using Bayesian network model 2. Here, the probabilities of risk levels, high and low, are measured in terms of $P(D = T)$ and $P(D = F)$ and evaluated by passing the presence, true (T), or absence, false (F), of each attribute into the Bayesian network model 2. The final decision on the drug abuse detection is made by considering the highest probability among the two risk levels (high and low).

$$Given\ the\ Evidence: B = T, M = F, S = F, H = T, G = T(Male)$$
$$Evaluate: Bel(D = T)\ and\ Bel(D = F)$$
$$Solution:$$
$$Bel(D = T) = P(D = T/B = T, M = F, S = F, H = T, G = T)$$
$$Bel(D = T) = 0.591$$
$$Bel(D = F) = 1 - Bel(D = T) = 0.409$$
$$Final\ Decision: High\ Risk\ (Abuser)$$

The BN model 2 was used for Stage 2 (decision making) of the proposed framework.

d) Performance Evaluation of Proposed Framework: Stage 1 - Drug Affect Face Attribute Classification and Stage 2 - BN Model 2 for Decision Making.: The best performing pre-trained model from Experiment 1 (Case 1) is ResNet50 and, hence, feature extraction is performed using ResNet50 and classification with SVM.

The proposed framework (stage 1-ResNet50 and SVM; stage 2-BN model 2) was tested on 100 subjects which includes 56 subjects of non-abusers and 44 subjects of abusers (collected from internet) for performance evaluation. The confusion matrix with the number of correct and incorrect predictions of drug abuser and non-abuser classes is shown in Figure 5. Out of the 56 subjects of non-abusers, 50 subjects were correctly classified as non-abusers and 6 subjects were misclassified as abusers and of the 44 subjects of abusers, 34 were correctly classified as abusers and 10 were misclassified as non-abusers. The classification accuracy for drug abuser, non-abuser and combined (drug and non-abuser) classes are shown in Table IV. Sensitivity (True Positive Rate) was estimated as 77.27%, Specificity (True Negative Rate) was 89.29%, and Precision (Positive Predictive Value) was 85%.

TABLE IV
CLASSIFICATION ACCURACY OF DRUG ABUSER, NON-ABUSER AND COMBINED CLASSES: RESNET50 AND SVM FOR FACE ATTRIBUTE CLASSIFICATION, BN MODEL 2 FOR RISK ASSESSMENT

Class	Classification Accuracy
Non-Addict	89.29%
Addicy	77.27%
Combined (Addict and Non-Addict)	84.00%

Thus, the proposed framework shows better performance in predicting non drug abusers than in doing so for drug abusers, with an accuracy of 89.29%. Overall, the prediction accuracy of both drug abuser and non-abuser combined shows an average accuracy of 84%.

E. Experiment 3: Direct Approach

a) Case 1: Pre-trained CNN (ResNet50) for feature extraction and SVM for classification.: In the third experiment, no face attributes and Bayesian networks were used. Instead, we carried out drug abuse detection by training the classifier directly with "before" (non-abuser) and "after" (abuser) images of subjects. Here, the best performing pre-trained CNN from experiment 1 (Case 1) - ResNet50 ($pool5$) was used for feature extraction and SVM for classification. This approach was conducted to compare the performance of such "direct" approach with the proposed one that uses facial attributes and BN model 2 for risk inference. The classification accuracy for drug abuser and non-abuser classes are shown in Table VI.

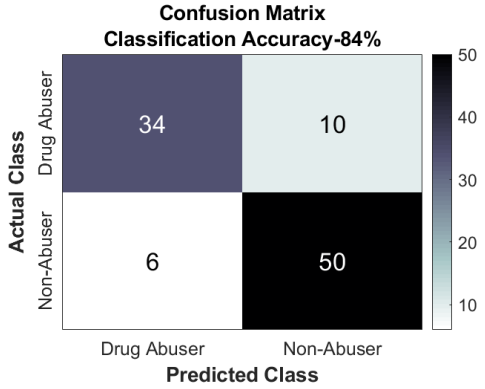

Fig. 5. Confusion Matrix for Drug Abuse Detection classifier

TABLE V
CLASSIFICATION ACCURACY OF DRUG ABUSER, NON-ABUSER AND COMBINED CLASSES: RESNET50 AND SVM FOR DRUG ABUSE CLASSIFICATION

Class	Classification Accuracy
Non-Abuser	70.54%
Drug Abuser	65.18%
Combined (Abuser and Non-Abuser)	67.86%

The results indicate that performance in case of the "direct" application of the pre-trained CNN is relatively weak compared to the results of the proposed approach (using facial attributes and the BN). Figure 6 shows the comparison of classification accuracy for drug abuser, non-abuser, both drug abuser and non-abuser combined using the proposed and the "direct" machine learning methods. The proposed approach (Stage 1: ResNet50 and SVM; Stage 2: BN model 2 for risk assessment) had better classification accuracy with an improvement of 18.75% for non-abuser class, 12.09% for drug abuser class and 16.14% for both drug abuser and non-abuser classes combined.

b) Case 2: End-to-End-Training: Pre-trained CNN (ResNet50) for feature extraction and classification.: Another experiment performed in the "direct" approach methodology is by using CNN for both feature extraction and classification, also referred to as end-to-end-training. Here, the best performing pre-trained CNN, ResNet50, was used to perform feature extraction and classification. This is carried out by training the network on "before" (non-abuser) and "after" (abuser) images of subjects for 30 epochs with training and validation split of 70% and 30% respectively. The learning rate and weight decay were set to $1 \times e^{-2}$ and $5 \times e^{-4}$ respectively. To avoid over-fitting of data, we performed data augmentation which includes jittering of images with random flipping and cropping.

The SoftMax function was used for classification. The results indicate that the network performs best on the training data as the training accuracy reaches 100% by epoch 16 but fails poorly in generalization of new data, in spite of the aggressive weight decay and data augmentation techniques performed. The best validation accuracy of 65.60% was obtained at epoch 29.

This study confirms that the proposed approach of using face attributes for initial analysis and Bayesian networks for decision making is robust, and supersedes the "direct" methods in terms of accuracy and performance.

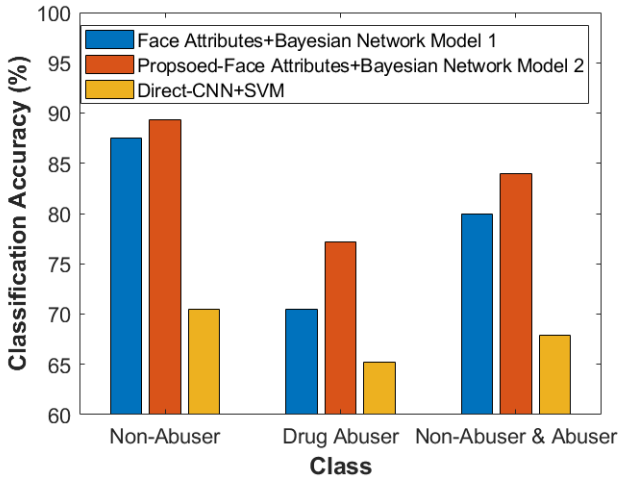

Fig. 6. Performance comparison using Proposed and Direct Methods

VI. Conclusion and Future work

In this research work, we investigated the use of face attributes as indicators of illicit drug addict detection, and how it can be applied for further probabilistic inference of overall risk of a person being an addict.

The overall contribution of this work is as follows: (1) We proposed to apply the deep neural networks pre-trained on large databases with the features extracted from the fully-connected layers of each network; we have selected five attributes that are the most prominent indicators of the methamphetamine and heroin affect for which the data for training was available. The feature extraction and classification of the five selected face attributes for the proposed approach is performed using ResNet50 and SVM. (2) We laid out the approach that integrates both the machine learning and the probabilistic machine reasoning to perform risk assessment given a subject image. We achieved an average accuracy of 84% using the proposed framework (Stage 1: ResNet50 and SVM; Stage 2: BN model 2 for inference) (3) For drug abuse classification using "direct" methodology, we experimented with two approaches: a combination of the CNN for feature extraction and SVM for classification, and the CNN applied for both tasks.

We further plan to: (1) Enhance the performance of attribute detection which can be achieved by increasing the size of the datasets for training the machine learning tools. This is difficult to achieve due to inaccessibility and legal restriction on collection of such databases. An alternative approach is to generate synthetic images, and possibly apply Generative Adversarial Network (GAN) approach to do so. (2) Conduct experiments on the full version of the proposed machine reasoning model, by using the computationally efficient tool for Bayesian inference which we are currently developing. (3) Implement a multi-layered approach to risk assessment of a subject being an addict; this approach includes not only analysis of face attributes, but also behavioral biometrics analysis of body attributes, gait and gestures.

Acknowledgment

This project was partially supported by the University of Calgary "Human Dynamics Research Strategy" VPR funds, and the Natural Science and Engineering Research Council Discovery Grant "'Biometric Contactless Interfaces".

References

[1] Faces of Meth. http://www.mcso.us/facesofmeth/ [Online; accessed August 17 2018].

[2] B. E. Boser, I. M. Guyon, and V. N. Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152. ACM, 1992.

[3] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[4] C.-W. Hsu, C.-C. Chang, C.-J. Lin, et al. A practical guide to support vector classification. 2003.

[5] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*, 2014.

[6] I. T. Jolliffe and J. Cadima. Principal component analysis: a review and recent developments. *Phil. Trans. R. Soc. A*, 374(2065):20150202, 2016.

[7] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K.-R. Mullers. Fisher discriminant analysis with kernels. In *Neural networks for signal processing IX: Proceedings of the 1999 IEEE signal processing society workshop (cat. no. 98th8468)*, pages 41–48. Ieee, 1999.

[8] P. Pandey, R. Singh, and M. Vatsa. Face recognition using scattering wavelet under illicit drug abuse variations. In *ICB*, pages 1–6, 2016.

[9] J. Pearl. Probabilistic reasoning in intelligent systems: Networks of plausible reasoning, 1988.

[10] R. Raghavendra, K. B. Raja, and C. Busch. Impact of drug abuse on face recognition systems: a preliminary study. In *Proceedings of the 9th International Conference on Security of Information and Networks*, pages 24–27. ACM, 2016.

[11] G. Roffo, S. Melzi, U. Castellani, and A. Vinciarelli. Infinite latent feature selection: A probabilistic latent graph-based ranking approach. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1398–1406, 2017.

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

[13] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[14] J. Tang, S. Alelyani, and H. Liu. Feature selection for classification: A review. *Data classification: algorithms and applications*, page 37, 2014.

[15] A. Vedaldi and K. Lenc. Matconvnet: Convolutional neural networks for matlab. In *Proceedings of the 23rd ACM international conference on Multimedia*, pages 689–692. ACM, 2015.

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

[17] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10):1499–1503, 2016.