

Gender and Ethnicity Classification of Iris Images using Deep Class-Encoder

Maneet Singh¹, Shruti Nagpal¹, Mayank Vatsa^{1,2}, Richa Singh^{1,2}, Afzel Noore², and Angshul Majumdar¹
¹IIT-Delhi, India, ²West Virginia University

{maneets, shrutin, mayank, rsingh, angshul}@iiitd.ac.in, afzel.noore@mail.wvu.edu

Abstract

Soft biometric modalities have shown their utility in different applications including reducing the search space significantly. This leads to improved recognition performance, reduced computation time, and faster processing of test samples. Some common soft biometric modalities are ethnicity, gender, age, hair color, iris color, presence of facial hair or moles, and markers. This research focuses on performing ethnicity and gender classification on iris images. We present a novel supervised autoencoder based approach, Deep Class-Encoder, which uses class labels to learn discriminative representation for the given sample by mapping the learned feature vector to its label. The proposed model is evaluated on two datasets each for ethnicity and gender classification. The results obtained using the proposed Deep Class-Encoder demonstrate its effectiveness in comparison to existing approaches and state-of-the-art methods.

1. Introduction

Soft biometric traits are physical or behavioral characteristics that can be extracted from the human body which help differentiate individuals from one another. However, they often lack information that is permanent or sufficiently discriminative to uniquely identify an individual. These traits are valuable as they aid in recognition and enhance the performance of automated biometric systems by reducing the search space. This results in a significant decrease of testing time as well. Some of the most commonly used soft biometric modalities are gender, ethnicity, gait, age, height, hair color, eye color, and presence of moles or mustache on the face [4, 16]. Figure 1 presents few sample soft biometric modalities that can be extracted from face and iris.

Iris texture has emerged as one of the most robust and widely used biometric modality for automated identity recognition [15]. Studies in literature have also established the presence of similar information in iris patterns to discriminate between ethnicities and gender [4, 10, 11, 21, 22]. That is, samples belonging to the same ethnicity, have some

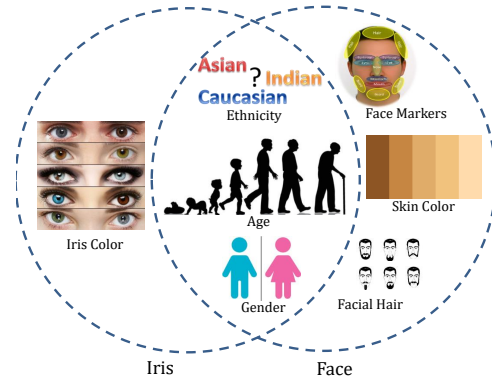


Figure 1: Sample soft biometric modalities that can be extracted from face and iris images. All images have been taken from the Internet.

commonalities in the iris pattern, and the ones belonging to the same gender also share certain information.

In literature, gender from iris images was first predicted by Thomas *et al.* [23] using an in-house dataset of left irises only. The iris images were segmented and normalized, onto which log-Gabor filter was applied to obtain the iris texture information. Features like mean, standard deviation, and local variation were computed for different neighborhoods of the texture pattern. Seven additional geometric features were also computed, followed by projection onto a random subspace ensemble of trees for classification. Lagree and Bowyer [11] extended this work by computing texture features for different horizontal and vertical bands of the normalized iris image. A total of 630 features were computed for 14 regions in a given image. These features were used as input to the Sequential Minimum Optimization algorithm for classification. The same extracted features were also used to perform ethnicity classification of iris images. Tapia *et al.* [21] used uniform LBP with concatenated histograms to perform gender classification on iris images on an in-house UND dataset. However, in this work, no mutual exclusivity of subjects was maintained between the training and testing partitions. Building upon this, in 2016 [22], the disjoint train-test sets in terms of the subject were created, and gender classification was performed using the

same binary texture pattern that is used for identity recognition. This work presented results on three datasets, the UND dataset used in [21], a novel ND-Gender From Iris (ND-GFI) dataset, and a subject disjoint validation set (UND_V).

Qiu *et al.* [18] first proposed the idea of classifying iris images into ethnicity based categories. The authors combined images from three datasets and performed classification between Asian and Non-Asian by computing the Gabor energy ratio between two regions of the iris. Continuing this work, a method to represent iris images using iris textons was developed [19]. The authors performed classification between Asian and Non-Asian on the CASIA BioSecure dataset. A filter bank of 40 Gabor filters was created and a 40-dimensional feature vector was computed. Each iris image was represented with a histogram of iris textons, which was used as a feature for SVMs to perform classification. The authors reported an accuracy of 91.02% to classify Asians v/s non-Asians. However, it is essential to note that these results were demonstrated on images collected from different datasets for different ethnicities, thus incorporating inter-database variations at the time of classification. Lagree and Bowyer [11] proposed to identify the ethnicity using features obtained from several bands for 14 subregions of the iris. In 2012, Zarei and Mou [25] used artificial neural networks to predict the ethnicity. The authors computed a total of 882 features from various regions of the iris, which were combined to create a feature vector used as input to a neural network for classification.

This research presents ethnicity and gender classification from iris images, using the proposed *Deep Class-Encoder*. It is an autoencoder based supervised model which utilizes the robust feature extraction capabilities of deep learning and discriminative capabilities of supervised models. The research contributions of this work are two-fold:

- An autoencoder based supervised model, *Deep Class-Encoder* is presented. The proposed model incorporates supervision by learning weights such that the hidden representation is mapped onto its class label. The proposed model is optimized using Alternating Direction Method of Multipliers [3], instead of a gradient descent approach, thus reducing the training time substantially.
- For gender classification, results are reported on existing protocols of ND-Iris-0405 [17] and ND-Gender From Iris [22] datasets. Ethnicity classification is performed on ND-Iris-0405 and the proposed Multi-Ethnicity dataset. A new protocol is also defined on the Multi-Ethnicity dataset for ethnicity classification. The proposed model achieves state-of-the-art results for each task.

2. Proposed Deep Class-Encoder

The success of unsupervised feature learning architectures has been well established over the past couple of years,

especially with the advent of representation learning techniques such as Deep Learning [2]. The performance of these architectures is not restricted to a particular domain, and spans across areas such as speech recognition, biometric recognition, computer vision, and natural language processing [13]. Unsupervised architectures aim to harness the information provided from unlabeled data to learn representative features used for classification at a later stage.

Autoencoders are deep learning architectures used to learn representations in an unsupervised manner. A traditional autoencoder consists of two sub-systems, encoder and decoder. For a given input $\mathbf{X}$, the encoder is used to obtain the representation $\mathbf{H}$ using the learned encoder weights $\mathbf{W}_e$. The decoder is used for reconstructing the input from the learned representation using the decoder weights, $\mathbf{W}_d$. The loss function of an autoencoder is optimized by minimizing the difference between the given input data, $\mathbf{X}$, and the reconstructed data, $\hat{\mathbf{X}}$. It is mathematically expressed as:

$$\min \|\mathbf{X} - \hat{\mathbf{X}}\|_F^2 \quad (1)$$

where, the input consists of n training samples, the hidden representation is computed as $\mathbf{H} = \phi(\mathbf{W}_e \mathbf{X})$, and the reconstructed data $\hat{\mathbf{X}} = \mathbf{W}_d \mathbf{H}$. ϕ is an activation function which can be any non-linear function such as *sigmoid*, or *tanh* or linear activation corresponding to unit activation. Upon expanding Equation 1, with unit activation function, the loss function can be expressed as:

$$\arg \min_{\mathbf{W}_d, \mathbf{W}_e} \|\mathbf{X} - \mathbf{W}_d \mathbf{W}_e \mathbf{X}\|_F^2 \quad (2)$$

A traditional biometrics pipeline consists of feature extraction, followed by classification. The feature extraction module extracts meaningful representations from the input, while the classifier learns a boundary in order to distinguish between the representations. Autoencoders have been used as unsupervised models for feature extraction in several applications. In order to learn meaningful representations for specific tasks, researchers have proposed incorporating supervision in the autoencoder model. Most of the existing supervised models do not explicitly encode the class label, but encode only the class information (same/different) for feature learning [6, 14, 26]. Encoding class information often leads to reducing the intra-class variability.

In this research, we incorporate supervision in the feature extraction module in an attempt to introduce discriminability at the time of feature extraction itself. It is our hypothesis that feature extraction from such a model would facilitate better classification. We propose a model which utilizes the robust feature extraction capability of autoencoders, while learning discriminative features to enhance classification. The proposed architecture, termed as *Class-Encoder* is built such that the class label of a given training sample is encoded in the learned feature representation, thereby making

the model supervised in nature. This is done by incorporating an additional mapping matrix $\mathbf{M}$, which maps the hidden representation $\mathbf{W}_e \mathbf{X}$ onto its corresponding class label $\mathbf{C}$. Mathematically, it can be expressed as:

$$\min_{\mathbf{M}} \|\mathbf{C} - \mathbf{M} \mathbf{W}_e \mathbf{X}\|_F^2 \quad (3)$$

where, $\mathbf{C} \in R^{l \times n}$ (l being the number of distinct classes) is a binary vector with its i^{th} element set to 1 if the data sample belongs to class i , rest being zero. Matrix $\mathbf{M}$ captures the linear mapping between the feature vector ($\mathbf{W}_e \mathbf{X}$) and its corresponding class label. Thus, under a fixed $\mathbf{M}$ and $\mathbf{W}_e$, samples belonging to the same class should map to the same class label. Extending Equation 2 to incorporate the above term, the loss function of the proposed Class-Encoder can thus be given as:

$$\min_{\mathbf{M}, \mathbf{W}_d, \mathbf{W}_e} \|\mathbf{X} - \mathbf{W}_d \mathbf{W}_e \mathbf{X}\|_F^2 + \lambda \|\mathbf{C} - \mathbf{M} \mathbf{W}_e \mathbf{X}\|_F^2 \quad (4)$$

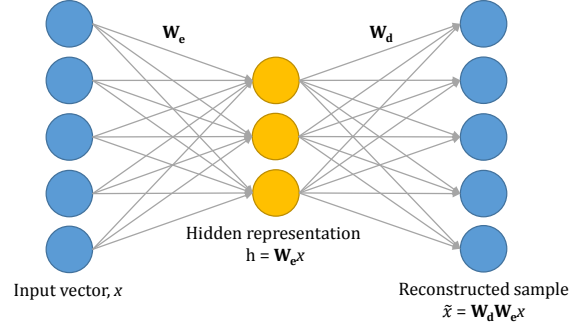
where, the regularization parameter, λ , controls the relative contribution of the two terms. In Equation 4, the first term corresponds to the loss function of the traditional autoencoder. It facilitates learning of parameters such that the reconstruction error is minimized and a meaningful representation is learned. The second term incorporates supervision in the autoencoder formulation and enables the model to learn a discriminative representation of the input data. The weight matrix is learned such that upon projecting the input data on it, the representation maps to a specific class label, thus encoding class-specific information in the feature learning process. Figure 2 gives an overview of the proposed model, and its comparison with the traditional unsupervised autoencoder.

Equation 4 represents a least square formulation, however, the large size of matrices involved and the non-convexity of the problem increases the computational costs. To mitigate the same, the Majorization-Minimization (MM) technique [5] along with Alternating Direction Method of Multipliers (ADMM) [3] is utilized. The aim of the MM technique is to replace complex equations with simpler and easier optimization steps. Moreover, ADMM does not involve computing the derivatives at each epoch, which leads to significant lower training time as compared to a traditional Stacked Autoencoder.

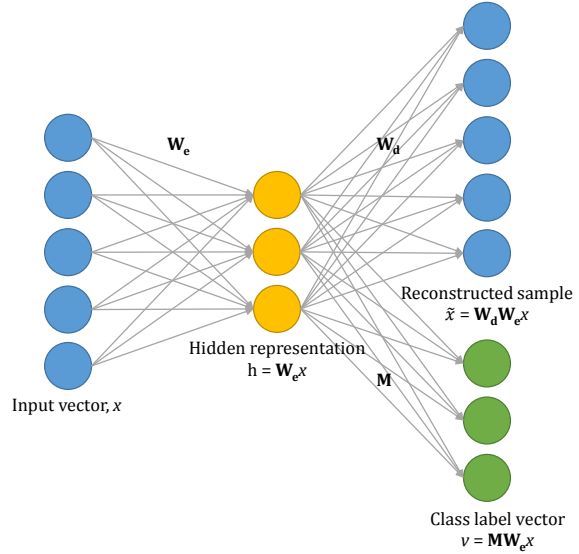
For a k layer model of Deep Class-Encoder, Equation 4 can be extended as follows:

$$\min_{\mathbf{M}, \mathbf{W}_d, \mathbf{W}_e} \|\mathbf{X} - (\mathbf{W}_d^1 \mathbf{W}_d^2 \dots \mathbf{W}_d^k (\mathbf{W}_e^k \dots \mathbf{W}_e^2 \mathbf{W}_e^1 \mathbf{X}))\|_F^2 + \sum_{i=1}^k \lambda \|\mathbf{C} - \mathbf{M}^i (\mathbf{W}_e^i \mathbf{X}^i)\|_F^2 \quad (5)$$

where, $\mathbf{W}_e^k$, $\mathbf{W}_d^k$, and $\mathbf{M}^k$ correspond to the encoding weights, decoding weights and the mapping matrix of the



(a) Traditional Unsupervised Autoencoder



(b) Proposed Class-Encoder

Figure 2: $\mathbf{W}_e$ and $\mathbf{W}_d$ correspond to the encoding and decoding weights respectively, while $\mathbf{M}$ captures the linear mapping between the feature vector and its class label.

k^{th} layer, respectively. $\mathbf{X}^k$ refers to the input to the k^{th} layer of the model, which is defined as $\mathbf{W}_e^{k-1} \mathbf{X}^{k-1}$. At the first layer, $\mathbf{X}^1 = \mathbf{X}$, that is, the input data. Thus, for a k -layer model, the proposed Class-Encoder aims to learn discriminative representations at each layer by modeling the class label information during the feature learning process.

3. Ethnicity and Gender Classification using Deep Class-Encoder

Deep Class-Encoder is used for the task of gender and ethnicity classification of iris images. For both the problems, experiments are performed on two datasets each. Details regarding the algorithm, datasets used, and experimental protocols are given below.

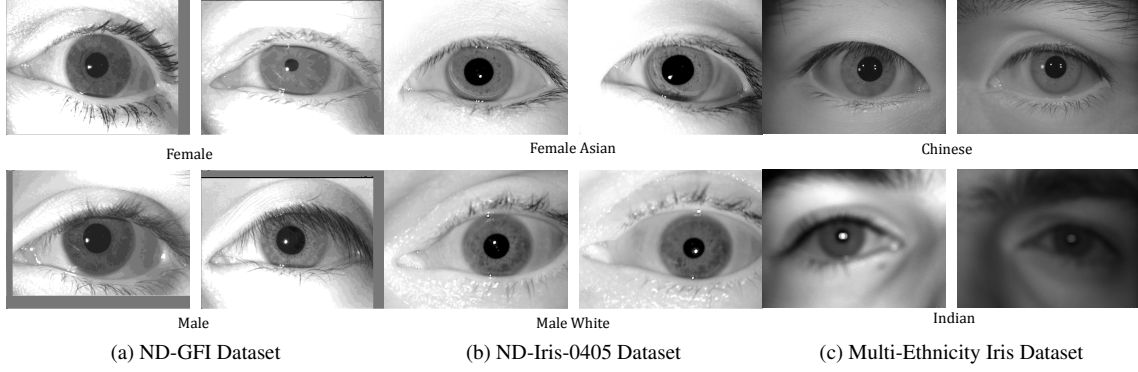


Figure 3: Sample images from the three datasets used for performing gender and ethnicity classification on iris images.

3.1. Ethnicity Classification

The proposed *Deep Class-Encoder* is used to perform ethnicity classification using Equation 5. Here, $\mathbf{X}$ is a $m \times n$ matrix which contains the input images, where m is the size of the input image and n is the number of input samples. In this case, iris images captured in the NIR spectrum are given as input, resized to 48×64 , therefore m being equal to 3,072. $\mathbf{W}_d$, $\mathbf{W}_e$ are the weight parameters learned at the time of training. $\mathbf{C}$ is the class label matrix of dimension $l \times n$, where l is the number of distinct classes. A two layer Deep Class-Encoder of dimension [3072, 768, 768] is learned. In order to evaluate the efficacy of the proposed model, results and comparison have been shown with two classifiers: Random Decision Forest (RDF) and a two layer Neural Network (NNet) having dimension [768, 192, 48].

3.1.1 Datasets and Experimental Protocol

Two datasets, ND-Iris-0405 [17] and a combined Multi-Ethnicity Iris Dataset, have been used in this research.

ND-Iris-0405 dataset consists of 64,980 images corresponding to 356 subjects, out of which 158 subjects are Caucasian (White), 82 are Asian, and the remaining 24 belong to other ethnicities. Since the number of samples belonging to *other* ethnicity are very limited, images corresponding to the Caucasian and Asian subjects have been chosen for evaluation, thereby resulting in a two-class problem (Caucasian or Asian), consisting of 60,259 images. Out of these, 26,272 images (containing equal number of Caucasian and Asian images) are used for training, while the remaining images form the test set.

Multi-Ethnicity Iris dataset is created by combining three existing datasets, due to the lack of publicly available iris datasets containing images of multiple ethnicities. The dataset consists of images from:

- CASIA-Iris V3 dataset [1] collected by the Chinese Academy of Sciences' Institute of Automation (CASIA): The dataset consists of three parts, CASIA-Iris-Interval,

Table 1: Experimental protocols for ethnicity and gender classification.

Dataset	No. of Training Images	No. of Testing Images
Ethnicity Classification		
ND-Iris-0405 [17]	26,272	33,987
Multi-Ethnicity Iris	1,302	59,008
Gender Classification		
ND-Iris-0405 [17]	42,899	22,081
ND-GFI [22]	2,399	600
UND_V [22]	-	1,944

CASIA-Iris-Lamp, and CASIA-Iris-Twin. Out of these, CASIA-Iris Lamp and CASIA-Iris Twin contain 18,197 images of only Chinese ethnicity, that have been used.

- IMP dataset [20]: This dataset contains images collected in the visible and NIR spectrum, and night-time mode. Out of these, 630 images of Indian ethnicity captured in the NIR spectrum are used.
- ND-Iris-0405 dataset [17]: As mentioned previously, this dataset contains images of subjects belonging to Asian, Caucasian, or other ethnicities. Since Asian might correspond to Chinese or Indian as well, and further categorization has not been provided in the dataset, we only use 41,518 images pertaining to Caucasian subjects in the Multi-Ethnicity Iris dataset.

Therefore, the Multi-Ethnicity Iris dataset consists of 60,310 images of Chinese, Indian, and Caucasian ethnicities. 1,302 images are chosen for creating the training partition, which consists of equal number of images from all three classes (Chinese, Indian, and Caucasian), and the remaining images are used to create the test set.

3.2. Gender Classification

Deep Class-Encoder is also used for performing gender classification on iris images. Similar to the previous experiments, the images are down-sampled to 48×64 . Thus, $\mathbf{X}$ is

Table 2: Ethnicity classification accuracy (%) on the ND-Iris-0405 dataset with two classifiers: Random Decision Forest (RDF) and Neural Network (NNet).

Algorithm	RDF	NNet
Stacked Autoencoder [8]	87.24	80.05
Stacked Denoising Autoencoder [24]	85.20	64.51
Deep Belief Network [7]	85.12	87.43
Discriminative RBM [12]	90.33	
Proposed Deep Class-Encoder	89.35	94.33

of dimension $3072 \times n$, where n is the number of input images. $\mathbf{C}$ is a $2 \times n$ binary-matrix containing the class labels. The architecture of the feature extractors and classifiers is kept consistent with that described in Section 3.1.

3.2.1 Datasets and Experimental Protocol

Experiments have been performed on two datasets: ND-Iris-0405 [17] and ND-Gender From Iris (GFI) [22].

ND-Iris-0405 dataset consists of 64,980 images corresponding to 158 females and 198 males. 70% data of each class is used for training (resulting in 42,899 images), while the remaining form the test set.

ND-Gender From Iris (ND-GFI) dataset consists of 3,000 images from 750 males and 750 females. The dataset consists of a pre-defined protocol [22], where 80% of the data (per class) is used for training, and the remaining 20% is used for testing. A separate subject-disjoint set (UND_V) containing 1,944 images has also been provided for evaluating the performance of the trained model. The same protocol is followed for the evaluation of the proposed model. All protocols ensure mutually exclusive training and testing sets, such that there is no image which occurs in both the partitions. Figure 3 presents sample images from all three datasets, and Table 1 summarizes the protocols followed for the experimental evaluation.

4. Experimental Results and Analysis

The proposed model has been compared with existing Deep Learning models for performing gender and ethnicity classification on iris images. Comparison has been performed with Stacked Autoencoder (SAE), Stacked Denoising Autoencoder (SDAE) [24], Deep Belief Network (DBN) [7], Discriminative Restricted Boltzmann Machine (DRBM) [12], and AlexNet [9], i.e. a Convolutional Neural Network based model. To be consistent with the proposed model, feature extraction models (SAE, SDAE, DBN) have the same architecture as that of the proposed Deep Class-Encoder (described in Section 3.1). This is followed by learning a classifier (RDF or NNet) for classification. Owing to the class imbalance in the test samples, mean class-wise accuracy has been reported throughout this paper.

Table 3: Ethnicity classification accuracy (%) on the Multi-Ethnicity Iris dataset with two classifiers: RDF and NNet.

Algorithm	RDF	NNet
Stacked Autoencoder [8]	95.76	90.42
Stacked Denoising Autoencoder [24]	95.13	87.3
Deep Belief Network [7]	96.97	95.22
Discriminative RBM [12]	95.92	
Proposed Deep Class-Encoder	97.22	97.38

Table 4: Confusion matrix of Deep Class-Encoder on the ND-Iris-0405 dataset for ethnicity classification.

		Predicted	
Actual		Asian	Caucasian
	Asian	93.78%	6.22%
	Caucasian	5.12%	94.88%

Since no existing Commercial-Off-The-Shelf system predicts the attributes of gender and ethnicity for iris images, a comparison could not have been drawn for the same. Task specific analysis for the proposed Deep Class-Encoder model is given below.

4.1. Ethnicity Classification

Tables 2 and 3 present the accuracy for ethnicity classification on ND-Iris-0405 and Multi-Ethnicity Iris dataset, respectively. Figure 4 also presents the Receiver Operative Characteristic (ROC) curve obtained on the ND-Iris-0405 dataset for ethnicity classification of Asians versus Caucasians. On the ND-Iris-0405 dataset, the proposed model achieves a classification accuracy of 89.35% and 94.35% with RDF and NNet classifier, respectively (Table 2). Compared with other unsupervised feature learning models (SAE, SDAE, and DBM), Deep Class-Encoder presents an improvement of at least 7%. Direct comparison with SAE (difference of around 14% in accuracy) shows the benefit of encoding class labels in the proposed model. Comparison with Discriminative RBM, which is a supervised Deep Learning architecture, also presents an improvement of 4%. The key difference between Discriminative RBM and proposed Deep Class-Encoder is that the former models the joint probability of the sample and the label, while the later tries to learn representations such that under a fixed learned mapping, the hidden representations map to the class labels. It can thus be observed that the introduction of the mapping matrix facilitates learning of discriminative features. Table 4 gives the confusion matrix obtained with the proposed Deep Class-Encoder and Neural Network. The low mis-classification rates for both the classes suggest that the proposed model is able to encode meaningful and discriminative representations across classes.

Similar results can be observed from Table 3 for eth-

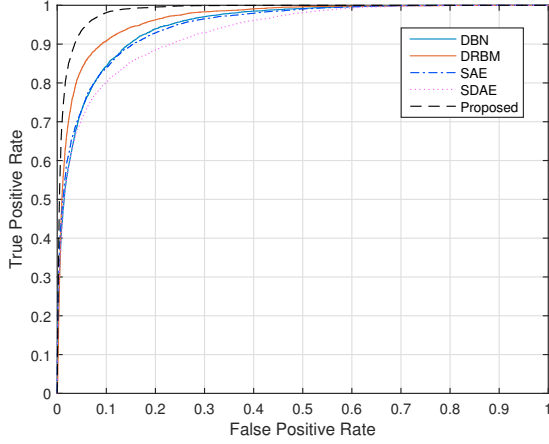


Figure 4: Ethnicity classification on the ND-Iris-0405 dataset. For all models, the best performance between RDF and NNet is plotted.

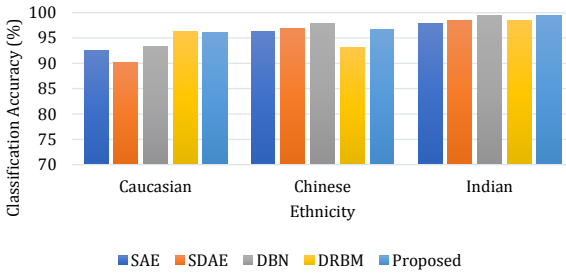


Figure 5: Classification performance (%) of all algorithms on all three classes for the Multi-Ethnicity Iris dataset.

nicity classification on Multi-Ethnicity Iris dataset. In this case, the classifier aims to classify the images as either Caucasian, Chinese, or Indian. The proposed model achieves a classification accuracy of 97.38% (with Neural Network), which shows improvement over other comparative feature extraction models. Moreover, the variation in classification performance obtained with RDF and NNet for Deep Class-Encoder is within 1%, whereas the difference is as high as 8% for other comparative models (SDAE). This observation further instantiates that the proposed model learns robust features which are independent of the classifier used in the pipeline. Figure 5 gives the bar graph with classification accuracies of all models for all three classes. The proposed model achieves at least 96% classification accuracy for all three classes, thus promoting the feature learning process of Deep Class-Encoder. Interestingly, the model mis-classifies only *one* sample of Indian ethnicity. As shown in Figure 6, overall illumination and size variations along with the presence of artifacts such as hair bangs render samples challenging for ethnicity classification.

It is important to note that no direct comparison could be drawn with existing work on ethnicity classification, since there is no fixed protocol or datasets which are used for

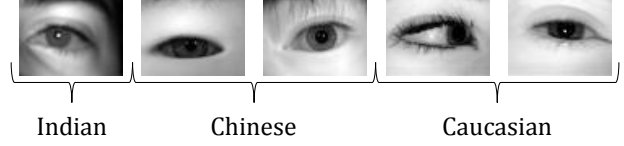


Figure 6: Sample images mis-classified by the proposed Deep Class-Encoder and Neural Network model for the Multi-Ethnicity Iris dataset. Artifacts such as hair (bangs) and incomplete iris information result in mis-classification. Table 5: Gender classification accuracy (%) on the ND-Iris 0405 dataset with two classifiers: RDF and NNet.

Algorithm	RDF	NNet
Stacked Autoencoder [8]	74.73	66.71
Stacked Denoising Autoencoder [24]	69.23	56.07
Deep Belief Network [7]	75.74	76.03
Discriminative RBM [12]	70.76	
Proposed Deep Class-Encoder	80.06	82.53

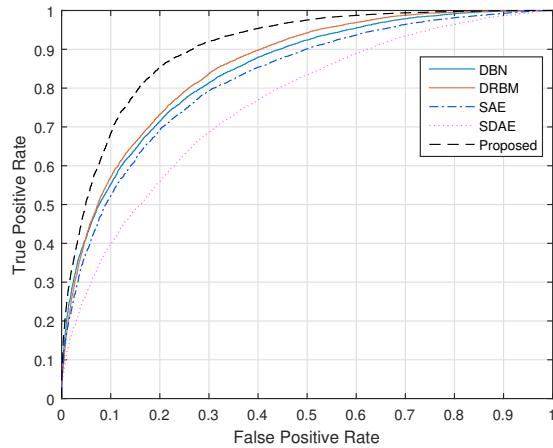
evaluation. While there exists works utilizing the ND-Iris-0405 dataset, however, experiments are performed on a very small subset of the dataset [10].

4.2. Gender Classification

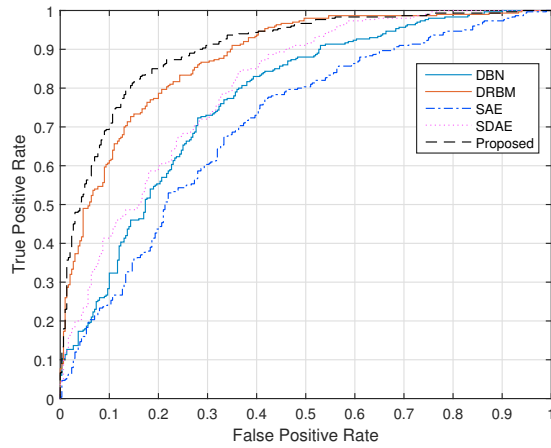
Tables 5 and 6 present the gender classification accuracy on ND-Iris-0405, and ND-GFI dataset, respectively. Figure 7 also presents the ROC curves obtained on both the datasets. On the ND-Iris-0405 dataset, the proposed Deep Class-Encoder achieves a classification accuracy of 82.53% (with Neural Network), which is at least 5.5% better than other comparative models. Figure 9 presents sample images mis-classified by the proposed model. It can be observed that iris images having poor illumination or partial information serve as challenging samples.

Experiments are also performed on the ND-GFI dataset, where, the Deep Class-Encoder achieved an accuracy of 83.17%, showing at least 5% improvement over other deep learning based comparative models. In literature, Tapai *et al.* [22] achieve an accuracy of 84.83% on the same protocol, which means that the proposed model performs 10 extra mis-classifications. The proposed Deep Class-Encoder achieves a classification accuracy of 79.25% on the UND_V dataset. This dataset is provided with the ND-GFI dataset and models the real world scenario of disjoint subjects in the training and testing set. The protocol mentioned by the authors [22] has been followed for evaluation, using which they report a classification accuracy of 77.5% (state-of-the-art for this dataset). This results in a difference of 34 correctly classified samples, thereby promoting the generalization abilities of Deep Class-Encoder.

Figure 8 presents a bar graph with the classification accuracies of a CNN based feature extractor, AlexNet [9], Tapia *et al.* [22], and the proposed model. Comparison with Con-



(a) ND-Iris-0405 Dataset



(b) ND-GFI Dataset

Figure 7: ROC curves for gender classification. For each model, the best performing classifier (RDF or NNet) is plotted.

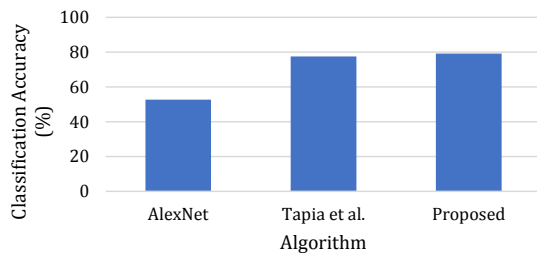


Figure 8: Bar graph presenting the gender classification accuracy obtained on the UND_V dataset [22].

Table 6: Gender classification accuracy (%) on the ND-GFI dataset with two classifiers: RDF and NNet.

Algorithm	RDF	NNet
Stacked Autoencoder [8]	64.17	65.33
Stacked Denoising Autoencoder [24]	71.33	66.00
Deep Belief Network [7]	67.17	71.83
Discriminative RBM [12]	78.67	
Proposed Deep Class-Encoder	78.17	83.17

volutional Neural Network based feature extractor, AlexNet [9] on the UND_V dataset further promotes the utility of the proposed model. Upon using the pre-trained model for feature extraction only, the proposed model achieves at least 16% improved performance. A major advantage of the proposed Deep Class-Encoder is in terms of the training time. As compared to Stacked Autoencoders, Deep Class-Encoder takes one fourth the total training time. This computational advantage is primarily due to the Majorization-Minimization and Alternating Direction Method of Multipliers based optimization which does not involve computing derivatives at each epoch.

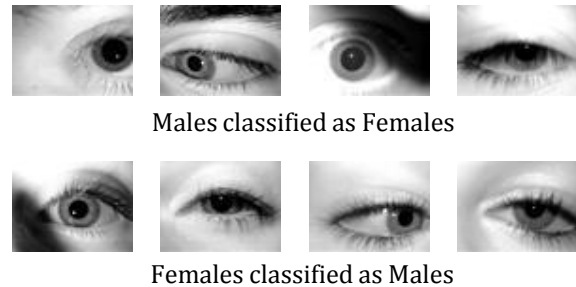


Figure 9: Sample mis-classifications by the proposed model. It can be observed that incomplete, poorly illuminated, and hidden iris tend to be more difficult to classify.

5. Conclusion

This research addresses the task of gender and ethnicity classification of iris images. A supervised autoencoder model, termed as Deep Class-Encoder has been proposed for the given task. Deep Class-Encoder utilizes the class labels at the time of feature learning, in order to learn discriminative features. The efficacy of the proposed model is evaluated on two datasets each, for gender and ethnicity classification. Experimental evaluation and results further promote the utility of the proposed model for learning class-specific discriminative features.

6. Acknowledgment

This research is partially supported by MEITY (Government of India), India. S. Nagpal is partially supported through TCS PhD fellowship. M. Vatsa, R. Singh, and A. Majumdar are partially supported through Infosys Center for Artificial Intelligence. We also thank NVIDIA Corp. for Tesla K40 GPU for research.

References

- [1] CASIA iris image database. <http://biometrics.idealtest.org/>.
- [2] Y. Bengio, A. Courville, and P. Vincent. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828, 2013.
- [3] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011.
- [4] A. Dantcheva, P. Elia, and A. Ross. What else does your biometric data reveal? a survey on soft biometrics. *IEEE Transactions on Information Forensics and Security*, 11(3):441–467, 2016.
- [5] M. Figueiredo, J. Bioucas-Dias, and R. Nowak. Majorization-minimization algorithms for wavelet-based image restoration. *IEEE Transactions on Image Processing*, 16(12):2980–2991, 2007.
- [6] S. Gao, Y. Zhang, K. Jia, J. Lu, and Y. Zhang. Single sample face recognition via learning deep supervised autoencoders. *IEEE Transactions on Information Forensics and Security*, 10(10):2108–2118, 2015.
- [7] G. E. Hinton, S. Osindero, and Y.-W. Teh. A fast learning algorithm for deep belief nets. *Neural computation*, 18(7):1527–1554, 2006.
- [8] G. E. Hinton and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 2006.
- [9] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105, 2012.
- [10] S. Lagree and K. W. Bowyer. Ethnicity prediction based on iris texture features. In *Midwest Artificial Intelligence and Cognitive Science*, pages 225–230, 2011.
- [11] S. Lagree and K. W. Bowyer. Predicting ethnicity and gender from iris texture. In *IEEE International Conference on Technologies for Homeland Security*, pages 440–445, 2011.
- [12] H. Larochelle and Y. Bengio. Classification using discriminative restricted boltzmann machines. In *International Conference on Machine Learning*, pages 536–543, 2008.
- [13] Y. Lecun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 5 2015.
- [14] A. Majumdar, R. Singh, and M. Vatsa. Face verification via class sparsity based supervised encoding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6):1273–1280, 2017.
- [15] I. Nigam, M. Vatsa, and R. Singh. Ocular biometrics: A survey of modalities and fusion approaches. *Information Fusion*, 26:1–35, 2015.
- [16] U. Park and A. K. Jain. Face matching and retrieval using soft biometrics. *IEEE Transactions on Information Forensics and Security*, 5(3):406–415, 2010.
- [17] P. J. Phillips, W. T. Scruggs, A. J. O’Toole, P. J. Flynn, K. W. Bowyer, C. L. Schott, and M. Sharpe. FRVT 2006 and ICE 2006 large-scale experimental results. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(5):831–846, 2010.
- [18] X. Qiu, Z. Sun, and T. Tan. Global texture analysis of iris images for ethnic classification. In *International Conference on Biometrics*, pages 411–418, 2005.
- [19] X. Qiu, Z. Sun, and T. Tan. Learning appearance primitives of iris images for ethnic classification. In *IEEE International Conference on Image Processing*, volume 2, pages II – 405–II – 408, 2007.
- [20] A. Sharma, S. Verma, M. Vatsa, and R. Singh. On cross spectral periocular recognition. In *IEEE International Conference on Image Processing*, pages 5007–5011, 2014.
- [21] J. E. Tapia, C. A. Perez, and K. W. Bowyer. Gender Classification from Iris Images Using Fusion of Uniform Local Binary Patterns, pages 751–763. 2015.
- [22] J. E. Tapia, C. A. Perez, and K. W. Bowyer. Gender classification from the same iris code used for recognition. *IEEE Transactions on Information Forensics and Security*, 11(8):1760–1770, 2016.
- [23] V. Thomas, N. V. Chawla, K. W. Bowyer, and P. J. Flynn. Learning to predict gender from iris images. In *IEEE International Conference on Biometrics: Theory, Applications, and Systems*, pages 1–5, 2007.
- [24] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research*, 11:3371–3408, 2010.
- [25] A. Zarei and D. Mou. Artificial neural network for prediction of ethnicity based on iris texture. In *International Conference on Machine Learning and Applications*, volume 1, pages 514–519, 2012.
- [26] X. Zheng, Z. Wu, H. Meng, and L. Cai. Contrastive auto-encoder for phoneme recognition. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2529–2533, 2014.