REPUBLIC OF TURKEY
YILDIZ TECHNICAL UNIVERSITY
DEPARTMENT OF COMPUTER ENGINEERING

IDENTITY AND BEAUTY ANALYSIS

14011052 — Rahmi Cemre ÜNAL

SENIOR PROJECT

Advisor
Assist. Prof. Dr. Hamza Osman İLHAN

January, 2021

ACKNOWLEDGEMENTS

I would like to thank for his devotion and all the helps to Assist. Prof. Dr. Hamza Osman İLHAN, who followed the study closely and guided me when it was needed.

Rahmi Cemre ÜNAL

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

CNN	Convolutional Neural Network
HOG	Histogram of Oriented Gradients

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Identity and Beauty Analysis

Rahmi Cemre ÜNAL

Department of Computer Engineering
Senior Project

Advisor: Assist. Prof. Dr. Hamza Osman İLHAN

This project consists of 4 main modules which are age, gender, race prediction and beauty analysis. Within the scope of the project, 4 different analysis of the 4 main modules are realized by determining the face of the person whose photograph is given. The gender and race of the person whose face is determined from the given photo are estimated with the help of trained CNN-based models. Afterwards, the person's photogenicity is predicted by selecting the appropriate one from 12 different model combinations trained according to gender and race.

The main idea adopted in this study is that each race has a different perception of beauty according to their ethnic origins. Instead of creating a general beauty perception model, it is thought that training a separate beauty model for each race can give results closer to the real beauty perception of people from different cultures. At the same time, within the scope of this study, a data set consisting of images of people who participated in beauty contests belonging to 6 different races as Asian, Indian, Latino-hispanic, Middle Eastern, Black and White were created.As a result of the study, 3 age, gender and race prediction models and models that can predict beauty with high accuracy in 12 different combinations of 2 different genders, male and female, for 6 different ethnic groups were obtained.

Keywords: Convolutional neural networks, face detection, age prediction, gender prediction, race prediction, beauty analysis

ÖZET

Kişilik ve Güzellik Analizi

Rahmi Cemre ÜNAL

Bilgisayar Mühendisliği Bölümü
Bitirme Projesi

Danışman: Dr. Ögr. Üyesi Hamza Osman İLHAN

Bu proje, yaş, cinsiyet, ırk tahmini ve güzellik analizi olarak 4 ana modülden oluşmaktadır. Proje kapsamında, fotoğrafı verilen kişinin yüzü tespit edilerek belirtilen 4 ana modüle ait 4 farklı analiz yapılmaktadır. Verilen fotoğraftan yüzü tespit edilen kişinin, cinsiyeti ve ırkı eğitilen CNN tabanlı modeller yardımıyla tahmin edilmektedir. Daha sonrasında cinsiyet ve ırka göre eğitilmiş 12 farklı model kombinasyonundan uygun olan seçilip kişinin fotojeniklik tahmini yapılmaktadır. Bu çalışmada benimsenen ana fikir, etnik kökenlere göre her bir ırkın farklı bir güzellik algısı olduğudur. Genel bir güzellik algısı modeli oluşturmak yerine, her ırk için ayrı bir güzellik modeli eğitmenin, farklı kültürdeki kişilerin gerçek güzellik algısına göre daha doğru sonuçlar verebileceği düşünülmüştür. Aynı zamanda bu çalışma kapsamında, Asyalı, Hindistanlı, Latin-ispanyol, Orta Doğulu, Siyahi ve Beyaz olacak şekilde 6 farklı ırka ait güzellik yarışmalarına katılmış kişilerin görüntülerden oluşan bir veri seti oluşturuldu.Çalışma sonucunda, 3 adet yaş, cinsiyet ve ırk tahmini modeli ve 6 farklı etnik grup ve erkek ve kadın olarak 2 farklı cinsiyete ait 12 farklı kombinasyonda yüksek doğruluk oranıyla güzellik tahmini yapabilen modeller elde edildi.

Anahtar Kelimeler: Konvolüsyonel yapay sinir ağları, yaş tahmini, cinsiyet tahmini, etnik köken tahmini, güzellik analizi

1
Introduction

1.1 Subject of the Study

The beauty and identity analysis project is a project that tries to predict the gender, age, ethnic origin, photogenicity characteristics of the person in the given image.

1.2 Motivation of the Study

Photogenicity is actually a relative concept related to people's perception of beauty. In this project, a separate beauty model for each ethnic group will be trained in data sets consisting of beauty contest participants, grouped according to ethnic origins.

In this way, it is expected that a more accurate result will be obtained since the photogenicity of the people will be determined according to the beauty perceptions of their own society.

1.3 Scope of the Study

In the project, it is aimed to analyze the age, gender, ethnic origin and beauty of the person whose image is given using image processing and image classification methods. Face detection will be required within the scope of this application.

Dlib HOG, which contains facial landmark information, is the appropriate face capture method for the project, since it is predicted that aligning the eyes to the same level will increase the accuracy rate in the case of capturing images taken at oblique angle after the faces are detected.

After the faces in the given images are correctly identified and aligned, age, gender and ethnicity information will be estimated with the help of models to be trained on the Convolutional Neural Network. Then, the relative beauty level of the person whose ethnic origin is determined will be predicted.

2
Preliminary Research

In this section, previous studies regarding the targeted parts of the project has been examined.

2.1 Face Detection

With the marvelous increase in video and image database there is an incredible need of automatic understanding and examination of information by the intelligent systems as manually it is getting to be plainly distant[1]. There are several ways to detect face on the images and we will look into some of them.

1. **Haar Cascade Classifiers**, a new human face detection algorithm by primitive Haar cascade algorithm combined with three additional weak classifiers is proposed in that paper [2]. The three weak classifiers are based on skin hue histogram matching, eyes detection and mouth detection. First, images of people are processed by a primitive Haar cascade classifier, nearly without wrong human face rejection (very low rate of false negative) but with some wrong acceptance (false positive). Secondly, to get rid of these wrongly accepted non-human faces, a weak classifier based on face skin hue histogram matching is applied and a majority of non-human faces are removed. Next, another weak classifier based on eyes detection is appended and some residual non-human faces are determined and rejected. Finally, a mouth detection operation is utilized to the remaining non-human faces and the false positive rate is further decreased. With the help of OpenCV, test results on images of people under different occlusions and illuminations and some degree of orientations and rotations, in both training set and test set show that the proposed algorithm is effective and achieves state-of-the-art performance. Furthermore, it is efficient because of its easiness and simplicity of implementation.

2. **Histogram of oriented gradients** is a feature descriptor used in computer vision and image processing for the purpose of object detection. The technique counts occurrences of gradient orientation in localized portions of an image. This method is similar to that of edge orientation histograms, scale-invariant feature transform descriptors, and shape contexts, but differs in that it is computed on a dense grid of uniformly spaced cells and uses overlapping local contrast normalization for improved accuracy.

   The success of the HOG SVM human detection algorithm lies in its discriminative HOG features and margin-based linear SVM classifier. The HOG SVM algorithm concentrates on the contrast of silhouette contours against the background. Different humans may have different appearances of wears but their contours are similar. Therefore the contours are discriminative for distinguishing humans from non-humans[3].

2.2 Age Prediction

The absolute first association did with age grouping from facial pictures in 1999 by Y. H. Kwon et, al [4].

In a study conducted in 2017[5], a novel Convolutional Neural Network (CNN)-based framework, ranking-CNN proposed for age estimation.Ranking-CNN contains a series of basic CNNs, each of which is trained with ordinal age labels. Then, their binary outputs are aggregated for the final age prediction. They consider the age estimation problem in the range between 16 and 66 years old and compare ranking-CNN with other state-of-the-art feature extractors and age estimators.The results show that Ranking-CNN can reach the accuracy of 89.90% for L (age error tolerance) = 6, and 92.93% for L = 7

In 2015, another research study published for apperent age prediction[6]. Their proposed method, Deep EXpectation (DEX) of apparent age, first detects the face in the test image and then extracts the CNN predictions from an ensemble of 20 networks on the cropped face. The CNNs of DEX were finetuned on the crawled images and then on the provided images with apparent age annotations. Their convolutional neural networks (CNNs) use the VGG-16 architecture and are pretrained on ImageNet for image classification. As a result of the study, standard mean absolute error was measured as 3.221.

2.3 Gender Prediction

K. Jhang and J. Cho published a study about training the CNN model for gender and age group prediction with camera in 2019 [7]. It appears that CNN for camera-based age and gender prediction is usually trained with RGB color images. However, it is difficult to say that CNN trained with RGB color images always produces good results in an environment where testing is performed with camera rather than with image files. They observe that in camera-based testing CNN trained with grayscale images shows better gender and age group prediction accuracy than CNN trained with RGB color images. Their best gender models reached 90.4% accuracy for file based RGB images and 89.7% accuracy for gray scale images.

K. Jhang also published another study on gender prediction based on voting of CNN models [8]. He proposed voting schemes to utilize the already developed CNN models to further improve gender prediction accuracy.In his study, he showed that using softmax-based models can improve the accuracy rather than using the single or majority voters models. The results of the models shown in 2.1

Table 2.1 Results of voters with cnn models

models	accuracy	majority voters	accuracy	softmax voters	accuracy
densenet	93.93	v3_012	95.00	sv_01	94.64
res50	93.97	v3_013	94.66	sv_02	95.10
vgg19	94.27	v3_023	94.55	sv_0-2	95.24
vgg16	93.16	v3_123	94.67	sv_0-3	95.19
res152	94.08	v5	95.23	sv_0-4	95.44

2.4 Race Prediction

Automatic verification and identification of face from facial image to obtain good accuracy with huge dataset of training and testing to using face attributes from images is still challengeable. Hence proposing efficient and accurate facial image identification and classification based of facial attributes is important task. The prediction from human face image is much complex. The proposed research [9] work for automatic gender, age and race classification is based on facial features and Convolutional Neural Network (CNN). The proposed study uses the physical appearance of human face to predict age, gender and race. The proposed methodology consists of three sub systems, Gender, Ageing and Race. The proposed race identification methods are used to extract standardized information from face pictures: LPB and WLD [10] in this [9] research. The study is based on 4 races as Mongolian, Indian, Black and White. The model accuracy was 99.2% for race prediction.

2.5 Beauty Prediction

Since the golden rules in the era of Leonardo Da Vinci, decoding the human perception of facial beauty has been a significant research. The perception of facial beauty for a human is involved with the attributes of facial appearance, which provides some significant visual cues for facial beauty prediction. In a study conducted in 2015[11], researchers propose a deep learning method to address the challenging facial attractiveness prediction problem. The method constructs a convolutional neural network for facial beauty prediction using a new deep cascaded fine tuning scheme with various face inputting channels, such as the original RGB face image, the detail layer image, and the lighting layer image.we have achieved a high prediction correlation of 0.88. This result convinces us that the problem of the facial attractiveness prediction can be solved by deep learning approach, and it also shows the important roles of the facial smoothness, lightness, and color information that involve in facial beauty evaluation, which is consistent with the result of recent psychology studies.

3 Feasibility

Feasibility for the "Identity and Beauty Analysis" project were conducted under four subjects as technique, time, legal and economics.

3.1 Technical Feasibility

It is necessary to determine the hardware and software features required for the realization of the identity and beauty analysis project.

3.1.1 Software Feasibility

Since the environment where the project will be implemented is the desktop, application environment will be the Windows platform.

The model training part of the project will be done using Convolutional Neural Network structure in python language and Tensorflow [12], Keras [13], OpenCV [14] libraries.

In the image processing part, Opencv, Dlib libraries will be used.

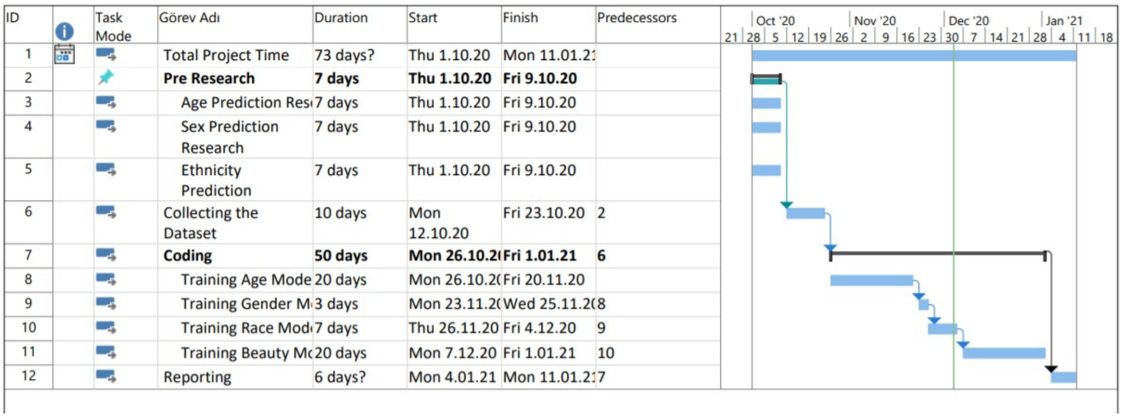
Windows 10 operating system, JetBrains Pycharm and Google Colabs will be used for the environment in which the project will be developed.

3.1.2 Hardware Feasibility

The environment in which the project will be carried out must meet the system requirements and software feasibility. Hardware features of the device generally used within the scope of the project: Intel i7 3.70 GHz processor speed, 32 GB DDR4 RAM, GTX1080 graphics card.

3.2 Time Feasibility

The path followed for time planning is shown in the 3.1.



ID	Task Mode	Görev Adı	Duration	Start	Finish	Predecessors
1		Total Project Time	73 days?	Thu 1.10.20	Mon 11.01.2	
2		**Pre Research**	**7 days**	**Thu 1.10.20**	**Fri 9.10.20**	
3		Age Prediction Res	7 days	Thu 1.10.20	Fri 9.10.20	
4		Sex Prediction Research	7 days	Thu 1.10.20	Fri 9.10.20	
5		Ethnicity Prediction	7 days	Thu 1.10.20	Fri 9.10.20	
6		Collecting the Dataset	10 days	Mon 12.10.20	Fri 23.10.20	2
7		**Coding**	**50 days**	**Mon 26.10.2**	**Fri 1.01.21**	**6**
8		Training Age Mode	20 days	Mon 26.10.2	Fri 20.11.20	
9		Training Gender M	3 days	Mon 23.11.2	Wed 25.11.2	8
10		Training Race Mod	7 days	Thu 26.11.20	Fri 4.12.20	9
11		Training Beauty M	20 days	Mon 7.12.20	Fri 1.01.21	10
12		Reporting	6 days?	Mon 4.01.21	Mon 11.01.2	7

Figure 3.1 Gantt Diagram

3.3 Legal Feasibility

The work to be done within the scope of the "Identity and Beauty analysis" project has not been prevented or restricted by law. Since the technologies planned to be used are free software environments, there is no need to obtain permission from any institution. Since the product obtained within the scope of the project is developed for research purposes and it is not commercial, it does not cause any violation of rights [15].

3.4 Economic Feasibility

The software planned to be used for the "identity and beauty analysis" project can be obtained free of charge.

In order to develop the project, it is sufficient for a software developer to work with a computer with sufficient hardware for the prescribed time.

The cost that will occur is shown in the 3.1.

Table 3.1 The cost of the project

	Cost
Computer	5.000TL
Developer	10.000TL
Total	**15.000TL**

4
System Analysis

The "identity and beauty analysis" project is planned to be created under 5 main modules. These modules are:

1. Face detection
2. Age prediction
3. Gender prediction
4. Race prediction
5. Beauty prediction

In this section, information is given about requirements and performance criteria as well as the modules.

4.1 System Modules

4.1.1 Face Detection

It has been determined that the HOG method can be used to detect the faces required for the later stages of the project.

Since it is predicted that aligning the eyes to the same level will increase the accuracy rate against the possibility of capturing the images that have been turned sideways after the faces are detected, it is thought that Dlib HOG, which contains the facial landmark information, is the appropriate method.

4.1.2 Age Prediction

After identifying and aligning the faces in the images given, it was thought that age estimation could be made with the help of a CNN-based model to be trained.

As can be seen from the research examined, accurate age prediction from the given image is a challenging task.

Since obtaining the estimate with softmax does not seem to be very accurate, estimating the 101 age group between 0 and 100 requires training a model with 101 output class.

Later, instead of choosing the one with the highest probability of these output classes, it was found to be more accurate to estimate the total result by multiplying the probabilities of each class with its own label as shown in 4.1.

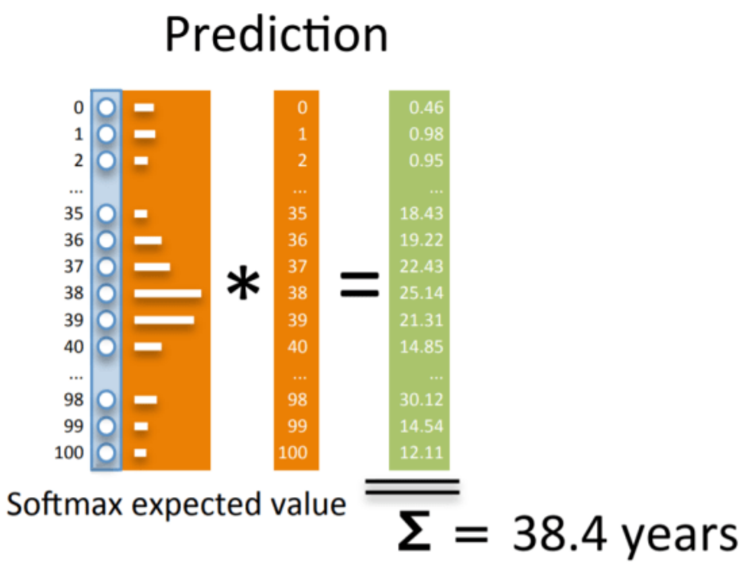


Figure 4.1 Age Prediction Approach

4.1.3 Gender Prediction

A cnn-based model trained with face photos will be used for gender prediction. The model will have 2 output classes to denote man and woman labels.

4.1.4 Race Prediction

Race prediction will be made by training a model with 6 output classes, trained with face images and ethnic background information.

The output labels will be East Asian, Southeast Asian, Indian, Black, White, Middle-Eastern and Latino-Hispanic.

4.1.5 Beauty Prediction

For the race prediction, multiple models trained with face photographs of people who have been placed in beauty contests will be used. Accordingly, for 6 races and 2 genders, a total of 12 different beauty models will be trained. Consequently, models that make an objective judgment of photogenicity of each ethnicity will be builded.

4.2 Requirements

We can list the project requirements as user, architectural, and functional requirements.

4.2.1 User Requirements

The basic user requirement in the "Identity and Beauty Analysis" project is to be able to give a photo that can be detected face as input data to the models.

4.2.2 Architectural Requirements

The project will be developed to work in a desktop environment.Tensorflow and Keras libraries will be used to train and test the CNN based models so, these libraries should be downloaded to properly use the project features.

4.2.3 Functional Requirements

In order for the project to work correctly from beginning to end, the gender and race prediction models must be accurate, respectively. If a model makes the wrong decision, the later stages of the project will also lead to wrong directions. Accordingly, in functionality, the project is expected to determine the person in the given image and estimate gender, age, race and beauty.

4.2.4 Diagrams

The work flow and data flow diagrams of the project are shown in figure 4.2 and figure 4.3.

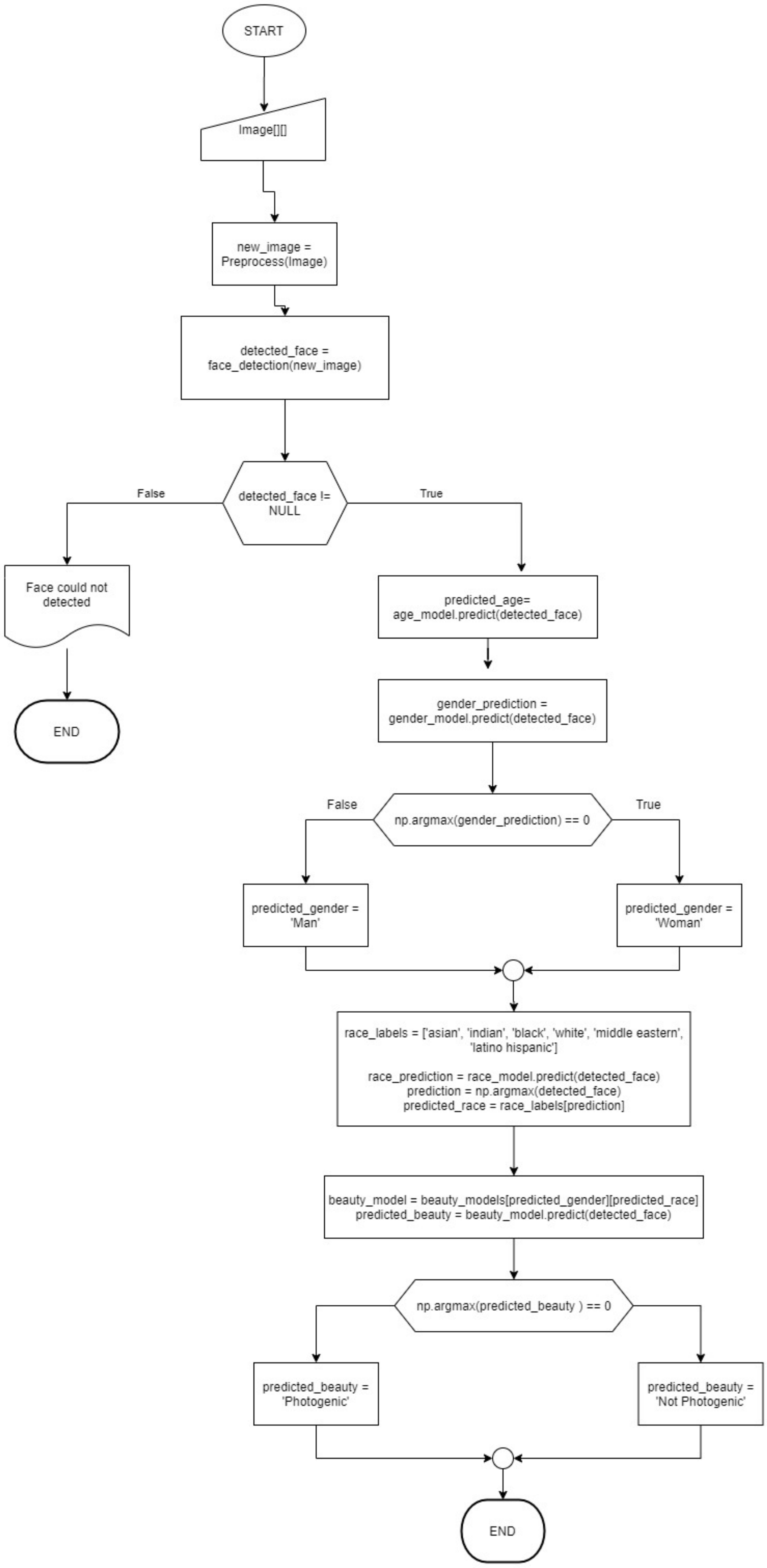


Figure 4.2 Flowchart of processes

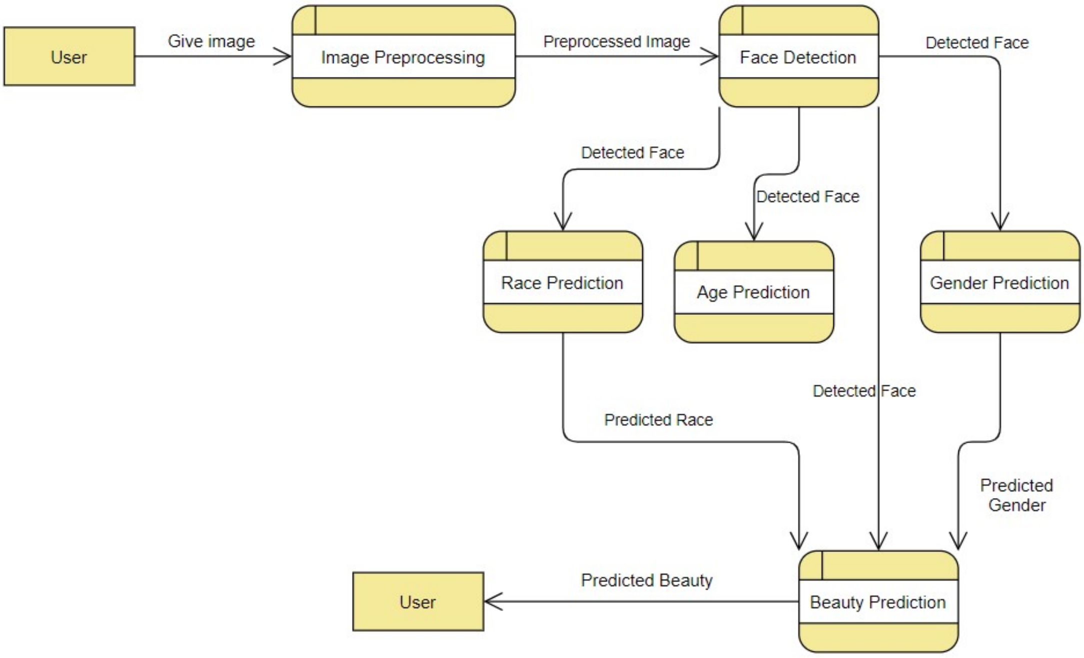


Figure 4.3 Data flow diagram

4.3 Performance Metrics

The performance criteria of the identity and beauty analysis project will be evaluated on the basis of 2 types of accuracy.

1. Individual accuracies of the modules
2. Overall accuracy of the system

Accuracy:

$$(a/b) * 100$$

Where:

a: correctly predicted label count

b: total number of test instances

5
System Design

The design phase realized after the analysis phase is over is reported in this section. This section basically consists of software design, database design and input-output design subsections.

5.1 Software Design

5.1.1 Convolutional Neural Network Structure

The Convolutional Neural Network is the structure to be used for the realization and classification of the modules in the Identity and Beauty Analysis project. This structure is used with deep learning in many areas such as image classification, object detection, face recognition.

In image classification with CNN, an input photo is taken, processed and classified under certain categories. (For example Dog, Cat, Tiger, Lion). Computers view an input image as an array of pixels, and that array depends on the resolution of the image. Depending on the image resolution, it will see y x w x b (y = Height, w = Width, b = Size). For example, an image of a 6 x 6 x 3 RGB matrix array (3 corresponds to channels of RGB values) shown in 5.1

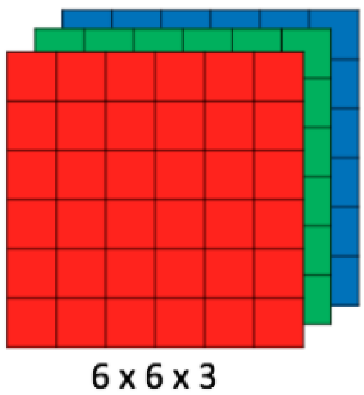


Figure 5.1 RGB Image Matrix

To train and test deep learning CNN models, each input image passes through a series of convolution layers containing filtering, merging, and fully connected layers. A vector representing the probability distribution of a list of potential results called Softmax is used to classify an object with probability values between 0 and 1. Figure 5.2 below is a complete CNN stream executed to process an input image and classify objects by their value.

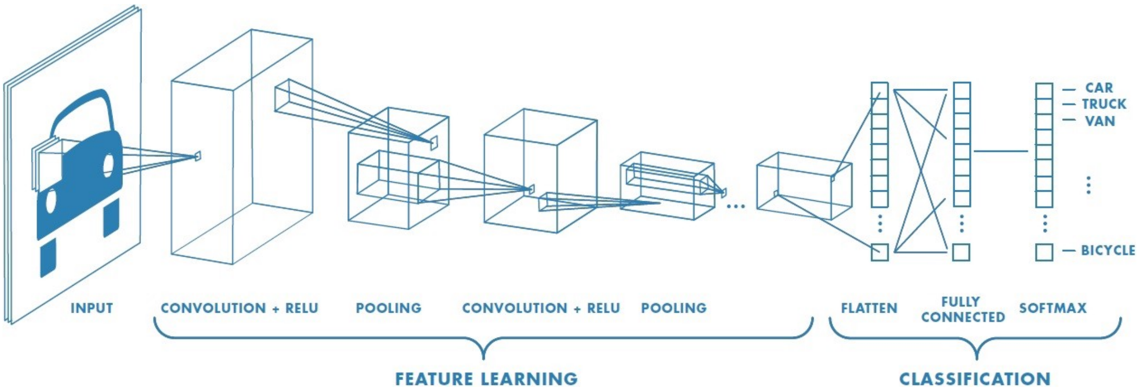


Figure 5.2 CNN Flow

5.1.2 Building a CNN Model

Since the models to be developed to make classification predictions will take the face photos as input data, it was decided to take the VGG-Face model, which has successful results with face photos, as the base model architecture.

VGG-Face [16] is a very deep CNN architecture which learned on a large scale database, is used as feature extractor to extract the activation vector of the fully connected layer in the CNN architecture .

Important Features of the VGG-Face

1. Learning a multi-way classifier
   - Softmax Objective
   - 2622 way classification
   - 4096d descriptor
2. Learning Task Specific Embedding
   - The embedding is learned by minimizing the triplet loss 5.3.
   - Learning a projection from 4096 to 1024 dimensions.

$$\sum_{(a,p,n)\in T} \max\{0, \alpha - \|\mathbf{x}_a - \mathbf{x}_n\|_2^2 + \|\mathbf{x}_a - \mathbf{x}_p\|_2^2\}$$

Figure 5.3

Figure 5.4 VGG-Face Structure

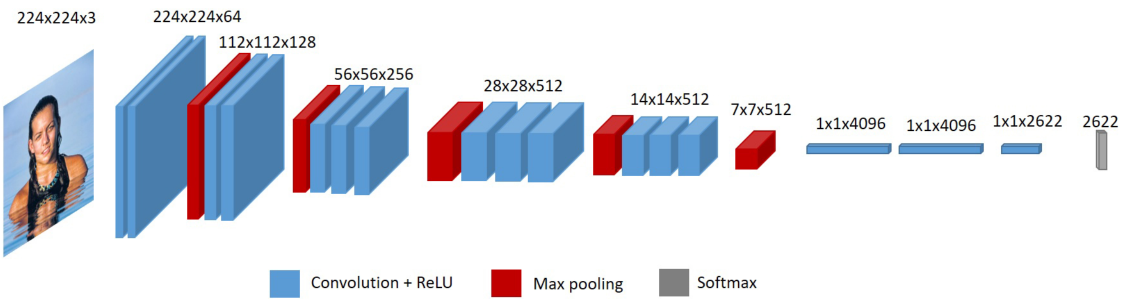


Figure 5.5 VGG-Face Structure Visualization

The structure of the VGG-Face model is demonstrated in 5.4 and 5.5.

The research paper [17] denotes the layer structre of VGG-Face as shown in 5.6.

layer	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
type	input	conv	relu	conv	relu	mpool	conv	relu	conv	relu	mpool	conv	relu	conv	relu	conv	relu	mpool	conv
name	–	conv1_1	relu1_1	conv1_2	relu1_2	pool1	conv2_1	relu2_1	conv2_2	relu2_2	pool2	conv3_1	relu3_1	conv3_2	relu3_2	conv3_3	relu3_3	pool3	conv4_1
support	–	3	1	3	1	2	3	1	3	1	2	3	1	3	1	3	1	2	3
filt dim	–	3	–	64	–	–	64	–	128	–	–	128	–	256	–	256	–	–	256
num filts	–	64	–	64	–	–	128	–	128	–	–	256	–	256	–	256	–	–	512
stride	–	1	1	1	1	2	1	1	1	1	2	1	1	1	1	1	1	2	1
pad	–	1	0	1	0	0	1	0	1	0	0	1	0	1	0	1	0	0	1

layer	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
type	relu	conv	relu	conv	relu	mpool	conv	relu	conv	relu	conv	relu	mpool	conv	relu	conv	relu	conv	softmx
name	relu4_1	conv4_2	relu4_2	conv4_3	relu4_3	pool4	conv5_1	relu5_1	conv5_2	relu5_2	conv5_3	relu5_3	pool5	fc6	relu6	fc7	relu7	fc8	prob
support	1	3	1	3	1	2	3	1	3	1	3	1	2	7	1	1	1	1	1
filt dim	–	512	–	512	–	–	512	–	512	–	512	–	–	512	–	4096	–	4096	–
num filts	–	512	–	512	–	–	512	–	512	–	512	–	–	4096	–	4096	–	2622	–
stride	1	1	1	1	1	2	1	1	1	1	1	1	2	1	1	1	1	1	1
pad	0	1	0	1	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0

Figure 5.6 VGG-Face Layers

5.2 Database Design

It was deemed appropriate to use more than one data set within the scope of the project, since the data sets required for training the models that will realize the modules of the project have different contents.

5.2.1 Dataset for the Age and Gender models

For the Age and Gender models, WIKI dataset [18] used to train and test the models. It is a large dataset that which contains face images with age and gender labels. Due to the large size of the entire data set, 1GB part of the wiki dataset has been used.

The first version of the dataset is shown in the figure 5.7.

	dob	photo_taken	full_path	gender	name	face_location	face_score	second_face_score
0	723671	2009	[17/10000217_1981-05-05_2009.jpg]	1.0	[Sami Jauhojärvi]	[[111.29109473290997, 111.29109473290997, 252....	4.300962	NaN
1	703186	1964	[48/10000548_1925-04-04_1964.jpg]	1.0	[Dettmar Cramer]	[[252.48330229530742, 126.68165114765371, 354....	2.645639	1.949248
2	711677	2008	[12/100012_1948-07-03_2008.jpg]	1.0	[Marc Okrand]	[[113.52, 169.83999999999997, 366.08, 422.4]]	4.329329	NaN
3	705061	1961	[65/10001965_1930-05-23_1961.jpg]	1.0	[Aleksandar Matanović]	[[1, 1, 634, 440]]	-inf	NaN
4	720044	2012	[16/10002116_1971-05-31_2012.jpg]	0.0	[Diana Damrau]	[[171.61031405173117, 75.57451239763239, 266.7...	3.408442	NaN

Figure 5.7 Wiki Dataset - Unprocessed

Data set contains date of birth (dob) in Matlab datenum format. The date of birth has been converted into the appropriate format so that the age can be calculated correctly. The converted dataset is shown in 5.8.

	dob	photo_taken	full_path	gender	name	face_location	face_score	second_face_score	date_of_birth
0	723671	2009	[17/10000217_1981-05-05_2009.jpg]	1.0	[Sami Jauhojärvi]	[[111.29109473290997, 111.29109473290997, 252....	4.300962	NaN	1981
1	703186	1964	[48/10000548_1925-04-04_1964.jpg]	1.0	[Dettmar Cramer]	[[252.48330229530742, 126.68165114765371, 354....	2.645639	1.949248	1925
2	711677	2008	[12/100012_1948-07-03_2008.jpg]	1.0	[Marc Okrand]	[[113.52, 169.83999999999997, 366.08, 422.4]]	4.329329	NaN	1948
3	705061	1961	[65/10001965_1930-05-23_1961.jpg]	1.0	[Aleksandar Matanović]	[[1, 1, 634, 440]]	-inf	NaN	1930
4	720044	2012	[16/10002116_1971-05-31_2012.jpg]	0.0	[Diana Damrau]	[[171.61031405173117, 75.57451239763239, 266.7...	3.408442	NaN	1971

Figure 5.8 Wiki Dataset - Date of birth converted

Some pictures don't include people in the Wiki dataset. For example, a vase picture exists in the data set. Moreover, some pictures might include two person. Furthermore, some of them are taken from distant. Face score value can help us to understand the picture has enough quality or not. Also, age labels are missing for some records. They all might confuse the model so, they should be ignored. Finally, unnecessary columns should be dropped to occupy less memory.

Some photos in the data set were taken from unborn people in an ultrasonic environment. Some records has negative age value. Within the scope of this study, the age problem is limited between 0 and 100 so considering these, the data set has been cleared from the mentioned situations and made as shown in Figure 5.9.

	full_path	gender	age
0	[17/10000217_1981-05-05_2009.jpg]	1.0	28
2	[12/100012_1948-07-03_2008.jpg]	1.0	60
4	[16/10002116_1971-05-31_2012.jpg]	0.0	41
5	[02/10002702_1960-11-09_2012.jpg]	0.0	52
6	[41/10003541_1937-09-27_1971.jpg]	1.0	34

Figure 5.9 Wiki Dataset - Cleared

The final data set consists of 22578 instances. The distributions of the labels to be predicted are shown in 5.10.

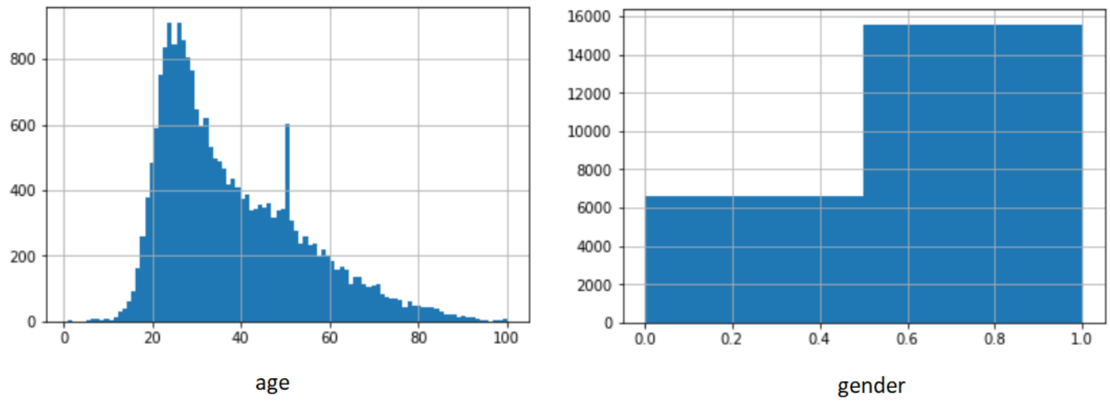


Figure 5.10 Age and gender distributions of the data set

It is splitted into 15905 train instances and 6673 test instances as 70% for training and 30% for testing.

5.2.2 Data set for the Race Prediction

For the race prediction, FairFace [19] data set used to train and test the model. It is a large scale data set and it contains of 86745 train and 10955 test instances. There are 7 race labels including Black, East Asian, Southeast Asian, Lationa_hispanic, Middle Eastern, Indian and White.

The distributions of the race labels in data set are shown in 5.11.

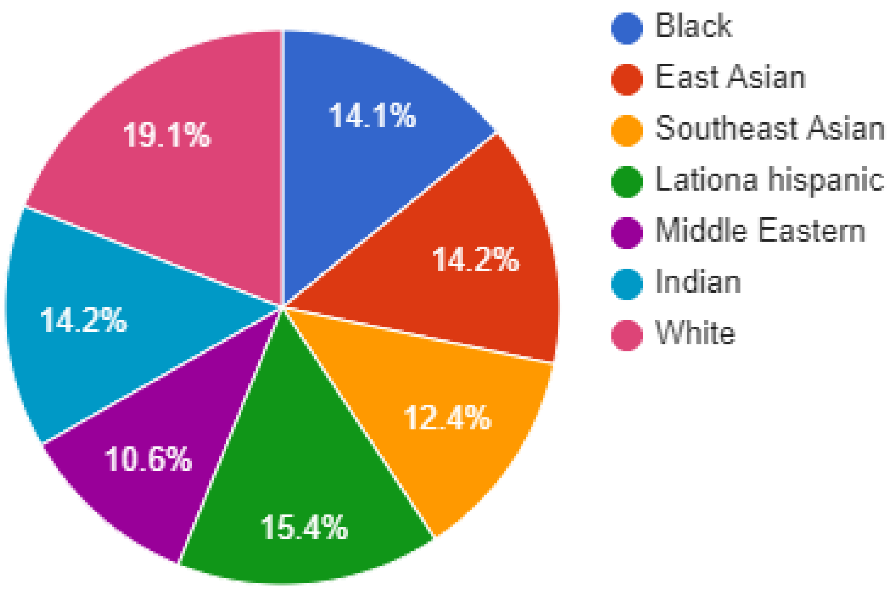


Figure 5.11 Race distribution of the data set

The distribution of labels after merging two Asian races to reduce complexity is shown in figure 5.12.

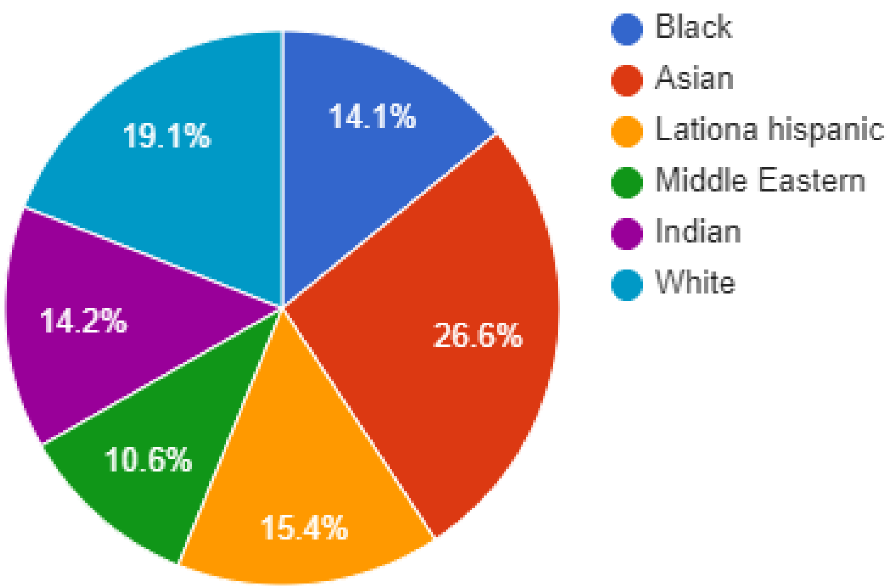


Figure 5.12 Race distribution of the data set - Asians merged

5.2.3 Beauty Prediction

For beauty prediction, since model training was desired with the photographs of people who participated in beauty contests, a suitable data set for the project objectives could not be found.

For this module of the project, it is aimed to train beauty models for each gender and race. Since the data sets of the models to be trained are not readily available, they are collected online from the relevant websites. Details of the data set collected are shown in 5.1.

Table 5.1 Beauty Pageants Contestant Dataset

Races	Men	Women
Asian	62	195
Black	50	139
Indian	47	214
Latino-hispanic	45	191
Middle Eastern	40	198
White	50	110

5.3 Input Output Design

The system basically expects the photo containing a human face to be given as input data.

After the image is given, face detection, age, gender, race and beauty prediction will be done by the trained models and the result will be reported to the user.

The final outpul will contain these predictions:

1. Age: Numeric value
2. Gender: [Man, Woman]
3. Race: [Black, Asian, Lationa_hispanic, Middle Eastern, Indian, White]
4. Beauty: [Photogenic, Not photogenic]

6
Application

In this chapter, input-output examples of the project will be shared.

Since a data set is not created where all modules of the project can be tested at the same time, the modules will be tested separately. "?" symbol will represent the unknown data.

- Example 1

Figure 6.1 Original Image

Figure 6.2 Detected Face

Table 6.1 Results of the example 1

	Gender	Age	Race	Photogenicity
True Label	Woman	?	Indian	True
Predicted Label	Woman	26	Middle Eastern	True

- Example 2

Figure 6.3 Original Image

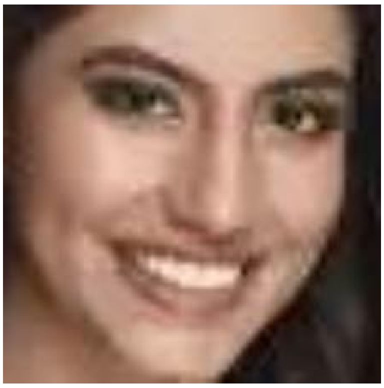


Figure 6.4 Detected Face

Table 6.2 Results of the example 2

	Gender	Age	Race	Photogenicity
True Label	Woman	?	Indian	True
Predicted Label	Woman	25	Indian	True

- Example 3

Figure 6.5 Original Image

Figure 6.6 Detected Face

Table 6.3 Results of the example 3

	Gender	Age	Race	Photogenicity
True Label	Man	?	Indian	True
Predicted Label	Man	31	Indian	True

- Example 4

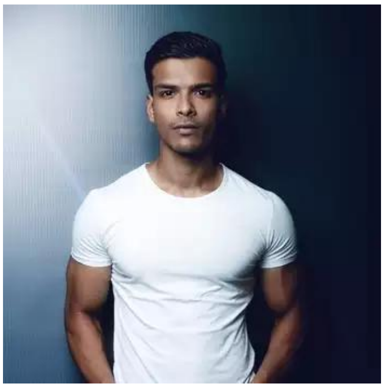


Figure 6.7 Original Image

Figure 6.8 Detected Face

Table 6.4 Results of the example 4

	Gender	**Age**	**Race**	**Photogenicity**
True Label	Man	?	Indian	True
Predicted Label	Man	24	Latino_hispanic	True

- Example 5

Figure 6.9 Original Image

Figure 6.10 Detected Face

Table 6.5 Results of the example 5

	Gender	**Age**	**Race**	**Photogenicity**
True Label	Man	28	?	?
Predicted Label	Man	32	White	True

- Example 6

Figure 6.11 Original Image

Figure 6.12 Detected Face

Table 6.6 Results of the example 6

	Gender	Age	Race	Photogenicity
True Label	Woman	52	?	?
Predicted Label	Woman	37	White	False

- Example 7

Figure 6.13 Original Image

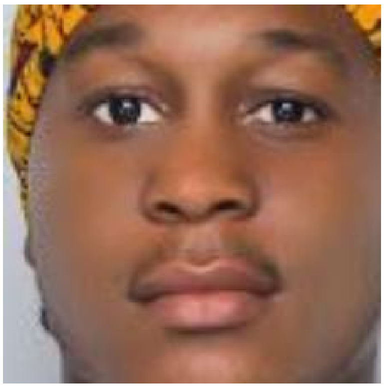


Figure 6.14 Detected Face

Table 6.7 Results of the example 7

	Gender	Age	Race	Photogenic
True Label	Man	?	Black	True
Predicted Label	Man	27	Black	True

7
Experimental Results

The project started to be developed in line with the decisions taken as a result of the researches made within the scope of feasibility.However, some unforeseen problems were encountered during the development phase.

Table 7.1 shows the success rates of 15 models in total, including gender, age, race and 12 beauty prediction models.

Table 7.1 Accuracies of the models

Models	**Accuracies**
Gender	96.67%
Age	59.24%
Race	69.77%
Beauty-Indian-male	98.95%
Beauty-Indian-female	99.76%
Beauty-Asian-female	99.23%
Beauty-Asian-male	100%
Beauty-Black-male	99.03%
Beauty-Black-female	99.28%
Beauty-Latino-female	99.47%
Beauty-Latino-male	98.95%
Beauty-Middle-eastern-male	98.74%
Beauty-Middle-eastern-female	99.75%
Beauty-White-female	99.10%
Beauty-White-male	99.03%

Under the chapter of experimental results, the difficulties encountered in the project development process and the solutions brought to these will be detailed.

1. In the model training processes, the HOG-based face detection and face alignment algorithms of the dlib library were used to standardize the images taken with oblique angles. This approach was adopted, as it was seen in preliminary studies that facial alignment increased the success rate of the

model.In the experiment, it was observed that if the original photo did not have enough size, the alignment algorithm left the edges black because it could not find pixels to place on the edges of the photo. For this reason, the alignment step in face detection was canceled.

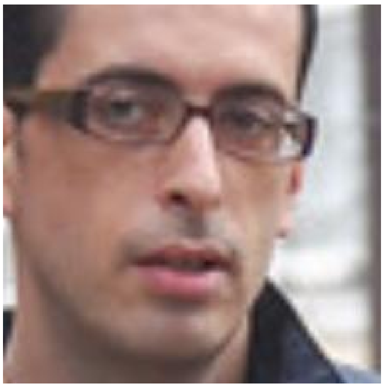


Figure 7.1 Original Image

Figure 7.2 After the Alignment

2. In the age prediction part, which is one of the main modules of the project, it was observed that the accuracy was low when the predicted age was compared directly with the actual age. Since the age estimation from the photograph depends on many different parameters such as the quality of the photograph, the light of the environment, make-up and genetics, it is a very difficult area to estimate points as a single number.For this reason, the predictions have been designed to estimate an age range.The age estimated during the test process was considered correct if it was between 5 minus and 5 more than the real age.

3. Another problem was encountered in the beauty prediction part of the project.The Evaluate results shows that the accuracy of all models were 100%. It was thought that the model achieved 100% success with data it had never seen, which was a bit suspicious. It was tested to what extent it was successful in randomly taken photos from the internet that were not in the data set. As a result of the experiments, it was seen that the model produced very bad results with real world data.When the steps in the development process were examined again, it was seen that the problem was in the face detection process.

   The HOG based face detection algorithm of the used Dlib library is designed to return the coordinates of the corner points of the frame of the captured face.If an edge of the frame of the detected face extends beyond the original photo, the coordinates of the relevant corners of the frame are returned with negative values.Accordingly, when the frame boundaries are out of the photo and the region with detected face is to be cropped from the original photo, the cropping process is formatted incorrectly.In order to avoid this extreme situation, when the edges of the detected frame extend beyond the size of the original photo, the frame was rearranged to fit within the boundaries of the original photo.After this arrangement, the models trained with correctly detected faces were able to achieve the expected success with both test samples selected from the dataset and real world data.

8
Performance Analysis

Within the scope of this study, a total of 15 models with CNN architecture were trained.Considering that the implemented modules try to solve the difficult problems realistically, it can be said that satisfactory results have been obtained. The least successful models were the race prediction and age prediction models.It was mentioned in the previous chapter that an age range is accepted to measure the accuracy rate for age prediction.The change in this acceptance range directly affects the success of the model.

Another point to note is that predicting race is a difficult problem, as is age estimation. When the class confusion matrix of the race model is examined, it has been observed that the most incorrectly predicted class is the middle-eastern class. As shown in Figure 8.14, a significant portion of the Middle Eastern class has been misclassified as latin-hispanic and white classes.Indeed, the results are acceptable due to the presence of people similar to other classes within the middle eastern class.

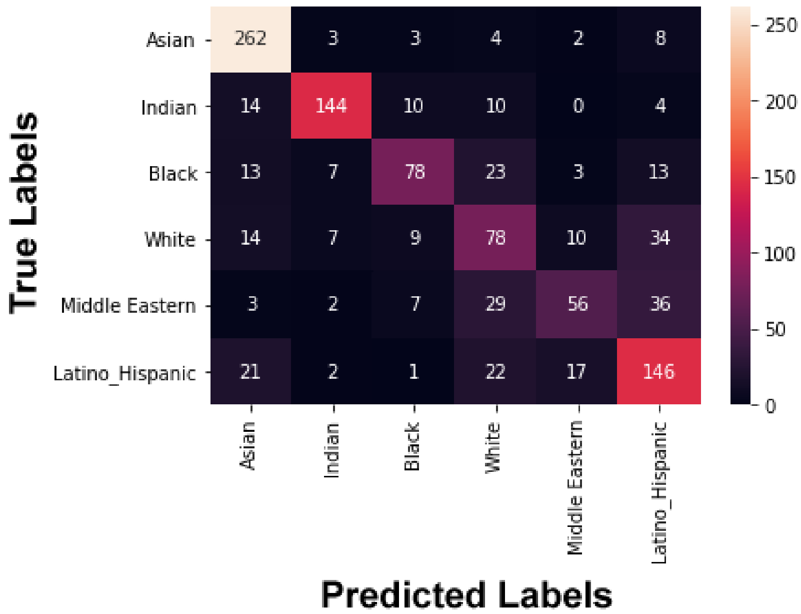


Figure 8.1 Class confusion matrix of race model

Gender model results:

Figure 8.2 Class confusion matrix of gender model

8.1 Confusion Matrixes of Beauty Models

1. Test results for beauty model of indian-female.

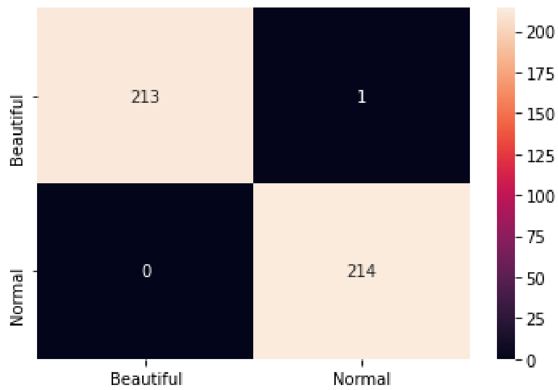


Figure 8.3 Class confusion matrix of beauty-indian-female model

2. Test results for beauty model of indian-male.

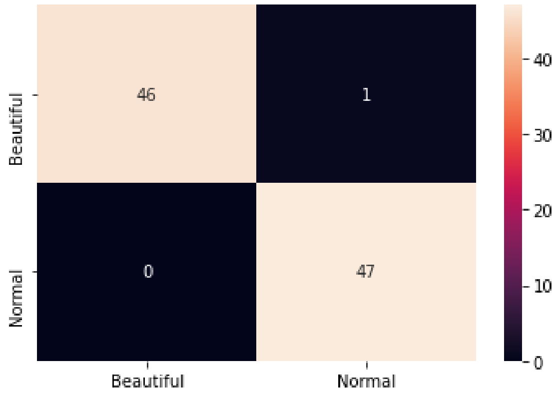


Figure 8.4 Class confusion matrix of beauty-indian-male model

3. Test results for beauty model of asian-female.

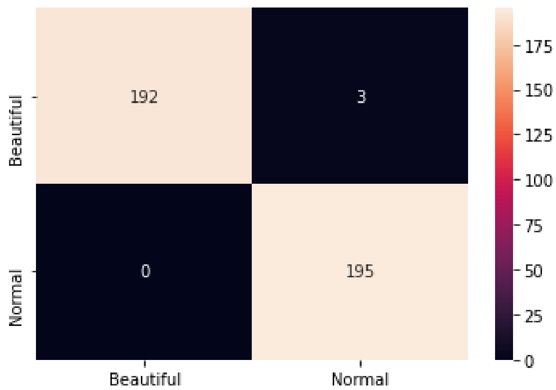


Figure 8.5 Class confusion matrix of beauty-asian-female model

4. Test results for beauty model of asian-male.

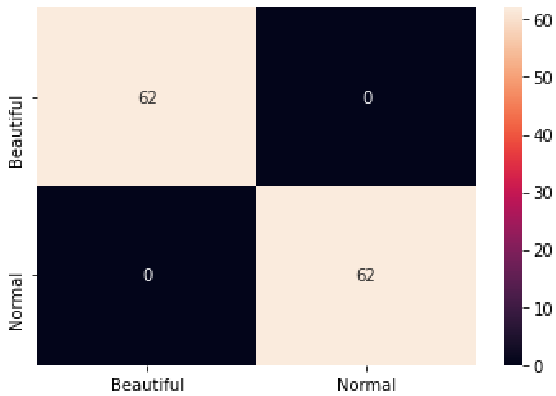


Figure 8.6 Class confusion matrix of beauty-asian-male model

5. Test results for beauty model of black-male.

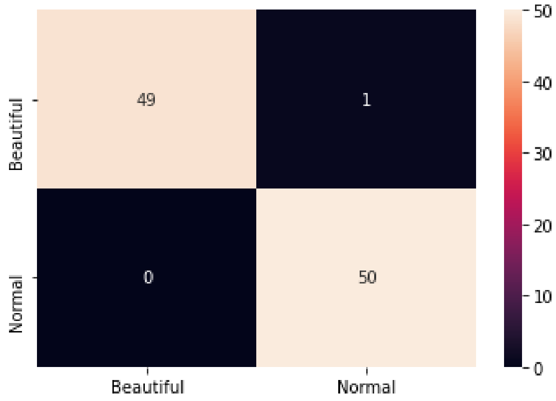


Figure 8.7 Class confusion matrix of beauty-black-male model

6. Test results for beauty model of black-female.

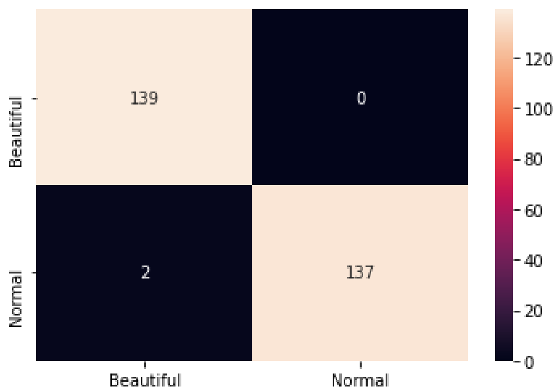


Figure 8.8 Class confusion matrix of beauty-black-female model

7. Test results for beauty model of latino-female.

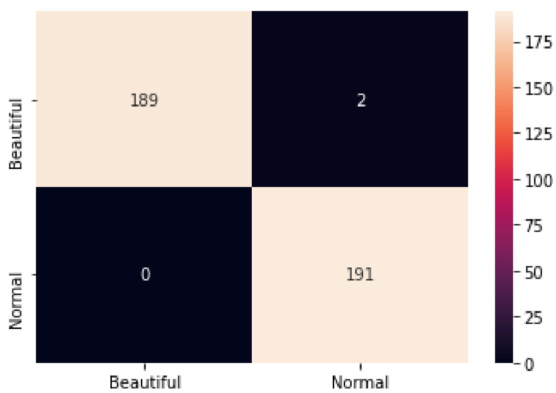


Figure 8.9 Class confusion matrix of beauty-latino-female model

8. Test results for beauty model of latino-male.

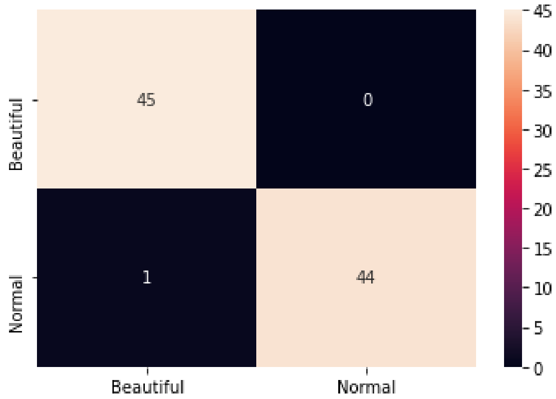


Figure 8.10 Class confusion matrix of beauty-latino-male model

9. Test results for beauty model of middle eastern-male.

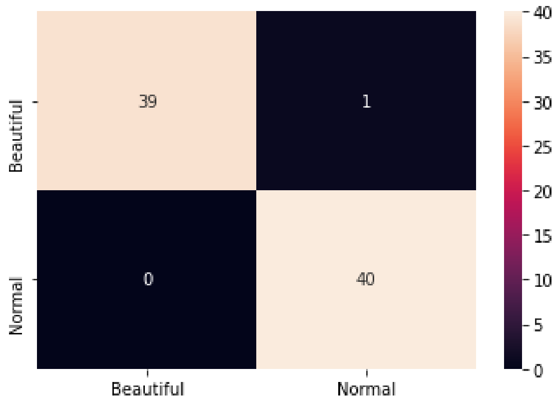


Figure 8.11 Class confusion matrix of beauty-middle eastern-male model

10. Test results for beauty model of middle eastern-female.

Figure 8.12 Class confusion matrix of beauty-middle eastern-female model

11. Test results for beauty model of white-female.

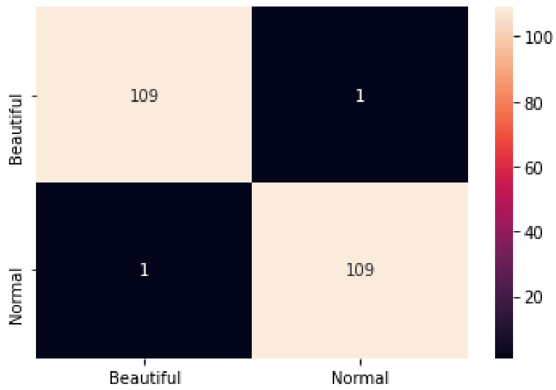


Figure 8.13 Class confusion matrix of beauty-white-female model

12. Test results for beauty model of white-male.

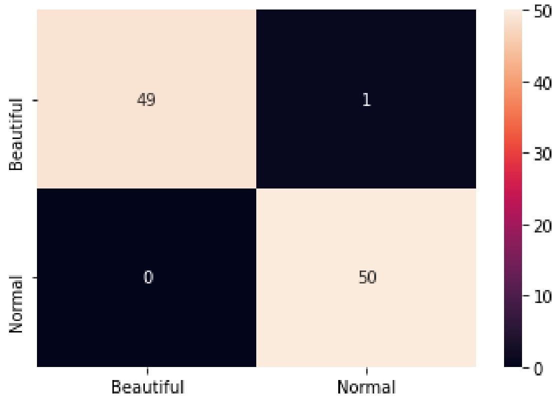


Figure 8.14 Class confusion matrix of beauty-white-male model

9 Result

An overview of deep learning and how methods and techniques can be used for convolution neural networks were examined in general within the scope of this project. In this context, all of the targeted gender, age, race and beauty prediction modules were realized.

The main goal of the project which is the obtaininig a separate beauty model for each ethnic group that have been trained in data sets consisting of beauty contest participants, grouped according to ethnic origins has been accomplished.In this way, since the photogenicity of the people is determined according to the beauty perceptions of their own society, a closer to the real beauty rating was made. More than 90% accuracy percentile has been achieved in all beauty models.

In order to take the study a step further, model training can be done by applying face alignment with different datasets.Aligning low-size photos that are already cropped, as described in the previous sections, has reduced success.Face detection is critical in all facial prediction problems.Instead of a HOG-based algorithm for face detection and alignment, a better face detection method such as CNN-based models can be used.

References

[1] M. K. Ashu Kumar Amandeep Kaur, "Face detection techniques: A review," 2019.

[2] L. Cuimei, Q. Zhiliang, J. Nan, and W. Jianhua, "Human face detection algorithm via haar cascade classifier combined with three additional classifiers," 2018.

[3] Y. Pang, Y. Yuan, X. Li, and J. Pan, "Efficient hog human detection," *Signal Processing*, vol. 91, no. 4, pp. 773–781, 2011, ISSN: 0165-1684. DOI: https://doi.org/10.1016/j.sigpro.2010.08.010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0165168410003476.

[4] Y. H. Kwon and N. da Vitoria Lobo, "Age classification from facial images," *Computer Vision and Image Understanding*, vol. 74, no. 1, pp. 1–21, 1999, ISSN: 1077-3142. DOI: https://doi.org/10.1006/cviu.1997.0549. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S107731429790549X.

[5] S. Chen, C. Zhang, M. Dong, J. Le, and M. Rao, "Using ranking-cnn for age estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017.

[6] R. Rothe, R. Timofte, and L. Van Gool, "Dex: Deep expectation of apparent age from a single image," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV) Workshops*, Dec. 2015.

[7] K. Jhang and J. Cho, "Cnn training for face photo based gender and age group prediction with camera," in *2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, 2019, pp. 548–551. DOI: 10.1109/ICAIIC.2019.8669039.

[8] K. Jhang, "Gender prediction based on voting of cnn models," in *2019 International Conference on Green and Human Information Technology (ICGHIT)*, 2019, pp. 89–92. DOI: 10.1109/ICGHIT.2019.00028.

[9] Z. Abbas, S. Ali, M. A. Baloch, H. Ilyas, M. Ahmad, M. H. Malik, N. Javaid, and T. F. N. Bukht, "Joint demographic features extraction for gender, age and race classification based on cnn,"

[10] J. Chen, S. Shan, C. He, G. Zhao, M. Pietikäinen, X. Chen, and W. Gao, "Wld: A robust local image descriptor," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1705–1720, 2010. DOI: 10.1109/TPAMI.2009.155.

[11] J. Xu, L. Jin, L. Liang, Z. Feng, and D. Xie, *A new humanlike facial attractiveness predictor with cascaded fine-tuning deep learning model*, 2015. arXiv: 1511.02465 [cs.CV].

[12] Martin Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Y. Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mane, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viegas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng, *TensorFlow: Large-scale machine learning on heterogeneous systems*, Software available from tensorflow.org, 2015. [Online]. Available: https://www.tensorflow.org/.

[13] F. Chollet *et al.* (2015). "Keras," [Online]. Available: https://github.com/fchollet/keras.

[14] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.

[15] (2020), [Online]. Available: https://ppp-certification.com/ppp-certification-guide/151-process-analyzing-legal-feasibility-and-conducting-legal-due-diligence (visited on 12/04/2020).

[16] Lihong Wan, Na Liu, Hong Huo, and Tao Fang, "Face recognition with convolutional neural networks and subspace learning," in *2017 2nd International Conference on Image, Vision and Computing (ICIVC)*, 2017, pp. 228–233. DOI: 10.1109/ICIVC.2017.7984551.

[17] O. M. Parkhi, A. Vedaldi, and A. Zisserman. (2015). "Deep face recognition," [Online]. Available: https://www.robots.ox.ac.uk/~vgg/publications/2015/Parkhi15/parkhi15.pdf.

[18] R. Rothe, R. Timofte, and L. V. Gool. (2015). "Imdb-wiki – 500k+ face images with age and gender labels," [Online]. Available: https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki/.

[19] K. Kärkkäinen and J. Joo, "Fairface: Face attribute dataset for balanced race, gender, and age," *arXiv preprint arXiv:1908.04913*, 2019.

Curriculum Vitae

FIRST MEMBER

Name-Surname: Rahmi Cemre ÜNAL
Birthdate and Place of Birth: 15.09.1996, Konya
E-mail: rahmicemreunal@gmail.com
Phone: 0531 465 35 15
Practical Training: Kafein Software Solutions - AI Department

Project System Informations

System and Software: Windows 10, Python
Required RAM: 8GB
Required Disk: 10GB