Age, mood and gender prediction using deep learning

Filip Byberg
Universitat Politècnica de Catalunya
BarcelonaTech, Barcelona, Spain

CONTENTS

Abstract **- Predicting the age, gender and other characteristics from a human face has turned into a growing area of research due to its numerous applications in marketing or law enforcement. This paper presents the development of convolutional neural networks (CNN) designed to predict the age, gender and emotion from a face, by using large datasets of labelled facial images. The CNNs are then used to simulate a real-life application of predicting whether a person is a potential customer or not based on their appearance. The results of the simulations are extensively discussed.**

I. Introduction

Face analysis of images in the wild still pose a challenge for automatic age, mood and gender recognition tasks, mainly due to their high variability in resolution, deformation, and occlusion. Face verification has turned into an area of dynamic research and the applications are important in law enforcement because it can be done without involving the subject. These types of applications are increasingly more used in our daily lives on a wide spectre, all the way from social media filter use, to visual surveillance and biometric analysis. The existing methods have quite satisfying performance on real-world images, however, due to vast differences of facial structure and features between different ethnic groups, can produce unexpected results and confusion in the application. For this reason, the project is centered around having a good dataset with a variety in ethnicity, age and moods as a good fundamental for the model-training in the application.

The aim of this project is to build a sufficiently accurate application that can predict the age, gender and emotion of a person based on their facial features. The paper is organized as follows : section 2 explains the state of the art related to age, gender and emotion detection. Section 3 describes the proposed approach. Section 4 and 5 present the model and the experiments. Finally, section 6 includes the conclusions and discussions around the project.

II. State of the Art

Age, gender and emotion predictions from facial features are important problems in computer vision and pattern recognition, due to their extensive applications in human-computer interaction, surveillance monitoring, biometric analysis and so on.

A. Age prediction

Estimating the age of a person from facial features has received great interest in the recent years [2],[3]. There are multiple computational challenges to accurately estimate age, as facial appearances due to ageing greatly vary depending on gender, ethnicity and lifestyle.

Age estimation based on facial features is most commonly formulated as a classification problem, where each face image is assigned to a class label that represents a range of ages; intuitively, it's easier to predict whether someone falls into a certain age range, than to predict their exact age based on their appearance.

The earliest studies used the size and proportions of the human face, but they were quite limited to young ages and not appropriate for adult faces. In the last decade, age prediction based on machine learning took over, with methods based on convolutional neural networks and support vector machines being the most used.

B. Gender prediction

Gender prediction is a problem closely related to age prediction and face recognition in general. One of the first methods for gender classification was developed by Golomb. et al [4]. It is an early example of a fully connected neural network with 3 layers of 90, 40 and 90 neurons that classified images samples at 30x30 pixels.

Another method uses Support Vector Machines to classify gender [3],[5].

Most recently, gender prediction has been treated as a similar problem to age prediction, using deep learning to classify gender.

C. Emotion prediction

Emotion recognition is the process of identifying human emotion. There are several subfields of emotion recognition : text, audio, video and physiology, which includes predicting emotions based on facial features. Face emotion recognition typically uses Support Vector Machines [6] or neural networks.

D. Customer Prediction

Customer prediction is often related to comprehensive strategies, such as Customer Relationship Management (CRM). This typically aims at building, managing and strengthening loyal and long-lasting customer relationships. Several, very popular in research community machine learning algorithms have been proposed in order to tackle the customer prediction problem. Such methods include Artificial Neural Networks [7], Regression Analysis [8], Logistic Regression [9], Support Vector Machines [10], Linear Discriminant Analysis [11] and Decision Trees learning [7]. A recently published paper *"A comparison of machine learning techniques for customer churn prediction"* [12], presents a comparative study on the most popular machine learning methods, many of which is mentioned above.

III. Proposed approach

The objective to be achieved is to build a highly accurate application that can predict a person's age range, gender and mood by their facial features. The goal is to use this application for further studies in the marketing industry, where the classification of different people is used to determine if a person entering a store is likely to buy a product or not, solely on their facial appearance. For this reason, the project is split into two main parts, where the first part revolves around the application, and the second part is testing the application for the marketing purposes.

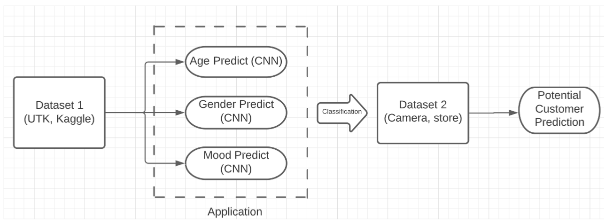

Figure 1: General Idea

Figure 1 shows an illustration of the general concept for this project. Dataset 1 represents the face-images used to train and implement the application for classifying the faces by prediction, for the second part. Dataset 2 is supposed to simulate the gathered face-images in an arbitrary store, classified by the application and containing information about predicted age group, predicted gender, predicted mood as they enter the store and whether they are a buying customer or not.

A. Choosing dataset

For the application to be able to perform as accurately as possible, a good dataset with a lot of diversity and variety is necessary. The considerations that were taken, when choosing a dataset, was the following:

- Variety in ethnicity
- A wide age range
- Even gender distribution
- A variety of facial expressions
- Augmented/rotated face-images

To be able to represent all the criteria shown above, the two datasets; *UTK Faces* [13] with 23 708 images, and *Kaggle Facial Age* [14] with 9 778 images was chosen, resulting in a total of 33 486 images.

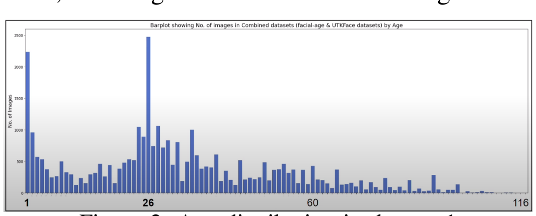

Figure 2: Age distribution in dataset 1

Figure 2 shows the distribution of ages in dataset 1, with a range from 1 to 116. The majority of data is distributed around the ages 1 and 26. There are also significantly less data for ages 70 and above. However, since the majority of people entering a store is below 70 years old, the dataset is considered acceptable. For the gender prediction, only the *UTK faces* dataset was used, due to poor gender labeling in Kaggle's dataset. For the mood prediction, an alternative dataset called *CK+* was used [15]. This dataset provides sub-folders of seven expression classes, saving time for manual labeling.

B. Image pre-processing

Most of the images from the given datasets is portrait-like face-images in upright positioning. In the real world, this is often not the case, where faces in images might be tilted. To address this issue, a data augmentation can be used on each image to slightly modify the image with rotation. A process like this creates a total of 10 augmented images, per original image. This works in the favor of the application, as it gives a wider selection in the amount of training dataset. More about the method of image augmentation used can be found in [16]. Figure 3 shows the image pre-processing for age prediction. The combined set of images are split into a training set and a testing set. This can simply be done using *sklearn train_test_split* method [17].

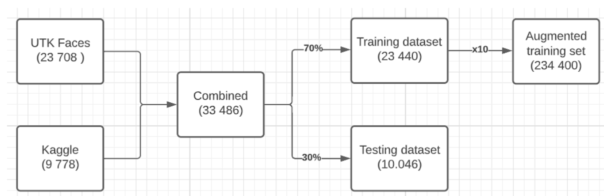

Figure 3: Dataset 1 image pre-processing

Since the images from the dataset are *RGB*, a grayscale function is necessary to reduce computation [18].

IV. Classification

So far, the basic procedure for the application is addressed. The application consists of the three separate predictions, for later use as classification in the experimentation part. Since age, mood and in some cases gender, can be difficult to determine even by humans, a limited selection of classes are made for each prediction model. This helps the program make more generalized decision when classifying each dataset. For the three prediction models, the following classes was made.

Age:	**Mood:**	**Gender:**
Age 1-2	Positive	Male
Age 3-9	Negative	Female
Age 10-20	Neutral	-
Age 21-27	-	-
Age 28-45	-	-
Age 46-65	-	-
Age 65-112	-	-
Ref: [13, 14]	Ref: [15]	Ref: [19]

The distribution of ages in the different classes was decided after many trials, where the aim was to not make overly vast generalizations, as this would serve no good purpose to an age prediction model. At the same time, it was also so that it wouldn't be too specific, as this would lead to a higher rate of misclassification. The resulting classes that seemed the best took into consideration the available number of images per age-range and human intuition as main factor, meaning age groups that humans would likely classify a person into. The references for the labeling of each dataset, used for the three prediction methods are given at the bottom of the tabular.
There are a variety of techniques available out there, that deal with feature extraction from images for classification modelling. For this project, deep learning using *CNN* was seemingly the best option, according to previous studies mentioned earlier. To avoid high *RAM* consumption by converting the images into a *Pandas* Dataframe, dataset pipelines was created, using *TensorFlow* [22]. The usefulness of using *TensorFlow* when dealing with a limited *GPU*, is demonstrated in previous work (Zuppichini, 2018, [23]).

A. *CNN Architecture*

The three prediction models are to be trained using the above defined classes. What remains for the application is to define the architecture from the sequential neural network. The amount of convolutions and *Pooling* in the sequence determine the feature extraction in the architecture. Where a total of four convolutional layers with filters increasing by a factor of 2, for every successive layer and an *Average Pooling*. From here, *densing* is applied to reduce the number of output layers to the amount of classes for each prediction model.
Figure 4 shows the CNN architecture for the age prediction, and results in an output layer of seven nodes and a total of 422 695 trainable parameters. The construction and design of this architecture was inspired by a scientific article on the subject (Mishra, 2020, [25]).

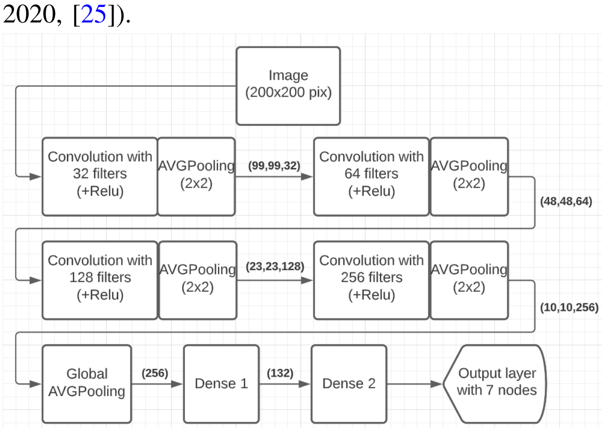

Figure 4: CNN architecture for age prediction

The other two prediction models, will have quite similar architecture, with the exception of the output layer, with the corresponding amount of nodes for each model. As for now, these models can be fitted to complete the trained model sets. Each model is to be trained and tested with their according datasets as shown in figure 5, where the performance for each model fitting is displayed over the set amount of epochs. Finally, the tabular shows the best scores.

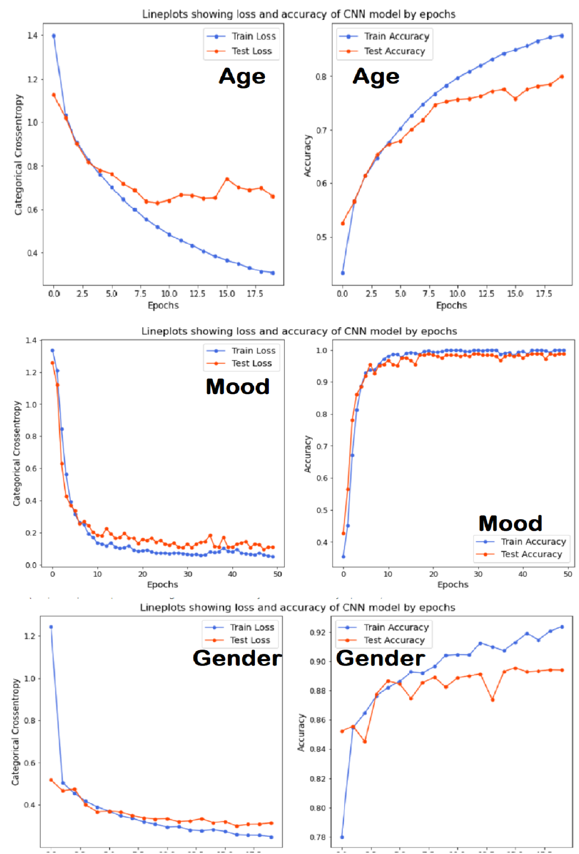

Figure 5: Loss and accuracy for applications

Prediction	Test Accuracy	Test Loss	Epoch
Age	0.7988	0.6608	20
Mood	0.9769	0.1358	50
Gender	0.9237	0.24978	20

As shown in the tabular, mood and gender prediction shows promising results, while the age model has decently lower performance. This is due to the GPU used, not being able to handle the large dataset and complex CNN architecture optimally. Resulting in a lower amount of epochs used, then initially planned. However, after plenty of testing, this was the best result obtainable within the limited time availible.

B. *Customer classification*

The customer classification is done in two parts. Firstly, the existing application is used on dataset 2 shown in figure 1. This will label each subject in the dataset, with an estimation of their age, gender and mood. However, another label is needed to indicate whether the subject is a customer or not. To that end, every image is randomly labelled as customer (1) or non-customer (0). The reason for this comes from multiple factors : it's very hard to find labelled datasets of customers/non-customers and manually labelling thousands of images is a laborious task. To avoid too much randomness in the labelling process, 50% of subjects between the ages of 15-45 will be labeled as customers, while the rest will be labelled as non-customers. As a result, about 30% of the total dataset is labelled as customers. Considering that the program chooses half of the subject in this age class at randome, the accuracy for the customer prediction is not expected to be above 50% at best.

Secondly, now that the dataset has been pre-processed, the neural network used for customer classification is implemented and trained. Below are the loss and accuracy for the training and testing datasets :

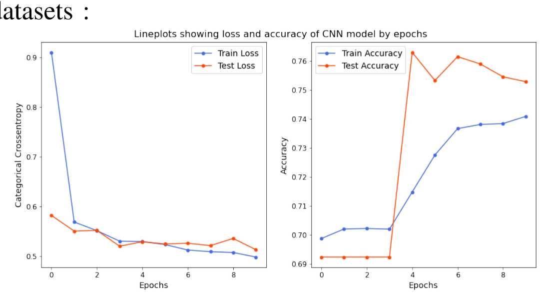

Figure 6: Loss and accuracy for customer classification

Loss and accuracy performance over 10 epochs are shown in figure 6. As you can see, the peak performance is at the 4th epoch for the accuracy with the value 0.76, while the loss slightly keeps improving for each epoch with the final value 0.53.

V. Experiment

In this section, experiments will be conducted such as testing the application and using it for the customer classification process.

A. Application experiment

To analyze the performance of the application, its necessary to take a closer look at the confusion matrix of the classifications. Considering that the age prediction model has the biggest input dataset and the lowest performance of the three, there will be a closer inspection of its prediction performance.

The confusion matrix for the age prediction model is shown in figure 7.

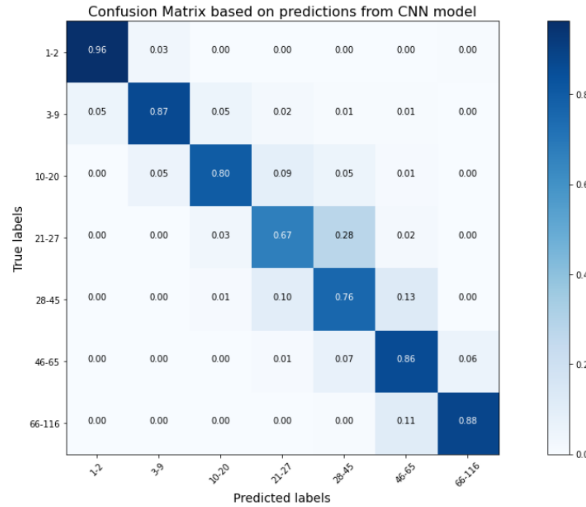

Figure 7: Confusion matrix for age prediction model

Even though the model performs quite good from ages 1-20 and 45-116 with a success classification rate between 80-96%. The middle part of the age groups performs significantly lower. A closer look at the worst performance class, 21-27 with a success rate of 67%, shows that the majority of the miss-classification falls in the above age class 28-45, 28% of the time.

There are multiple factors that might play a role in the low performance in this class, such as:

- Most of the subjects in the dataset are clustered around this age group ie. more variety in the class (ref. figure 2).
- The class has a short age span of only 7 ages.
- Naturally, its often hard to distinguish people in their mid 20s and early 30s.
- The training-set probably could use a better model fitting process with a larger amount of epochs, as the accuracy for each epoch was still improving. (ref. figure 5)

To further explore the performance, and take a closer look at instances where the application misclassifies, a couple of face images was used to analyze the performance. The goal here is to get a result that a human agrees with by first glance of the faces. These test images is shown in figure 8, with a corresponding tabular showing the result for each labeled subject.

Figure 8: Test faces used to analyze the application

Face label	Age	Mood	Gender
1	28-45	Negative	Male
2	28-45	Positive	Male
3	10-20	Positive	Female
4	10-20	Neutral	Female
5	46-65	Negative	Male
6	28-45	Neutral	Male
7	1-2	Neutral	Male
8	66-116	positive	Female
9	28-45	Positive	Male
10	46-65	Negative	Female

At first glance the test shows generally good results, with very few major flaws in the classifications. With the exception that faces 3 and 9 are misgendered. There are also some minor age misclassifications, however since reasonable arguments can be made that the subjects could also possibly belong to this particular age group, it's decided that this is to be neglected from further analysis.

Multiple other tests with similar faces were conducted, with the aim of finding correlations between certain features and repetitive misclassifications. The following observations were found.

- Beards and mustaches on younger men, often makes the application think the subject is older then they are.
- Complete lack of beard and/or mustache on men, occasionally makes the program think the subject is younger then they are.
- Glasses tend to confuse the application.
- Asian people are often misclassified in gender and/or age.
- If the image only contains a closeup of the face, instead of their full head, the program sometimes has major misclassifications.
- Women with short hair are sometimes classified as males.
- Resized images tend to have major misclassifications.
- Images of people smiling with teeth sometimes misclassifies the mood, as the training dataset does not contain a lot these type of subjects.

B. Customer experiment

The performance of the customer classification model can be visualized with the confusion matrix in figure 9.

To analyze the performance, the appropriate metrics are computed, such as the true positive rate or sensitivity (TPR) and true negative rate or specificity (TFR):

$$TPR = \frac{TP}{TP+FN} = \frac{0.1298}{0.1298+0.1778} = 0.42 = 42\%$$

$$TFR = \frac{TF}{TF+FP} = \frac{0.6231}{0.6231+0.0693} = 0.90 = 90\%$$

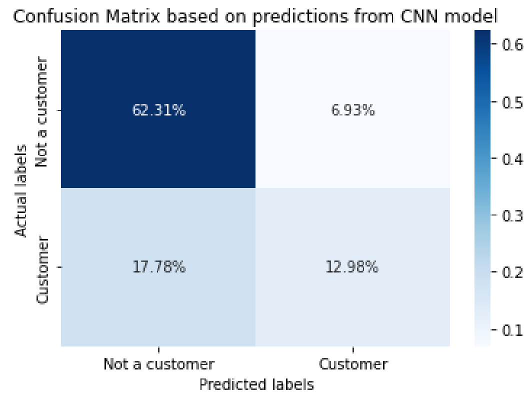

Figure 9: Confusion matrix for customer classification model

The customer classification model is quite accurate for predicting non-customers, showing a 90% sensitivity. As for predicting customers, since the labelling of customers was a random process, a result better than 50% accuracy is highly unlikely, so 40% accuracy is a passable result.

VI. Conclusion

In this paper, we presented our development of neural networks designed to predict the age, gender and mood of a person from an image of their face. We obtained high accuracies for each neural network as shown in figure 5 and we explored their limitations, such as misclassifications due to features like facial hair, glasses or the pictures being close-ups or resized images. The networks were then used to simulate a real-life application, in the form of a customer prediction model. Despite the random labelling, it showed decent results and could be significantly improved with access to real-life datasets.

A. Discussion

In this final section, aspects around the challenges we encountered during the projects will be discussed. This is in the form of stating a problem encountered, how it was solved and/or proposing a better solution.

- ***Google Colab Free*** **did not provide enough RAM to handle our CNN architecture optimally.**
  - This was mainly a problem for the age prediction model, and was resolved by lowering the amount of convolutions and Epochs, than initially planned, resulting in a model with lower accuracy.
  - *Kaggle* provides a free notebook with access to NVidia K80 GPUs in kernels. This can potentially solve this particular problem on a cost-free basis.
- **The chosen age classes might not be optimal.**

- A couple of strategies were considered for the choosing of our age classes, where the aim was mainly to not make too much generalizations and also not much specifications in the classes. The ending classes that seemed the best, took into consideration the available number of images per age-range and human intuition as main factors for determining the optimal classes. However this caused some accuracy drop for subjects between 20-30 years.
- Having a larger dataset with more even distribution in the different ages, would make the model more robust to changes in the chosen age classes. However, such a dataset can be challenging to find or create.

- **Importing the huge image datasets and converting them into *Pandas* Dataframe with individual pixel values as columns and the images as rows, resulted in a huge and heavy dataframe.**
  - This required all the pixel values to be scaled down from 0–255 integers to 0–1 floats, making the dataframe 33,486 rows X 40,000 columns. This led to constant "Out of Memory" errors. This was avoided by using *TensorFlow* dataset pipelines, to significantly reduce RAM consumption.
- **Issues surrounding the customer classification model.**
  - Finding a labelled dataset with informations about customers was not possible, so we had to settle with random labelling which is not ideal since it doesn't capture any potential relationship involving the customers and the product. The CNN's architecture had to be simplified several times because it was overfitting and predicting no customers at all on a testing dataset of thousands of people.
  - A better solution the group would prefer to use, would be to label more specific customers. An example of this would be to label all females between the age of 15-25 with a positive mood. This would be a good example to target a specific group of people with less randomness.

References

[1] H. Han, C. Otto, and A. K. Jain, \Age estimation from face images: Human vs. machine performance," in Proc. Int. Conf. BTAS, Jun. 2013, pp. 1{8.

[2] Hierarchical Attention-based Age Estimation and Bias Estimation, Shakediel Hiba, Yosi Keller, Bar-Ilan University

[3] Structured Output SVM Prediction of Apparent Age, Gender and Smile From Deep Features : Michal Uricar, Radu Timofte, Rasmus Rothe, Jiri Matas, Luc Van Gool; Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2016, pp. 25-33)

[4] GOLOMB, Beatrice A. ; LAWRENCE, David T; SEJNOWSKI, Terrence J. SEXNET: A Neural Network Identifies Sex From Human Faces. In: MPS. 1990.

[5] MOGHADDAM , B.; MING-HSUA N YANG . Gender classification with support vector machines. In: Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580). 2000, pp. 306-311.

[6] Rajesh, K. M.; Naveenkumar, M. (2016). [IEEE 2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT) - Mysuru, India (2016.12.9-2016.12.10)] 2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT) - A robust method for face recognition and face emotion detection system using support vector machines.

[7] C. Kirui, L. Hong, W. Cheruiyot, H. Kirui [Predicting customer churn in mobile telephony industry using probabilistic classifiers in data mining] Int. J. Comput. Sci. Iss. (IJCSI), 10 (2) (2013), pp. 165-172

[8] J. Hadden, A. Tiwari, R. Roy, D. Ruta [Churn prediction: does technology matter] Int. J. Intell. Technol., 1 (2) (2006), pp. 104-110

[9] Ş. Gürsoy, U. Tuğba [Customer churn analysis in telecommunication sector] J. School Bus. Admin. Istanbul Univ., 39 (1) (2010), pp. 35-49

[10] E. Shaaban, Y. Helmy, A. Khedr, M. Nasr [A proposed churn prediction model] J. Eng. Res. Appl., 2 (4) (2012), pp. 693-697

[11] Y. Xie, X. Li [Churn prediction with linear discriminant boosting algorithm] 2008 International Conference on Machine Learning and Cybernetics, vol. 1, IEEE (2008), pp. 228-233

[12] T.Vafeiadis, K.I.Diamantaras, G.Sarigiannidis, K.Ch.Chatzisavvas [A comparison of machine learning techniques for customer churn prediction] Simulation Modelling Practice and Theory, Vol. 55, (2015), pp. 1-9

[13] https://susanqq.github.io/UTKFace/

[14] https://www.kaggle.com/datasets/frabbisw/facial-age

[15] https://www.kaggle.com/code/shawon10/ck-facial-expression-detection/notebook

[16] https://towardsdatascience.com/image-augmentation-14a0aafd0498

[17] https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html

[18] https://towardsdatascience.com/transfer-learning-on-greyscale-images-how-to-fine-tune-pret

[19] https://github.com/ChibaniMohamed/cnn_age_gender/blob/main/train.py

[20] https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

[21] https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html
[22] https://www.tensorflow.org/api_docs/python/tf/data/Dataset#attributes_1
[23] https://towardsdatascience.com/how-to-use-dataset-in-tensorflow-c758ef9e4428
[24] https://towardsdatascience.com/convolutional-neural-networks-explained-9cc5188c4939
[25] https://www.kaggle.com/code/dansbecker/running-kaggle-kernels-with-a-gpu/notebook