

# Towards Continuous Health Diagnosis from Faces with Deep Learning

Victor Martin<sup>1,2</sup>[0000–0002–8619–6897], Renaud Séguier<sup>1</sup>, Aurélie Porcheron<sup>2</sup>, and Frédérique Morizot<sup>2</sup>

<sup>1</sup> CentraleSupélec, Avenue de la Boulaie, 35510 Cesson-Sévigné, France

<sup>2</sup> CHANEL Parfums Beauté, 8 Rue du Cheval Blanc, 93500 Pantin, France

Corresponding author: Victor Martin <[victor@vmartin.fr](mailto:victor@vmartin.fr)>

**Abstract.** Recent studies show that health perception from faces by humans is a good predictor of good health and healthy behaviors. We aimed to automatize human health perception by training a Convolutional Neural Network on a related task (age estimation) combined with a Ridge Regression to rate faces. Indeed, contrary to health ratings, large datasets with labels of biological age exist. The results show that our system outperforms average human judgments for health. The system could be used on a daily basis to detect early signs of sickness or a declining state. We are convinced that such a system will contribute to more extensively explore the use of holistic, fast, and non-invasive measures to improve the speed of diagnosis.

**Keywords:** Health Estimation, Non-Invasive Diagnosis, Convolutional Neural Network, Facial Features

## 1 Introduction

Judgments of a person’s health based on facial appearance are a daily occurrence in social interactions. Understanding how we perceive health from a face is important because this judgment drive a wide array of social behaviors. Looking healthy has many positive real-life outcomes such as preferential treatment in the professional context, in the justice system or in dating interactions [1,2,3,4]. Inversely, looking unhealthy is associated to lower self-esteem [5] and may lead to a risk of social stigmatization and isolation [6]. A better understanding of how health is perceived and which facial cues alter this perception is likely to help reducing the negative social consequences which can follow.

Scientific recent evidences also show that facial healthy appearance is a good predictor of healthy behaviors [7] and good health [8,9,10]. Faces with an increase of oxygenated blood skin coloration are perceived healthier, and blood oxygenation level is known to be associated with cardiovascular fitness [10]. People with a healthy diet, such as daily consumption of fruits and vegetables, have a more attractive skin color and are perceived healthier [7]. Sleep deprived people appear less healthy compared with when they are well rested [11]. And people would acutely detect signs of sickness from the face in an early phase after exposure to infectious stimuli and potentially contagious people [12]. Figure 1 shows two average faces of people perceived in good health and people perceived in bad health. As health perception and age are known to be correlated [13], health ratings are decorrelated with age.

We aim to develop an automatic system able to imitate human judgments of health. Such a



**Fig. 1.** Average face of the 10 faces with the greatest perceived health to the left, and with the lowest perceived health at right. Health ratings have been decorrelated from age.

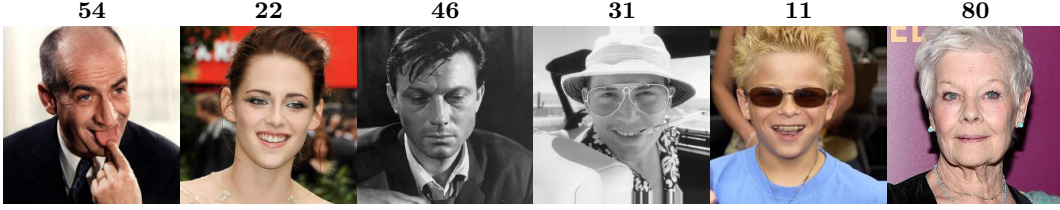
technology, when used over the long term, could enable fast and non-invasive detection of a declining state of a person. That’s why we introduce the first system able to estimate health scores from faces.

To the best of our knowledge, we introduce in this paper the first work on automatic health estimation from face. Lots of works have been made to estimate age from faces [14,15,16,17]. More recently, some researchers have begun to study whether it is possible to estimate less common attributes from the face such as intelligence [18], attractiveness [19,20,21,22] or social relation traits [23].

In view of the current state of art and our constraints, we use a Convolutional Neural Network trained on biological age combined with a Ridge Regression to assess health perception from faces (Sec. 2). Thereafter, we evaluate the system performance on our database and we compare it with human performance on the same database (Sec. 3).

## 2 Health Estimation

Based on the age estimation method of [17], we employ the Convolutional Neural Network VGG-16 pre-trained on the ImageNet database [24] to detect 1,000 classes of objects, and trained it on the Internet Movie Database (IMDb) of celebrities (Fig. 2). We filtered the  $\approx 500K$  images to keep only those containing faces with resolution greater than  $120 \times 120$  pixels, no more than one face detected in each image, and only picture depicting people from 11 to 85 years old. For each picture, we have the date of birth of the celebrity pictured and the date

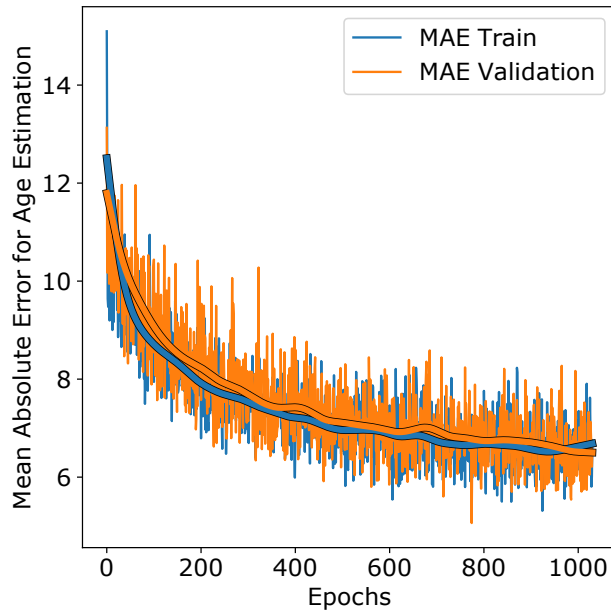


**Fig. 2.** An excerpt of the Internet Movie Database with their corresponding biological age. As we see it above, the database contains faces with large variations in pose, illumination and color distribution. Pictures are resized to 224x224 before training.

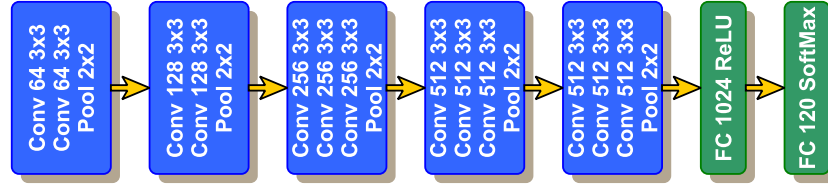
of the photo acquisition, thus we can deduce the biological age of the depicted person.

In addition, from the original VGG-16 architecture, we replace the final Multi Layer Perceptron containing a large part of the parameters, by a lighter one with one layer of 1024 units (Fig. 4) and an output layer of 120 units. The objective of doing so is to shift the learning effort onto the convolutional layers because the final Multi Layer Perceptron will be dropped as we want to estimate health and not biological age – thus, having the fastest training with the lowest score is not the main goal here.

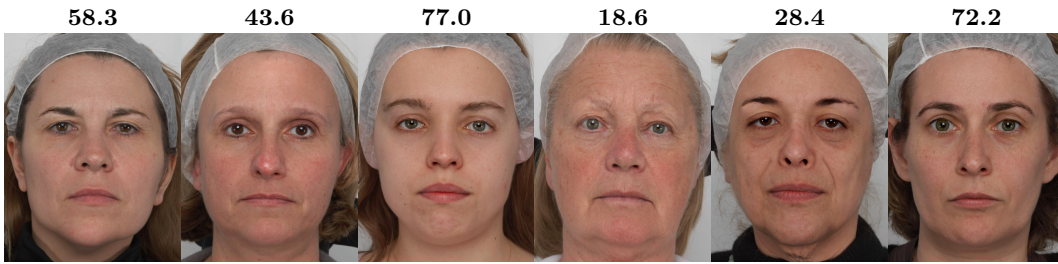
Thus, the last 3 convolutional blocks and the fully connected layers has been trained on IMDb with Stochastic Gradient Descent with a Learning Rate of  $10^{-4}$  on 1000 epochs with 10 steps per epoch and a batch size of 16. The decrease of the Mean Absolute Error for the training set and validation set can be seen in Figure 3.



**Fig. 3.** Decrease of the Mean Absolute Error during the training for the train set and the validation set.



**Fig. 4.** Our architecture takes a 224x224 image and produces a probability distribution over all possible ages. The blue part has not been modified from the original VGG-16 architecture.

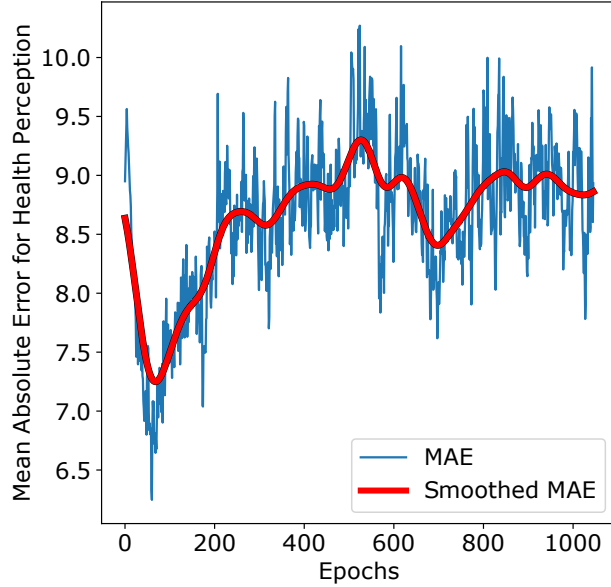


**Fig. 5.** An excerpt of our database with their corresponding perceived health scores. Our database contains 140 photos of women faces with a neutral expression in a controlled environment.

After that, we have to develop our system of health estimation with only 140 images annotated with health scores (Fig. 5). We want to compute a representation of our faces from the newly trained ConvNet using only the convolutions and pooling blocks, and use a regression to estimate health scores from representations. The question remains, at which epoch can we stop the training for health estimation? If we take the weights at an early epoch, the system will be underfitted. In the same way, as we do not want to predict biological age, taking the weights corresponding to an advanced epoch with a low MAE is not the go-to choice to make. We evaluate the suitability of ConvNet weights at each epoch for Health Estimation with a simple Linear Regression trained with a 40-fold configuration. We can see in Fig. 6 how the training on a different, but related, task can increase performance on our health estimation problem. At epoch 0, learning for biological age hasn't started yet and we get a relatively high MAE (9.0). In a second stage, learning for biological age greatly decreases Mean Absolute Error from 9.0 to 6.2. Finally, as learning progresses and the model specializes in biological age estimation, the error increases. An optimal period is found around epoch 60 to take the weights for health estimation.

Now that we found the ConvNet weights to compute representations from faces, we test several estimators to assess health scores from representations. For each estimator, we evaluate a broad range of parameters and report those producing the best performance in Table 1. In the table, the Multi Layer Perceptron is composed of two layers containing  $n$  neurons for the first layer and 120 for the output layer.





**Fig. 6.** Variation of the Mean Absolute Error in function of the epoch at which the weights are chosen. Epoch 0 corresponds to VGG-16 just trained on ImageNet. The red curve has been Gaussian smoothed with  $\sigma = 25$ .

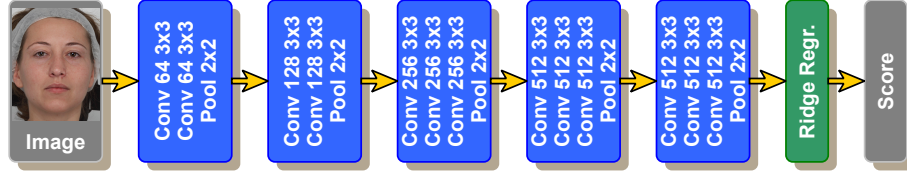
**Table 1.** List of tested estimators. The estimator with the lowest Mean Absolute Error is bolded.

Estimator	MAE
Linear Regression	6.240
<b>Ridge <math>\alpha = 10^{-3}</math></b>	<b>6.232</b>
Lasso $\alpha = 10^{-2}$	6.437
Linear SVR $C = 10^3$	6.355
RBF SVR $C = 10^4, \gamma = 10^{-4}$	6.269
PLS Regression $n = 100$	14.64
Multi Layer Perceptron $n = 2048$	8.543
Extremely Randomized Trees $n = 200$	8.446
K-NN $K = 15$	8.778

As we can see on Table 1, simple estimators as a Linear Regression or a Linear Regression regularized with a low  $\ell_2$  penalty (Ridge Regression) can achieve the best performance given our dataset and the feature extraction method we chose earlier. We can explain the fact that simpler estimators perform better than more complex estimators as Random Forests or Multi Layer Perceptron by the scarce number of samples  $n = 140$  in regard of the dimensionality of our features  $d = 512 * 7 * 7 = 25088$ . The final architecture of our system is described in Figure 7.

### 3 Experiment: System versus Human performance

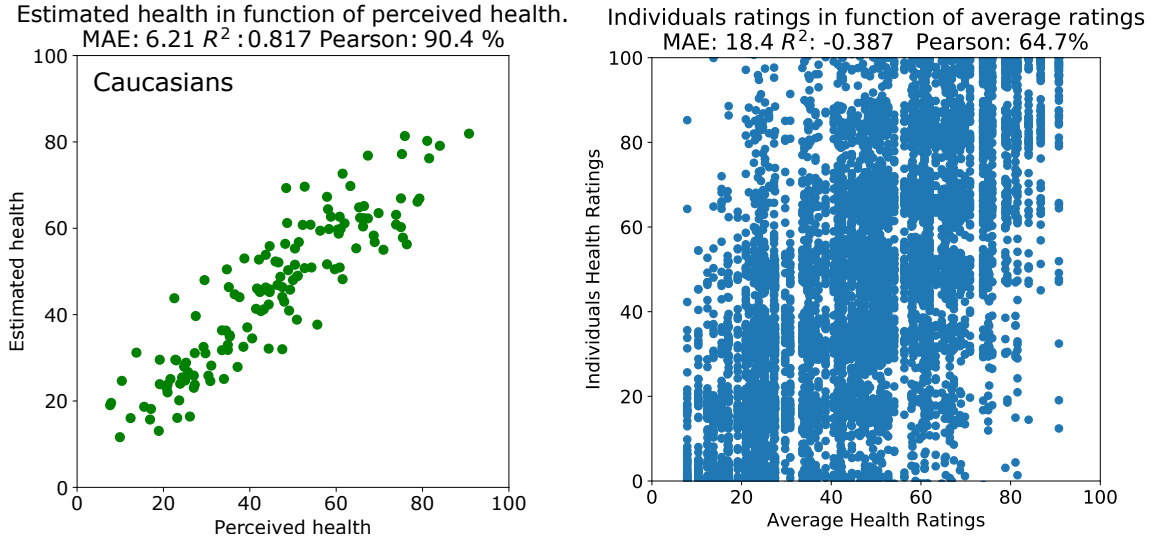
We have 140 images of faces and each of them had been rated by 74 judges. For every picture, we asked them to evaluate health and to give a score from 0 to 100; 0 being perceived in very



**Fig. 7.** The whole computation chain. The blue part and the green part are trained separately on different datasets.

bad health and 100 being perceived in very good health. Finally, for each image, we took the average of the 74 ratings to determine a reliable perceived health score. In this database, the health scores obtained are 60% correlated with biological ages.

Exploiting the previously described system, we trained the Ridge Regression in a 140-fold manner to assess its performance.



**Fig. 8.** Left: The predictions of our system compared to the perceived health scores, which is the average health ratings from humans. Right: all individual ratings from humans in function of average ratings. This 2<sup>nd</sup> graph shows the relatively high variance of human ratings for each image.

As we can see on Figure 8, we can achieve good performance on our dataset with a scarce amount of data. Using mean absolute error MAE, coefficient of determination  $R^2$  and Pearson correlation PC, Table 2 shows that our system estimates health more accurately than an average human working on the same dataset.

In addition, among the 74 judges, one judge with the lowest MAE (i.e smallest difference in average between his ratings and the average ratings) is selected and placed in the table below under the name Best Human.

**Table 2.** Performance of our health estimation system compared to human performance.

	MAE	PC	R <sup>2</sup>
<b>System</b>	<b>6.21</b>	<b>90.4%</b>	<b>0.817</b>
Average Human	18.4	64.7%	-0.387
Best Human	9.37	81.3%	0.637

As an additional note, we can observe that health scores are 60% correlated with biological ages, and health estimates outputted by our system are 90% correlated with health scores. Hence, we confirm that our system estimates health from faces, and not just biological age.

## 4 Conclusion

This paper describes how we manage to develop an automatic system able to imitate human judgments of health. We trained a Convolutional Neural Network to estimate biological age and we used representations produced by the network of our scarce database to train a simpler estimator. We observed a very good performance of the system when we compared it to human judgments of health.

Nevertheless, we identified several areas of improvement.

First, the use of a Linear Regression to rank the different ConvNet weights (Fig. 6) tends to favor this type of estimators in the next step where we compare the performance of different estimators (Tab. 1). We could have ranked the different weights using a multitude of estimators.

Moreover, by using more images annotated with health ratings, we could improve the performance of our system and make it more robust to variations in pose and illumination.

Additional work will be necessary to test its performance on other demographic groups such as other ethnicities and men.

To conclude, we developed the first automatic health estimation system able to reproduce human judgments. Such a system could be used in institutions such as hospitals or retirement homes to automatically predict a potential future sickness from earlier visual signs present in a face. Similarly, it could be used for the remote monitoring of patients, to detect a sudden drop in health perception and prevent behaviors that negatively impact health.

## References

1. Efran, M.G.: The effect of physical appearance on the judgment of guilt, interpersonal attraction, and severity of recommended punishment in a simulated jury task. *Journal of Research in Personality* **8**(1) (1974) bibtex: efran\_effect\_1974.
2. Marlowe, C.M., Schneider, S.L., Nelson, C.E.: Gender and attractiveness biases in hiring decisions: Are more experienced managers less biased? *Journal of Applied Psychology*, 81(1). 11–21 (1996) bibtex: marlowe\_gender\_1996.
3. Ritts, V., Patterson, M.L., Tubbs, M.E.: Expectations, impressions, and judgments of physically attractive students: A review. *Review of Educational Research* **62**(4) (1992) bibtex: ritts\_expectations\_1992.

4. Spisak, B.R., Blaker, N.M., Lefevre, C.E., Moore, F.R., Krebbers, K.F.B.: A face for all seasons: Searching for context-specific leadership traits and discovering a general preference for perceived health. *Frontiers in Human Neuroscience* **8**(792.) (2014) bibtex: spisak\_face\_2014.
5. Feingold, A.: Good-looking people are not what we think. *Psychological Bulletin* **111**(2) (1992) bibtex: feingold\_good-looking\_1992.
6. Henderson, A.J., Holzleitner, I.J., Talamas, S.N., Perrett, D.I.: Perception of health from facial cues. *Philosophical Transactions of the Royal Society B: Biological Sciences* **371**(1693) (2016) 20150380. bibtex: henderson\_perception\_2016.
7. Whitehead, R.D., Re, D., Xiao, D., Ozakinci, G., Perrett, D.I.: You Are What You Eat: Within-Subject Increases in Fruit and Vegetable Consumption Confer Beneficial Skin-Color Changes. *PLOS ONE* **7**(3) (March 2012) e32988
8. Zebrowitz, L.A., Franklin, R.G., Boshyan, J., Luevano, V., Agrigoroaei, S., Milosavljevic, B., Lachman, M.E.: Older and younger adults' accuracy in discerning health and competence in older and younger faces. *Psychology and Aging* **29**(3) (2014) bibtex: zebrowitz\_older\_2014.
9. Stephen, I.D., Coetzee, V., Smith, L., M., Perrett, D.I.: Skin blood perfusion and oxygenation colour affect perceived human health. *PLoS ONE* **4**(4) (2009) bibtex: stephen\_skin\_2009.
10. Re, D.E., Whitehead, R.D., Xiao, D., Perrett, D.I.: Oxygenated-Blood Colour Change Thresholds for Perceived Facial Redness, Health, and Attractiveness. *PLoS ONE* **6**(3) (March 2011)
11. Axelsson, J., Sundelin, T., Ingre, M., Someren, E.J.W.V., Olsson, A., Lekander, M.: Beauty sleep: experimental study on the perceived health and attractiveness of sleep deprived people. *BMJ* **341** (December 2010) c6614
12. Axelsson, J., Sundelin, T., Axelsson, C., Lasselin, J., Lekander, M.: Identification of acutely sick people and facial cues of sickness. *Brain, Behavior, and Immunity* **66** (November 2017) e38
13. Fink, B., Matts, P., D'Emiliano, D., Bunse, L., Weege, B., Röder, S.: Colour homogeneity and visual perception of age, health and attractiveness of male facial skin: Perception of male skin colour. *Journal of the European Academy of Dermatology and Venereology* (November 2011) no-no bibtex: fink\_colour\_2011.
14. Lanitis, A., Taylor, C.J., Cootes, T.F.: Toward automatic simulation of aging effects on face images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(4) (2002) 442–455 bibtex: lanitis\_toward\_2002.
15. Lanitis, A., Draganova, C., Christodoulou, C.: Comparing Different Classifiers for Automatic Age Estimation. *IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics : a publication of the IEEE Systems, Man, and Cybernetics Society* **34** (March 2004) 621–8 bibtex: lanitis\_comparing\_2004.
16. Guo, G., Mu, G., Fu, Y., Huang, T.S.: Human age estimation using bio-inspired features. In: *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, IEEE* (2009) 112–119 bibtex: guo\_human\_2009.
17. Rothe, R., Timofte, R., Van Gool, L.: Dex: Deep expectation of apparent age from a single image. In: *Proceedings of the IEEE International Conference on Computer Vision Workshops.* (2015) 10–15 bibtex: rothe\_dex\_2015.
18. Qin, R., Gao, W., Xu, H., Hu, Z.: Modern Physiognomy: An Investigation on Predicting Personality Traits and Intelligence from the Human Face. *arXiv:1604.07499 [cs]* (April 2016) arXiv: 1604.07499 bibtex: qin\_modern\_2016.
19. Fan, Y.Y., Liu, S., Li, B., Guo, Z., Samal, A., Wan, J., Li, S.Z.: Label Distribution-Based Facial Attractiveness Computation by Deep Residual Learning. *IEEE Transactions on Multimedia* **PP**(99) (2017) 1–1 bibtex: fan\_label\_2017.
20. Chen, F., Zhang, D.: Combining a causal effect criterion for evaluation of facial attractiveness models. *Neurocomputing* **177** (February 2016) 98–109 bibtex: chen\_combining\_2016.
21. Liu, S., Fan, Y.Y., Samal, A., Guo, Z.: Advances in computational facial attractiveness methods. *Multi-media Tools and Applications* **75**(23) (December 2016) 16633–16663 bibtex: liu\_advances\_2016.
22. Chen, F., Xiao, X., Zhang, D.: Data-Driven Facial Beauty Analysis: Prediction, Retrieval and Manipulation. *IEEE Transactions on Affective Computing* **PP**(99) (2017) 1–1 bibtex: chen\_data-driven\_2017.
23. Zhang, Z., Luo, P., Loy, C.C., Tang, X.: Learning Social Relation Traits from Face Images. *arXiv:1509.03936 [cs]* (September 2015) arXiv: 1509.03936 bibtex: zhang\_learning\_2015.
24. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014) bibtex: simonyan\_very\_2014.