

Face biometrics as a potential predictor for COVID-19 susceptibility

Oleg Kit

Innovative medical technologies, Lviv, Ukraine

Email: oleg.kit.mail@gmail.com

Abstract

Fighting against the COVID-19 pandemic caused by the SARS-CoV-2 virus is one of the most critical challenges facing the global health system today. The possibility to identify the group of persons in the cohort of people under 50 years old, who are sensitive to the COVID-19 disease by non-invasive methods, is a very perspective approach for estimating the epidemiological state of the human population. The study aimed to identify the features of people's faces with COVID-19 that the most correlate with disease severity could serve as one of these approaches. For this aim, 525 photos of patients' faces with different outcomes of COVID-19 disease were analyzed using the Dlib face recognition convolutional neural network pre-trained for face recognition. Face descriptor vectors were obtained using the convolutional neural network. Facial features were found that predict a person's sensitivity to the SARS-CoV-2 virus (disease severity), and the contribution of each of the features to the risk of developing a severe form of COVID in a person was found. The accuracy of the binary classification of the individual severity of the COVID-19 course using the k-nearest neighbors algorithm on the test dataset was accuracy - 84%, AUC - 0.90.

Introduction

The disease COVID-19, caused by SARS-CoV-2 virus, is one of the most pressing challenges of modern medicine [1]. To date, the SARS-CoV-2 virus has infected about 160 million people and caused 3 million deaths [2]. In most cases, the severe course of COVID-19 with a fatal outcome is observed in older people [3]. In total, young and middle-aged people fall ill as a result of infection with the SARS-CoV-2 virus in the same percentage as the elderly (over 50 years old) [1], but in most cases, COVID-19 disease occurs in them in a milder form [1]; however, mortality among young and middle-aged patients is 26% [3]. Due to new mutations of the SARS-CoV-2 virus, the COVID-19 disease affects an increasingly young age group with a more severe course of the disease [4]. Identifying the risk group among young and middle-aged people susceptible to COVID-19 is an important prerequisite for tackling the pandemic. Modern methods of analyzing photographs of a face in order to determine the phenotypic features of various diseases are increasingly used in the diagnosis of such diseases as cardiovascular diseases [5], hypertension [6], and cancer [7].

Methods

Our study used randomly selected photographs of the faces of young, middle-aged, and older people (from 18 to 102 years old) of different sexes (264 women and 261 men) and different ethnic groups. 95% of the dataset was made up of photographs of faces of Caucasians who fell ill with COVID-19 with positive and negative outcomes. Among them were 261 with negative and 264 with positive anamneses taken from the freely available Internet resources [8, 9, 10]. The obtained dataset contains 525 photos. All photographs of the patients' faces were processed at the initial stage using the pre-trained convolutional neural network Dlib face recognition from the Deepface library [11, 12] to extract face descriptor vectors. Pre-trained neural network (modification of ResNet-34) has computed 128-dimensional face descriptor vectors [11]. After converting patients' faces to face descriptor vectors in MATLAB, using chi-square tests and the MRMR algorithm, six facial features 117, 100, 52, 92, 24, and 67 were selected that were most correlated with the outcome COVID-19 disease. After interpreting these features, it was found that the

values of 24 and 67 features are most correlated with the age and sex of the patient, respectively. For 52 and 92 features, the PCA method yielded a single generalized feature, the value of which most correlated with the patient's ethnicity and disease outcome (the highest values of this feature were in black and asians). Using the six obtained features, the k-nearest neighbors algorithm was trained for the binary classification problem. The classification accuracy on the test dataset was 84%, AUC - 0.90. These values characterize the good performance of this classifier in Fig. 2.

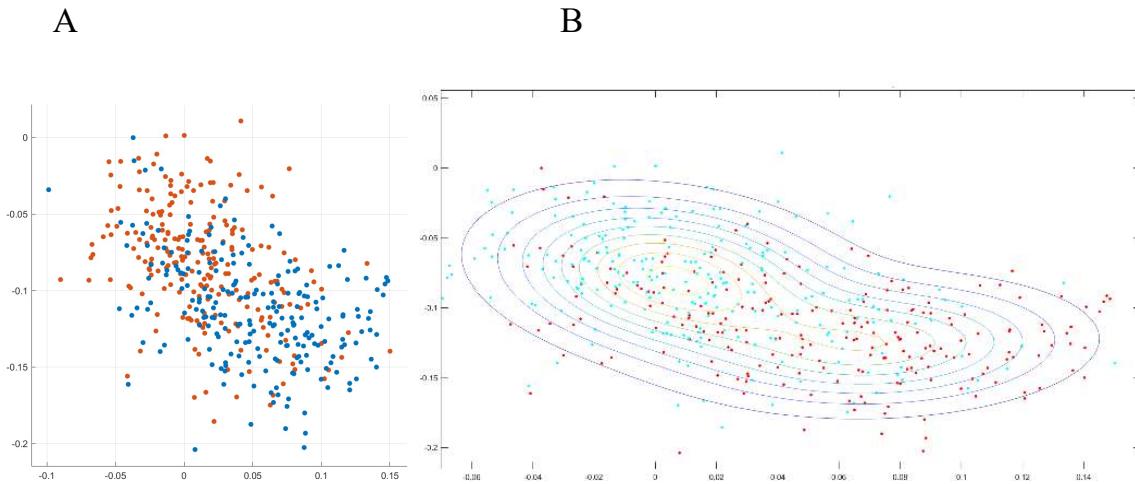


Fig.1 The distribution of disease outcomes by 24 and 67 features. A - Dependence of disease outcome on the values of 24 feature (X-axis) and 67 feature (Y-axis). Blue represents positive disease outcome, and red represents negative disease outcome. B - The distribution of COVID-19 cases with different disease outcomes forms two overlapping Gaussian.

Results and Discussion

Determination of phenotypic features associated with the anatomical features of the face and the search for their correlations with susceptibility (severity of the course of the disease) to the SARS-CoV-2 virus is an essential task for modern data analysis. The use of methods for analyzing biomedical data (in particular, facial images) makes it possible to find phenotypic features that are reliably associated with the individual severity of the course of the COVID-19 disease. As a result of the analysis of photographs of the faces of people with COVID-19, three features were found, listed in the order of their influence on the risk of negative disease outcome: age, ethnicity, and sex of the patient. The dependence of disease outcome on age (24 feature) and sex (67 feature) is shown in Fig. 1. This figure shows that the distribution of COVID-19 cases with different disease outcomes forms two overlapping Gaussian. If they take photographs of the face of the same patient at different ages, then the position of the point is shifted along the isolines in the space of these two features. Epidemiological data support our results that the severity of COVID significantly depends on the patient's age, sex, and ethnicity [3]. The risk of negative disease outcome strongly correlates with Asian ethnicity has been the research subject since the beginning of 2020. In particular, the level of ACE2 expression in the lung tissues of Asians [13] and the found genetic predisposition to coronaviruses in Asians [14] are mentioned. In Supplementary Materials, attached a dataset of 525 embeddings of COVID-19 patients to verify the results indicated here.

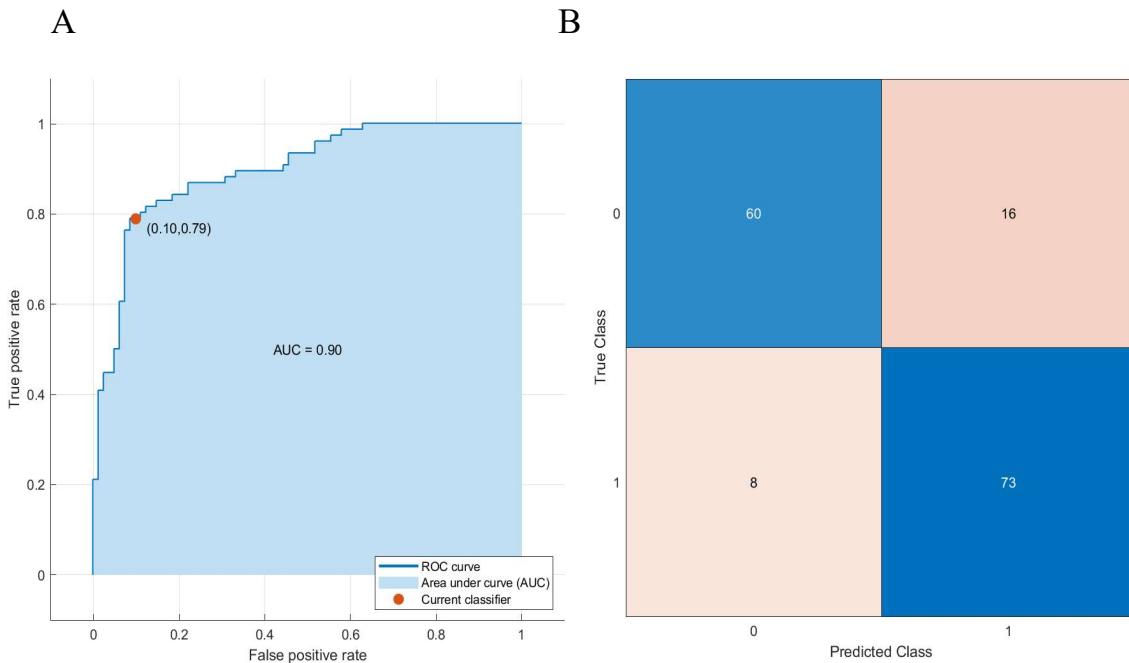


Fig.2 A) ROC-AUC curve for KNN classifier on a test B) Confusion matrix

Conclusions.

A method was developed for assessing the individual sensitivity of patients of different ages, sex, and ethnicity to the disease COVID-19. A classifier with a high predictive accuracy of disease outcome was built using the patient's facial features.

Supplementary Materials

This link can be used to download the dataset.

<https://drive.google.com/file/d/1Jwh3kMUBeHdYXwVRdilvZ5mpWh1PXLK-/view?usp=sharing>

References

1. Hu, B., Guo, H., Zhou, P. *et al.* Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol* **19**, 141–154 (2021). <https://doi.org/10.1038/s41579-020-00459-7>
2. Coronavirus cases:. (n.d.). Retrieved May 11, 2021, from <https://www.worldometers.info/coronavirus/>
3. Age, sex, existing conditions of covid-19 cases and deaths. (n.d.). Retrieved May 11, 2021, from <https://www.worldometers.info/coronavirus/coronavirus-age-sex-demographics/>
4. Thiagarajan K. Why is India having a covid-19 surge? *BMJ* 2021; 373 :n1124 doi:10.1136/bmj.n1124
5. Christos P Kotanidis, Charalambos Antoniades, Selfies in cardiovascular medicine: welcome to a new era of medical diagnostics, *European Heart Journal*, Volume 41, Issue 46, 7 December 2020, Pages 4412–4414, <https://doi.org/10.1093/eurheartj/ehaa608>

6. Ang L, Yim MH, Do J-H, Lee S. A Novel Method in Predicting Hypertension Using Facial Images. *Applied Sciences*. 2021; 11(5):2414. <https://doi.org/10.3390/app11052414>
7. Liang B, Yang N, He G, Huang P, Yang Y, Identification of the Facial Features of Patients With Cancer: A Deep Learning-Based Pilot Study, *J Med Internet Res* 2020;22(4):e17234 <https://doi.org/10.2196/17234>
8. Yee, L. (2021). Stars Who Have Tested Positive for Coronavirus (Photos). Retrieved 11 May 2021, from <https://www.thewrap.com/stars-tested-positive-coronavirus-photos/>
9. Wurzburger, A. (2021) Celebrities We Lost to COVID-19. Retrieved 11 May 2021, from <https://people.com/health/celebrity-coronavirus-deaths>
10. Memory list (in russian). Retrieved 11 May 2021, from <https://www.sites.google.com/view/covid-memory/home>
11. serengil/deepface. (2021). Retrieved 11 May 2021, from <https://github.com/serengil/deepface>
12. davisking/dlib-models. (2021). Retrieved 11 May 2021, from <https://github.com/davisking/dlib-models>
13. Zhao, Y., Zhao, Z., Wang, Y., Zhou, Y., Ma, Y., & Zuo, W. (2020, January 1). Single-cell RNA expression profiling of ACE2, the putative receptor of Wuhan 2019-nCov. *bioRxiv*. <https://www.biorxiv.org/content/10.1101/2020.01.26.919985v1.full>
14. Souilmi, Y., Lauterbur, M. E., Tobler, R., Huber, C. D., Johar, A. S. (2021, January 1). An ancient viral epidemic involving host coronavirus interacting genes more than 20,000 years ago in East Asia. *bioRxiv*. <https://www.biorxiv.org/content/10.1101/2020.11.16.385401v2>.