UDC 613.6.02, 616-036.22, 004.8
DOI: 10.21668/health.risk/2024.2.11.eng

Read online

Research article

USING INDUCTIVE MACHINE LEARNING TO IDENTIFY RISK FACTORS FOR HEALTHCARE WORKERS TO GET INFECTED WITH HIGHLY CONTAGIOUS VIRUSES (BASED ON COVID-19 MODEL)

I.A. Egorov[1], **S.S. Smirnova**[1,2], **A.V. Semenov**[1,2]

[1]Federal Scientific Research Institute of Viral Infections «Virome», 23 Letnyaya St., Ekaterinburg, 620030, Russian Federation
[2]Ural State Medical University, 3 Repina St., Ekaterinburg, 620028, Russian Federation

Epidemic and pandemic spread of highly contagious viruses (SARS-CoV, influenza A virus, Ebola virus, MERS-CoV, and SARS-CoV-2) has been a trend observed in the first two decades of the 21st century.

The predominant impact made by the biological occupational factor on healthcare workers determines high occupational risk of infection, a severe disease course and a fatal outcome. Epidemiological data mining based on machine learning algorithms is successfully used in epidemiological practice to identify factors (predictors) contributing to infection in various risk populations.

In this study, the database generated from a survey of 1312 healthcare workers was analyzed intelligently. A total of 6912 machine learning models were implemented. SARS-CoV-2 infection was found to be facilitated by providing medical care to a COVID-19 patient, using a full set of PPE after direct contact with a COVID-19 patient, direct contact with items in the external (hospital) environment, vaccination against COVID-19 after direct contact with a COVID-19 patient, acting as nursing staff (cleaners) and being present during aerosol-generating procedures.

The study identified four groups of predictors determining SARS-CoV-2 infection in healthcare workers: contact with a COVID-19 patient and environmental items, PPE quality and complexity, occupational affiliation of healthcare workers and their BMI values. One predictor was found in 56.2 % of healthcare workers; two, in 19.2 %; three, in 16.4 %; four, in 5.5 %; and five predictors, in 2.7 %.

Thus, epidemiological data mining is a modern stage in epidemiological analysis. The use of machine learning methods allows for multifactorial assessment of SARS-CoV-2 infection risks in healthcare workers and enables identifying and reliably estimating the most significant predictors. Intelligent data analysis has flexible architecture, which allows adjusting the model under study and supplementing new data to the existing database, detecting changes in an epidemiological situation and accomplishing relevant preventive and anti-epidemic activities.

Keywords: *data mining, artificial intelligence, machine learning, risk-based approach, occupational predictors of infection, highly contagious viruses, SARS-CoV-2, healthcare workers.*

Epidemic and pandemic spread of highly contagious viral pathogens has become a prominent trend in the first two decades of the 21st century. The beginning of it (2002) was marked with an outbreak of severe acute respiratory syndrome (SARS). The World Health

Ivan A. Egorov – Researcher at Ural-Siberian Research and Methodology Center for Prevention of Healthcare-Associated Infections (e-mail: egorov_ia@niivirom.ru; tel.: +7 (922) 127-66-88; ORCID: https://orcid.org/0000-0002-7153-2827).

Svetlana S. Smirnova – Candidate of Medical Sciences, Leading Researcher, Head of the Ural-Siberian Research and Methodology Center for Prevention of Healthcare-Associated Infections; Associate Professor of the Department of Epidemiology, Social Hygiene and Organization of Sanitary and Epidemiological Service (e-mail: smirnova_ss@niivirom.ru; tel.: +7 (908) 917-59-86; ORCID: https://orcid.org/0000-0002-9749-4611).

Aleksandr V. Semenov – Doctor of Biological Sciences, director; Professor of the Department of Medical Microbiology and Clinical Laboratory Diagnostics (e-mail: semenov_av@niivirom.ru; tel.: +7 (343) 261-99-47; ORCID: https://orcid.org/0000-0003-3223-8219).

Organization (WHO) declared the type A (H1N1)[1] influenza pandemic between 2009 and 2010. A major outbreak of Ebola virus disease (EVD) occurred in Western Africa in 2014–2016. In 2015 there was an outbreak of coronavirus-induced Middle East Respiratory Syndrome (MERS) [1, 2]. The COVID-19 pandemic has undoubtedly become one of the greatest global evens over two past decades [3].

Epidemic and pandemic spread of highly contagious viruses always involves a greater burden on the healthcare system [4]. The predominant impact made by the biological occupational factor on healthcare workers determines their high occupational risk of infection with highly contagious viruses, a severe disease course and a fatal outcome [5]. Various studies report rather high shares of healthcare workers among diseased during epidemics and pandemics reaching 30 % (SARS, 21.1 % [6]; influenza A virus (H1N1), 27.1–30.0 % [7]; EVD, 8.0 % [8]; MERS, 18.7 % [9]; COVID-19, 9.0–26.0 % [10]).

Artificial intelligence technologies are successfully used to predict incidence rates during outbreaks of various viral infections including Ebola virus disease [11], viral hepatitis and pneumonia, and type A influenza [12, 13]. During the pandemic and post-pandemic periods (COVID-19), intelligent data analysis made it possible to solve major epidemiological tasks. They include identification of COVID-19 infection risk territories, groups and factors (predictors); incidence forecasts and assessment of COVID-19 prevention effectiveness; forecasts of virus mutations; assessment of lung lesion severity; differential diagnostics based on instrumental patient examinations; modeling of molecular interactions typical for the SARS-CoV-2 virus [14–16]. In addition to that, practical use of machine learning algorithms helped identify risk factors able to cause SARS-CoV-2 infection among healthcare workers [17, 18]; establish priority occupational groups of healthcare workers for molecular-genetic examinations aimed at detecting SARS-CoV-2 RNA and their isolation [18, 19]; predict likelihood of SARS-CoV-2 infection by intelligent analysis of data collected from devices worn by healthcare workers [20].

Use of machine learning techniques to identify predictors of infection with highly contagious viruses in healthcare workers is a modern stage in the epidemiological analysis. It helps implement a risk-based approach to infection prevention not only under an already present epidemic or pandemic spread of known pathogens but also for potential epidemic threats.

The aim of this study was to build machine learning models based on data collected by a survey accomplished among healthcare workers and to identify predictors of infection with highly contagious viruses in this occupational group (using the COVID-19 model).

Materials and methods. The study was accomplished by the Urals-Siberia Scientific Methodical Center for Prevention of Healthcare-Associated Infections and the Federal Scientific Research Institute of Viral Infections “Virome” of the Federal Service for Surveillance over Consumer Rights Protection and Human Wellbeing. The study was approved by the local ethics committee of Rospotrebnadzor’s State Scientific Center for Virology and Biotechnology “Vector”, the meeting report No. 3 dated June 24, 2022 (the name of the institution was valid at the moment the study was approved; it was later changed in accordance with Rospotrebnadzor Order No. 599 issued on November 11, 2022).

A survey was conducted during the pandemic (2020–2021) in a large industrial city; overall, 1312 healthcare workers participated in it. The survey involved filling in a paper original non-personal questionnaire “Identification of occupational and non-occupational

[1] Influenza A (H1N1). pandemic 2009–2010. *WHO*. Available at: https://www.who.int/emergencies/situations/influenza-a-(h1n1)-outbreak (January 18, 2024).

factors influencing risks of SARS-CoV-2 infection for healthcare workers" developed by the authors[2]. The questionnaire included both open and closed questions and was divided into six subject items: sex and anthropometric parameters (height, weight), sex and age profile, occupation, COVID-19 infection risks, commitment to observing specific and non-specific prevention of SARS-CoV-2 infection, and circumstances of COVID-19 detection. Healthcare workers were included into the study only after providing personal voluntary consent to it.

Data taken from each paper filled-in questionnaire were put manually into an electronic table in Microsoft Excel (*.xlxs). An initial database contained 1312 lines according to the number of the respondents (including 366 lines for healthcare workers who got infected with COVID-19 and 946 COVID-19-intact healthcare workers) and 45 columns matched with the questions in the questionnaire. One column represented a dependent (target) variable where the value '1' meant a respondent got infected with the coronavirus infection and the value '0' meant they were COVID-19-intact.

Questionnaires filled in by administrative staff and questionnaires with data defects were excluded from the database at the preliminary stage in data analysis. Additionally, each questionnaire of a COVID-19 infected healthcare worker was matched with a questionnaire of a COVID-19-intact healthcare worker, both questionnaires being comparable as per the analyzed parameters. It was done to remove imbalance in classes of the dependent (target) variable. Body mass index (BMI) was calculated relying on the respondents' height and weight. BMI values were interpreted in accordance with the World Health Organization recommendations[3].

The ready-for-analysis database included 688 lines (questionnaires) and 28 columns containing non-personal data; 27 columns represented the analyzed predictors and 1 column represented the dependent (target) variable.

Predictors for training of machine learning models were selected by identifying an association between each predictor of the dependent (target) variable with calculating Pearson's χ^2 test. Overall, 22 predictors were selected for which the highest dependence was determined ($p < 0.05$).

The ultimate database was divided into a learning (2/3, $n = 460$) and test (1/3, $n = 228$) sub-sample. Data mining was accomplished by using five machine learning algorithms eligible for classification: extremely randomized trees, decision trees, random forest, logistic regression, and extreme gradient boosting. Original author settings were used for all five algorithms. The machine learning algorithms were reproduced in the Jupyter Notebook (v.6.0.0) interactive platform using Python (v.3.7.16). Data were preliminary processed and analyzed using *pandas*; mathematical and numeric operations were performed using *numpy*. We used *scikit-learn* library for machine learning to divide data into the learning and test sub-samples, normalize the data, calculate statistical indicators, build a discrepancy matrix, assort parameters for model training and select predictors. The algorithm functional was implemented using libraries with an open source code. The results were visualized using *matplotlib* and *seaborn* libraries; the visualization functional, *SHapley Additive exPlanations* libraries[4].

Statistical performance indicators obtained by machine learning models were interpreted by creating ROC-curves and computing ROC-AUC (area under the curve) with its 95 % confidence interval (95 % CI).

---

[2] Anketa dlya meditsinskikh rabotnikov [Questionnaire for healthcare workers]: a Yandex Disk document. Available at: https://disk.yandex.ru/i/nNFNjGaVLs5KDg (March 12, 2024) (in Russian).

[3] Body mass index (BMI). *WHO*. Available at: https://www.who.int/data/gho/data/themes/topics/topic-details/GHO/body-mass-index (January 18, 2024).

[4] Welcome to the SHAP documentation. *SHAP*. Available at: https://shap.readthedocs.io/en/latest/ (January 19, 2024).

Proportions of true-positive, true-negative, false-positive and false-negative forecasts were calculated based on the discrepancy matrix. We considered only models with statistical significance ($p < 0.05$) as well as sufficient sensitivity and specificity (above 60.0 %).

The importance of predictors was determined relying on the F-score, which was calculated using feature importance, a built-in method of the extreme gradient boosting library.

The power of impact exerted by each analyzed predictor on the model result was estimated using SHAP-values considering all possible combinations. Predictors with positive SHAP-values (above 0) were considered as able to determine SARS-CoV-2 infection of healthcare workers. In addition, the analyzed predictors were cauterized with the threshold 90.0 %. Figure 1 provides the general idea of the study design.

Results and discussion. Overall, 6912 machine learning models were trained; of them (Figure 2):

- extremely randomized trees algorithm (sensitivity is 66.0, specificity is 85.6, AUC (area under curve) is 69.9, 95 % CI [62.1–76.9]);
- decision tree algorithm (sensitivity is 66.0, specificity is 77.6, AUC is 73.5, 95 % CI [67.6–79.3]);
- random forest algorithm (sensitivity is 65.0, specificity is 80.8, AUC is 75.1, 95 % CI [68.1–81.5]);
- logistic regression algorithm (sensitivity is 69.9, specificity is 79.2, AUC is 79.4, 95 % CI [73.3– 85.4]);
- extreme gradient boosting algorithm (sensitivity is 70.9, specificity is 80.8, AUC is 80.4, 95 % CI [74.4–85.8]).

We comparatively assessed the statistical performance indicators of the trained models using the analyzed non-personal data set. As a

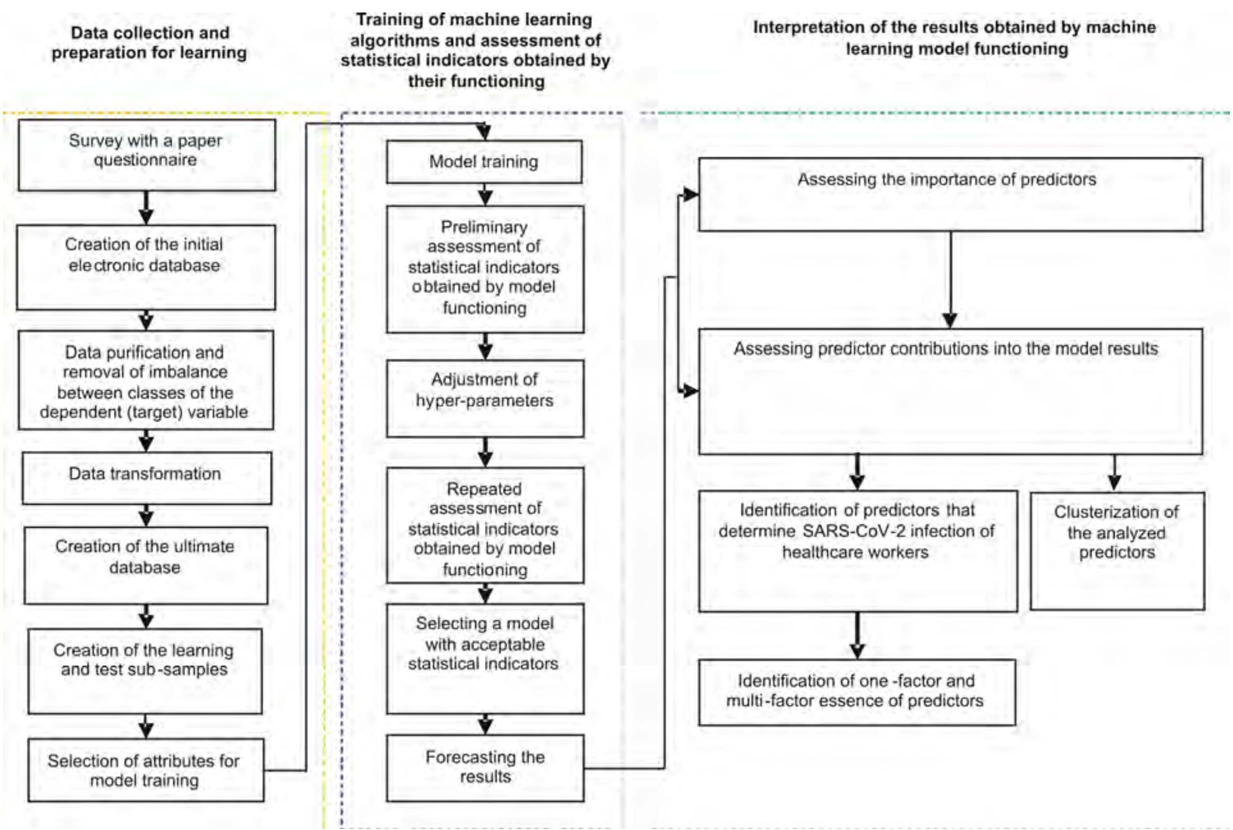

Figure 1. Study design

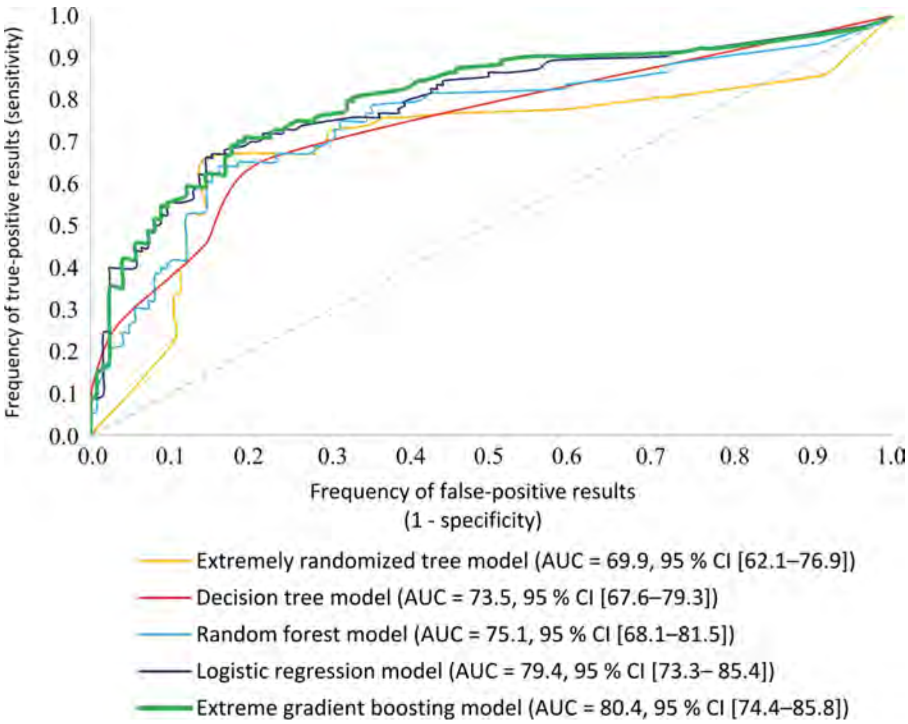

Figure 2. ROC-curves describing the statistical performance indicators of the machine learning algorithms

result, the extreme gradient boosting algorithm was established to have acceptable sensitivity, specificity and the AUC value. This algorithm was applied to identify predictors able to determine SARS-CoV-2 infection in healthcare workers.

The importance of predictors was analyzed using the built-in method of the extreme gradient boosting model (F-score). The analysis revealed 19 predictors (86.4 %) and made it possible to create several rank groups. The highest importance was identified for providing outpatient clinical healthcare, 56.0; COVID-19 infected among people close to a healthcare worker, 46.0; providing healthcare to COVID-19 patients, 44.0; normal body weight as per BMI, 38.0; use of personal protective equipment (PPE) without complete protection provided for the eyes or respiratory organs, 32.0. The second rank group included such predictors as a work shift longer than 24 hours, 29.0; acting as nursing staff, 24.0; vaccination against COVID-19 after a direct contact with a COVID-19 patient, 21.0. The predictors in the third rank group were emergencies involving exposure to patient biomaterials, 19.0; overweight (pre-obesity), 17.0; chronic somatic diseases, 16.0; direct contacts with environmental (hospital) objects, 13.0; use of the full PPE set after a direct contact with a COVID-19 patient and Class 1 obesity, both 12.0. The fourth rank group included the following predictors: acting as doctors and Class 2 obesity, both 7.0; acting as orderlies (cleaners), 6.0; being present during aerosol-generating procedures, 5.0; performing laboratory and pathological anatomy diagnostics, 1.0.

Assessment of the predictor importance per the F-score had certain limitations in our study when interpreting the performance of the analyzed model. Absence of underweight or overweight was identified as a predictor by the model due to high frequency of this attribute in the learning sample (65.4 %, $n = 301$).

A strategy for assessing the power of impact exerted by each predictor that involved calculating its SHAP-value was used at further stages in the research to more precisely

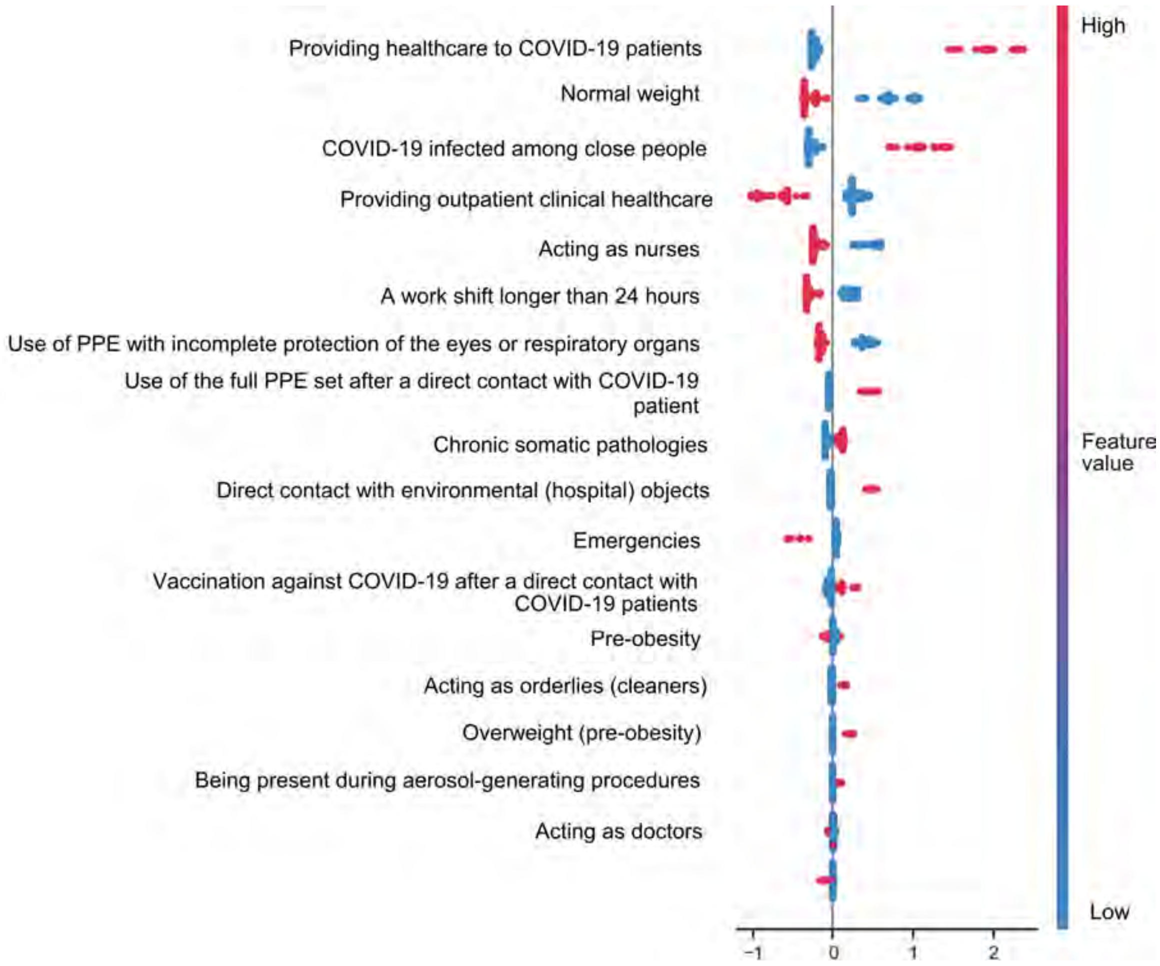

Figure 3. Predictors determining SARS-CoV-2 infection of healthcare workers

estimate the importance of predictors able to determine SARS-CoV-2 infection in healthcare workers. This approach revealed those predictors that determined the infection and made it possible to rank them depending on the SHAP-value.

The power of impact had different intensity for the analyzed predictors, the difference reaching 10.5 times (0.9904–0.0943, $p < 0.05$), including (Figure 3): providing healthcare to COVID-19 patients (2.378 ± 0.791, $p < 0.05$), use of the full PPE set after a direct contact with a COVID-19 patient (0.565 ± 0.17, $p < 0.05$), direct contacts with environmental (hospital) objects (0.547 ± 0.146, $p < 0.05$), vaccination against COVID-19 after a direct contact with a COVID-19 patient (0.304 ± 0.072, $p < 0.05$), acting as orderlies (cleaners) (0.162 ± 0.035, $p < 0.05$), being present during aerosol-generating procedures (0.109 ± 0.022, $p < 0.05$). It is worth noting that several predictors determining SARS-CoV-2 infection were not associated with occupation, for example, COVID-19 infected among people close to a healthcare worker (1.464 ± 0.58, $p < 0.05$), Class 2 obesity (0.259 ± 0.04, $p < 0.05$), chronic somatic diseases (0.148 ± 0.092, $p < 0.05$).

We performed a cluster analysis of predictors determining SARS-CoV-2 infection in healthcare workers. As a result, four different clusters were established (Figure 4):

- Cluster 1 was associated with a direct contact with COVID-19 patients or environmental (hospital) objects around them: providing healthcare to a COVID-19 patient ($p < 0.05$); being present during aerosol-generating procedures ($p < 0.05$); a direct

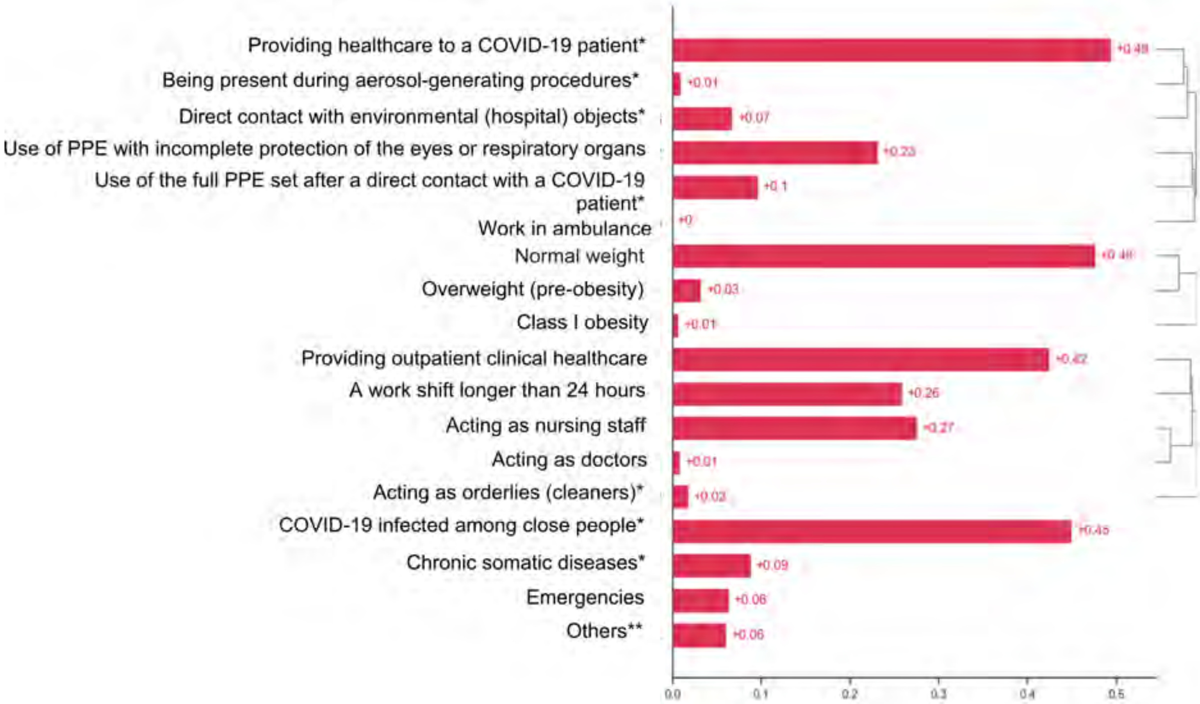

Figure 4. Clusterization of predictors that determine SARS-CoV-2 infection of healthcare workers: * means these predictors have a positive effect on SARS-CoV-2 infection of healthcare workers; ** means the sum of SHAP-values: vaccination against COVID-19 after a direct contact with a COVID-19 patient; Class II and III obesity; underweight; performing laboratory and pathological anatomy diagnostics

contact with environmental (hospital) objects ($p < 0.05$);

- Cluster 2 described quality and completeness of PPE use: use of PPE without complete protection provided for the eyes or respiratory organs ($p > 0.05$); use of the full PPE set after a direct contact with a COVID-19 patient ($p < 0.05$);
- Cluster 3 was related to healthcare workers' occupation: acting as orderlies (cleaners) ($p < 0.05$); acting as nursing staff ($p > 0.05$); acting as doctors ($p > 0.05$);
- Cluster 4 described a worker's personal characteristics such as BMI: normal weight as per body mass index (BMI) ($p > 0.05$); overweight (pre-obesity) ($p > 0.05$); Class I obesity ($p > 0.05$).

The next stage was to determine whether interactions between the analyzed occupational predictors were one-factor or multi-factorial ones.

In our study, one predictor determining SARS-CpV-2 infection was found in 56.2 % of healthcare workers; two, in 19.2 %; three, in 16.4 %; four, in 5.5 %; and five predictors, in 2.7 %.

Frequency of predictors was different under one-factor interaction. Vaccination against COVID-19 after a direct contact with COVID-19 patient occupied the leading place with 65.9 %; the second place belonged to acting as orderlies (cleaners), 22.0 %; providing healthcare to a COVID-19 patient took the third place, 12.2 %.

In case there were two predictors, the first rank place was taken by being present during aerosol-generating procedures and use of the full PPE set after a direct contact with a COVID-19 patient, both 32.1 %; the second place belonged to vaccination against COVID-19 after a direct contact with COVID-19 patient, 17.9 %; the third place, providing healthcare to a COVID-19 patient, 10.7 %; acting as orderlies (cleaners) took the fourth place with 7.1 %.

Multifactorial interactions in case of three predictors had the following structure:

being present during aerosol-generating procedures, 27.8 %; providing healthcare to a COVID-19 patient and use of the full PPE set after a direct contact with a COVID-19 patient 19.4 % both; vaccination against COVID-19 after a direct contact with COVID-19 patient and direct contact with environmental (hospital) objects 13.9 % both; acting as orderlies (cleaners), 5.6 %.

In case four predictors were present, multifactorial interactions were determined by being present during aerosol-generating procedures and a direct contact with a COVID-19 patient, both 25.0 %; providing healthcare to a COVID-19 patient and use of the full PPE set after a direct contact with a COVID-19 patient, 18.8 %; vaccination against COVID-19 after a direct contact with COVID-19 patient and acting as orderlies (cleaners), 6.3 % both.

The same frequency was established in case five predictors were identified simultaneously: vaccination against COVID-19 after a direct contact with COVID-19 patient, providing healthcare to a COVID-19 patient, being present during aerosol-generating procedures, use of the full PPE set after a direct contact with a COVID-19 patient, and direct contacts with environmental (hospital) objects, each item with 20.0 %.

Conclusion. Thus, epidemiological data mining is a modern stage in epidemiological analysis. The use of machine learning methods allows for multifactorial assessment of SARS-CoV-2 infection risks in healthcare workers, enables identifying and reliably estimating the most significant predictors and helps create relevant risk groups with an opportunity to implement a personalized approach to prevention of occupational infection with viral pathogens.

Intelligent analysis of epidemiological data has flexible architecture, which allows adjusting the model under study and supplementing new data to the existing database, detecting changes in an epidemiological situation and accomplishing relevant preventive and anti-epidemic activities. Qualitative and precise performance of a machine learning model is achieved by complete and qualitative initial data collection, preliminary data processing and using purified databases (datasets) in model training.

Funding. The study was accomplished within the scientific research work Investigation of Epidemic Process and Prevention of Healthcare-Associated Viral Infections (on the Example of Chickenpox, Norovirus and Rotavirus Infections etc.). Registration Number NIOKTR121040500099-5.

Competing interests. The authors declare no explicit or potential competing interests related to publication of this article.

References

1. Baker R.E., Mahmud A.S., Miller I.F., Rajeev M., Rasambainarivo F., Rice B.L., Takahashi S., Tatem A.J. [et al.]. Infectious disease in an era of global change. *Nat. Rev. Microbiol.*, 2022, vol. 20, no. 4, pp. 193–205. DOI: 10.1038/s41579-021-00639-z

2. Sagaya Jansi R., Khusro A., Agastian P., Alfarhan A., Al-Dhabi N.A., Arasu M.V., Rajagopal R., Barcelo D., Al-Tamimi A. Emerging paradigms of viral diseases and paramount role of natural resources as antiviral agents. *Sci. Total Environ.*, 2021, vol. 759, pp. 143539. DOI: 10.1016/j.scitotenv.2020.143539

3. Ashmore P., Sherwood E. An overview of COVID-19 global epidemiology and discussion of potential drivers of variable global pandemic impacts. *J. Antimicrob. Chemother.*, 2023, vol. 78, suppl. 2, pp. ii2–ii11. DOI: 10.1093/jac/dkad311

4. Tian C., Lovrics O., Vaisman A., Chin K.J., Tomlinson G., Lee Y., Englesakis M., Parotto M., Singh M. Risk factors and protective measures for healthcare worker infection during highly infectious viral respiratory epidemics: A systematic review and meta-analysis. *Infect. Control Hosp. Epidemiol.*, 2022, vol. 43, no. 5, pp. 639–650. DOI: 10.1017/ice.2021.18

5. Gholami M., Fawad I., Shadan S., Rowaiee R., Ghanem H., Hassan Khamis A., Ho S.B. COVID-19 and healthcare workers: A systematic review and meta-analysis. *Int. J. Infect. Dis.*, 2021, vol. 104, pp. 335–346. DOI: 10.1016/j.ijid.2021.01.013

6. Seto W.H., Tsang D., Yung R.W.H., Ching T.Y., Ng T.K., Ho M., Ho L.M., Peiris J.S.M., Advisors of Expert SARS group of Hospital Authority. Effectiveness of precautions against droplets and contact in prevention of nosocomial transmission of severe acute respiratory syndrome (SARS). *Lancet*, 2003, vol. 361, no. 9368, pp. 1519–1520. DOI: 10.1016/s0140-6736(03)13168-6

7. Zhang Y., Seale H., Yang P., MacIntyre C.R., Blackwell B., Tang S., Wang Q. Factors associated with the transmission of pandemic (H1N1) 2009 among hospital healthcare workers in Beijing, China. *Influenza Other Respir. Viruses*, 2013, vol. 7, no. 3, pp. 466–471. DOI: 10.1111/irv.12025

8. Olu O., Kargbo B., Kamara S., Wurie A.H., Amone J., Ganda L., Ntsama B., Poy A. [et al.]. Epidemiology of Ebola virus disease transmission among health care workers in Sierra Leone, May to December 2014: a retrospective descriptive study. *BMC Infect. Dis.*, 2015, vol. 15, pp. 416. DOI: 10.1186/s12879-015-1166-7

9. Alraddadi B.M., Al-Salmi H.S., Jacobs-Slifka K., Slayton R.B., Estivariz C.F., Geller A.I., Al-Turkistani H.H., Al-Rehily S.S. [et al.]. Risk Factors for Middle East Respiratory Syndrome Coronavirus Infection among Healthcare Personnel. *Emerg. Infect. Dis.*, 2016, vol. 22, no. 11, pp. 1915–1920. DOI: 10.3201/eid2211.160920

10. Smirnova S.S., Egorov I.A., Zhuikov N.N., Vyatkina L.G., Kharitonov A.N., Semenov A.V., Morova O.V. Risks of becoming infected with SARS-CoV-2 for medical personnel in a large industrial city during the pandemic: comparative assessment. *Health Risk Analysis*, 2022, no. 2, pp. 139–150. DOI: 10.21668/health.risk/2022.2.13.eng

11. Zhang P., Chen B., Ma L., Li Z., Song Z., Duan W., Qiu X. The Large Scale Machine Learning in an Artificial Society: Prediction of the Ebola Outbreak in Beijing. *Comput. Intell. Neurosci.*, 2015, vol. 2015, pp. 531650. DOI: 10.1155/2015/531650

12. Alqaissi E.Y., Alotaibi F.S., Ramzan M.S. Modern Machine-Learning Predictive Models for Diagnosing Infectious Diseases. *Comput. Math. Methods Med.*, 2022, vol. 2022, pp. 6902321. DOI: 10.1155/2022/6902321

13. Singh R., Singh R. Applications of sentiment analysis and machine learning techniques in disease outbreak prediction – A review. *Materials Today: Proceedings*, 2021, vol. 81, no. 2, pp. 1006–1011. DOI: 10.1016/j.matpr.2021.04.356

14. Abd-Alrazaq A., Alajlani M., Alhuwail D., Schneider J., Al-Kuwari S., Shah Z., Hamdi M., Househ M. Artificial Intelligence in the Fight Against COVID-19: Scoping Review. *J. Med. Internet Res.*, 2020, vol. 22, no. 12, pp. e20756. DOI: 10.2196/20756

15. Abd-Alrazaq A., Schneider J., Mifsud B., Alam T., Househ M., Hamdi M., Shah Z.A. Comprehensive Overview of the COVID-19 Literature: Machine Learning-Based Bibliometric Analysis. *J. Med. Internet Res.*, 2021, vol. 23, no. 3, pp. e23703. DOI: 10.2196/23703

16. Guo Y., Zhang Y., Lyu T., Prosperi M., Wang F., Xu H., Bian J. The application of artificial intelligence and data integration in COVID-19 studies: a scoping review. *J. Am. Med. Inform. Assoc.*, 2021, vol. 28, no. 9, pp. 2050–2067. DOI: 10.1093/jamia/ocab098

17. McCall B. COVID-19 and artificial intelligence: protecting health-care workers and curbing the spread. *Lancet Digit. Health*, 2020, vol. 2, no. 4, pp. e166–e167. DOI: 10.1016/S2589-7500(20)30054-6

18. Tostmann A., Bradley J., Bousema T., Yiek W.-K., Holwerda M., Bleeker-Rovers C., Ten Oever J., Meijer C. [et al.]. Strong associations and moderate predictive value of early symptoms for SARS-CoV-2 test positivity among healthcare workers, the Netherlands, March 2020. *Euro Surveill.*, 2020, vol. 25, no. 16, pp. 2000508. DOI: 10.2807/1560-7917.ES.2020.25.16.2000508

19. Yang H.S., Hou Y., Vasovic L.V., Steel P.A.D., Chadburn A., Racine-Brzostek S.E., Velu P., Cushing M.M. [et al.]. Routine Laboratory Blood Tests Predict SARS-CoV-2 Infection Using Machine Learning. *Clin. Chem.*, 2020, vol. 66, no. 11, pp. 1396–1404. DOI: 10.1093/clinchem/hvaa200

20. Hirten R.P., Tomalin L., Danieletto M., Golden E., Zweig M., Kaur S., Helmus D., Biello A. [et al.]. Evaluation of a machine learning approach utilizing wearable data for prediction of SARS-CoV-2 infection in healthcare workers. *JAMIA Open*, 2022, vol. 5, no. 2, pp. ooac041. DOI: 10.1093/jamiaopen/ooac041

Egorov I.A., Smirnova S.S., Semenov A.V. Using inductive machine learning to identify risk factors for healthcare workers to get infected with highly contagious viruses (based on COVID-19 model). Health Risk Analysis, 2024, no. 2, pp. 122–131. DOI: 10.21668/health.risk/2024.2.11.eng

Received: 02.02.2024
Approved: 22.04.2024
Accepted for publication: 20.06.2024