DOI: 10.63550/ICEIP.2025.84.56.054

Application of Artificial Intelligence in Solving Problems of Maintaining Public Health

Svyatoslav I. Mazilov[1], Vladimir N. Dolich[2], Nataliia E. Komleva[3], Valeria A.Savina[4], and Vladimir F. Spirin[5]

Saratov Hygiene Medical Research Center of the FBSI FSC Medical and Preventive Health Risk Management Technologies, 1a building 1, Zarechnaya str., Saratov, Russia

Keywords: artificial intelligence, machine learning,predictive modeling, risk assessment, working population, preventive medicine, social behavior, cardiovascular diseases

Abstract: This study addresses the pressing issue of cardiovascular diseases (CVD) prevalence among the working population, emphasizing the significant socio-economic and health burdens caused by chronic noncommunicable diseases. Leveraging advances in artificial intelligence and modern technologies, we developed a predictive machine learning model using XGBoost, trained on a dataset of 1142 adults including physical, clinical, social, and lifestyle factors. Our model aims to improve the accuracy of predicting individual CVD risk by integrating complex medical and behavioral data, thereby enhancing preventive strategies targeted at workers across various industries. The results highlight the potential of AI-driven approaches to support effective health monitoring and risk-oriented prevention policies.

[1] https://orcid.org/0000-0002-8220-145X
[2] https://orcid.org/0000-0002-8980-5117
[3] https://orcid.org/0000-0003-4099-9368
[4] https://orcid.org/0000-0001-7717-8235
[5] https://orcid.org/0000-0002-8980-5117

DOI: 10.63550/ICEIP.2025.84.56.054

1 INTRODUCTION

The socio-economic situation in the country that has developed over the past decades has reached a level where the st ate cannot but recognize the health of the nation as the highest national priority. In recent decades, chronic noncommunicable diseases (NCDs) have been of particular importance, among which diseases of the circulatory system (CVD) occupy a leading place. NCDs cause global socio-economic and demographic damage: they worsen the quality of life, cause disability, and pose a serious problem not only for the healthcare system, a significant proportion of whose resources are spent on their treatment, but also for the country's economy as a whole (Potemkina R.A., et al. 2021, Shlyakhto E.V., et al 2019), which indicates the need to improve the prevention system for this category of the population.

The development and application of modern technologies and artificial intelligence in modern medicine makes it possible to significantly improve the solution of preventive medicine problems, as they open up new opportunities for the development and optimization of predictive models capable of analyzing large amounts of data, identifying complex relationships between risk factors and predicting the population and individual probability of developing diseases with high accuracy and aimed at predicting the risks of development diseases and health monitoring. At the same time, the risk-based approach underlying the control and supervisory function of Rospotrebnadzor (the national public health agency of the Russian Federation) is aimed at reducing the risk of industrial emergencies and preventing occupational diseases among employees (Krasovsky V.O. et al., 2021, Tsareva G.R. et al., 2019), while a set of medical, social and behavioral factors plays an important role in the development of CVD.

Thus, given the high relevance of the problem of the prevalence of CVD among the working population, more and more data are appearing in the domestic and foreign scientific literature on the need to assess the combined effects of medical, social, and behavioral factors, taking into account the specifics of working conditions, in order to improve the level of predicting the risk of developing CVD among workers in various fields of activity (Kuznetsova Z.M. et al., 1993, Shastin A.S. et al., 2022, Perng W. et al., 2023, Filimonov E.S. et al., 2022, Zaitseva N.V. et al., 2022, Lebedeva-Nesevrya N.A. et al., 2010).

2 MATERIALS AND METHODS

To create a predictive machine learning model, the following Python programming language packages were used: pandas, nympy, matploth and seaborn for sorting data and visualizing graphs, taken from The Python Standard Library.The XGBoost package was used to train the model and visualize it, scikit-learn was used to evaluate the model performance, and the streamlit framework was used for its visualization.

The predictive model was trained on data (dataset) provided in the form of an Excel spreadsheet, which included indicators of physical, clinical laboratory and clinical instrumental studies, medical and social factors and lifestyle of the adult population (n = 1142), the average age of the surveyed was 48.0 ± 11.2 years.

3 RESULTS AND DISCUSSION

3.1 Existing Systems for Assessing the Risk of Developing CVD

Prototypes for a large number of existing predictive models for assessing the risk of developing CVD, such as: Probability of Clinical Event (PCE) (Anderson T.S. et al., 2024, Medina-Inojosa J.R. et al., 2023), Framingham-REGICOR (Ramos R. et al, 2003), Framingham Wilson and Framingham D'Agostino (Zhiting G. et al., 2022), etc. are the Framingham 10-year Risk Assessment system (Framingham Risk Score) (Kannel W.B. et al., 1961), and the SCORE (Systematic Coronary Risk Assessment) (Conroy R.M. et al., 2003). It is on the basis of SCORE that the russian assessment system was developed. The individual total 10-year risk of adverse outcomes from cardiovascular pathology for the population of the Russian Federation is "CardioRisk Expert" (Shalnova S.A. et al., 2013).

At the same time, the results of a number of studies indicate that the actual effectiveness of most predictive models for assessing the risk of CVD, derived from the SCORE scale and the Framingham system, remains unclear due to methodological shortcomings and lack of proper validation (Damen J.A. et al., 2016), which indicates the need to adapt these models taking into account population characteristics (Ramsay S.E. et al., 2011, Eichler K. et al., 2007, WHO CVD Risk Chart Working Group, 2019, Gopal D.P. et al., 2016), as well as quantitative and qualitative specifics of the analyzed samples (Kurogi K. et al., 2025).

3.2 The Importance of Model Adaptation

Adaptation of existing predictive models to specific study populations is a critical step to improve the accuracy and relevance of chronic NCD risk predictions, including CVD. Universal models developed based on some populations may demonstrate insufficient efficiency when applied to others with different genetic, socioeconomic, cultural and environmental characteristics. Therefore, it is necessary to calibrate and validate existing models on new populations taking into account both quantitative (age, body mass index, cholesterol level) and qualitative (education, social status, dietary habits) risk factors. In addition, it is necessary to consider the specific impact of these risk factors in a specific study population, since their influence may vary depending on the context and interaction with other factors. Only as a result of such adaptation and consideration of the characteristics of the study groups can reliable and practically significant predictions be obtained, facilitating the development of effective prevention strategies and personalized medicine.

3.3 On the Need to Develop a Predictive Model for Assessing the Risk of Developing CVD among Working Population, Taking into Account the Influencing Medical and Social Factors and the Production Environment

In the context of the prevention of CVD, the development of a predictive model aimed at industrial workers is particularly relevant. Given the specifics of professional activity, characterized by the impact of unique combinations of medical and social factors (for example, stress, sedentary lifestyle, irregular diet) and environmental factors (noise, vibration, exposure to chemicals, night shifts), the use of universal models for predicting the risk of CVD may not be effective enough. A specialized model is needed that can integrate data on the impact of these factors on the individual risk of developing CVD, which will more accurately identify high-risk workers, develop and implement targeted preventive measures aimed at minimizing the negative impact of the work environment and lifestyle correction, and, as a result, reduce morbidity and disability among the working-age population.

3.4 Development ofaPredictive Model forAssessingthe Risk of DevelopingCVD

At the Saratov Hygiene Medical Research Center a predictive model for assessing the risk of developing CVD among the adult population was developed and visualized.

To build the model, 1142 patients were examined and divided into 3 groups. Group (i) included 277 individuals with hypertension, group (ii) included 121 individuals with coronary heart disease, and group (iii) included 743 conditionally healthy individuals.The studied sample was distributed by classification type using an unsupervised machine learning model.

The first step in creating a predictive model was to analyze the data provided as an Excel spreadsheet using Python programming language packages: pandas, nympy, matploth, and seaborn to sort data and visualize graphs (Pavlova, 2003).

To determine the relationships, the strength of the correlation between the studied indicators were analyzed: gender, age, height, weight, body mass index (BMI), systolic blood pressure (BP), diastolic BP, heart rate (HR), ALT, AST, total bilirubin, direct bilirubin, glucose, glycated hemoglobin, creatinine, total cholesterol, high-density lipoprotein cholesterol (HDL-C) and low-density lipoprotein cholesterol (LDL-C), triglycerides, atherogenicity coefficient, total protein. Due to insufficient strength of the correlation, the following parameters were excluded from the study: HR, systolic BP, and diastolic BP.

As a method for solving the classification problem, a classical machine learning model (decision trees) was used – gradient boosting (XGBoost), which allows one to effectively find nonlinear dependencies in data from different samples (Pavlova, 2003).

Using the Python programming language, the dataset was divided into training and test samples in a ratio of 80% and 20%, respectively.

The following metrics were used to evaluate the model's performance: accuracy (the proportion of correctly classified objects out of their total number); recall (the proportion of correctly predicted positive objects out of all positive objects); precision (the proportion of correctly predicted positive objects out of all predicted positive objects); F1 measurement (the harmonic mean between accuracy and completeness).

The accuracy of the developed method was 65.2%, which justifies the need for further expansion of the dataset and the use of additional methods for configuring the model, such as sampling classes and changing hyperparameters.

To visualize the predictive model, the Streamlit framework, based on Python, was used.It allows the machine learning model to be transferred to a web page for public use.

3.5 On the Potential Area of Application of the Developed Model (Improvement of Social And Hygienic Monitoring, Implementation of the Model in Corporate Programs of Industrial Enterprises)

The developed predictive model for assessing the risk of developing NCD, in particular CVD, in the working population has significant potential for application in several key areas. Firstly, it can be used to improve the existing hygiene monitoring system, allowing for a more accurate assessment of the impact of environmental factors on workers' health and the development of targeted preventive measures. Secondly, the introduction of the model into corporate health programs of industrial enterprises will make it possible to identify people at high risk of developing diseases at an early stage, which will make it possible to timely implement personalized interventions, including lifestyle adjustments, dietary recommendations and, if necessary, drug therapy. In addition, given the understaffing of primary care facilities, the developed model can be implemented in the form of screening at the patient's initial request for medical help, in pre-medical reception rooms, during medical examinations, preliminary and periodic medical examinations. This approach, based on proactive identification and risk management, contributes to improving the health and well-being of employees, reducing economic losses associated with morbidity, and increasing overall labor productivity.

4 CONCLUSIONS

The development of an artificial intelligence-based predictive model for assessing cardiovascular disease risk demonstrates promising opportunities in processing multifaceted medical, social, and behavioral data related to the working population. By including a variety of factors affecting CVD the development, the model can support early detection of high-risk individuals, allowing timely intervention and reducing socio-economic consequences. This approach is in line with current trends in preventive medicine and regulatory frameworks that emphasize reducing occupational health risks. Further work should focus on expanding data sets, refining model parameters, and integrating such tools into conventional healthcare management systems to maximize their practical benefits.

REFERENCES

Anderson, T.S., Wilson, L.M., Sussman, J.B., 2024. Atherosclerotic Cardiovascular Disease Risk Estimates Using the Predicting Risk of Cardiovascular Disease Events Equations. *JAMA Internal Medicine*. Vol. 184(8), p. 963-970.

Conroy, R.M., Pyörälä, K., Fitzgerald, A.P., Sans, S., Menotti, A., De Backer, G., De Bacquer, D., Ducimetière, P., Jousilahti, P., Keil, U., Njølstad, I., Oganov, R.G., Thomsen, T., Tunstall-Pedoe, H., Tverdal, A., Wedel, H., Whincup, P., Wilhelmsen, L., Graham, I.M., SCORE project group, 2003. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *European Heart Journal. Vol.* 24(11), p. 987-1003.

Damen, J.A., Hooft, L., Schuit, E., Debray, T.P., Collins, G.S., Tzoulaki, I., Lassale, C.M., Siontis, G.C., Chiocchia, V., Roberts, C., Schlüssel, M.M., Gerry, S., Black J.A., Heus P., van der Schouw Y.T., Peelen L.M., Moons, K.G., 2016. Prediction models for cardiovascular disease risk in the general population: systematic review. *British Medical Journal*. Vol. 353, p. i2416.

Eichler, K., Puhan, M.A., Steurer, J., Bachmann, L.M. 2007. Prediction of first coronary events with the Framingham score: a systematic review. *American Heart Journal*. Vol. 153(5), p. 722-731. 10.1016/j.ahj.2007.02.027 40.

Filimonov, E.S., Korotenko, O.Yu., 2022. Atherosclerosis prediction system based on the identification of the most significant risk factors in workers of the main professions of the coal industry in the South of Kuzbass. *Medicina v Kuzbasse (Medicine in Kuzbass)*. Vol. 21(3), p. 80-85.

Gopal, D.P., Usher-Smith, J.A., 2016. Cardiovascular risk models for South Asian populations: a systematic review. *International Journal of Public Health*. Vol. 61(5), p. 525-534.

Kannel, W.B., Dawber, T.R., Kagan, A., Revotskie, N., Stokes J. 3rd., 1961. Factors of risk in the development of coronary heart disease – six year follow-up experience. The Framingham Study. Annals of Internal Medicine, Vol. 55, p. 33-50.

Krasovsky, V.O., Yakhina, M.R., 2021. Management of sanitary supervision of occupational health at oil refineries and petrochemical plants (analytical review). *Gigiena i sanitariya (Hygiene and Sanitation)*. Vol. 100(3), p. 246-253.

Kurogi, K., Yura, S., Moriyama, K., Tsuda, E., Yoshida, N., Ito, M., 2025. Assessment of the application of atherosclerotic disease risk scores in the workplace. *Sangyo Eiseigaku Zasshi (Journal of Occupational Health Ranking)*. Vol. 67(1), p. 9-25.

Kuznetsova, Z.M., Polzik, E.V., Danilenko, I.I., 1993. Forecasting the development of cardiovascular diseases in employees engaged in managerial work. *Gigiena i sanitariya (Hygiene and sanitation).* Vol.1, p. 23-25.

Lebedeva-Nesevrya, N.A., Kiryanov, D.A., Barg, A.O., 2010. Assessment of a combined impact of social and occupational risk factors on health of powder metallurgy workers according to data of epidemiological studies. *Zdorovye naseleniya i sreda obitaniya (Public Health and Life Environment).* Vol. 11(212), p. 44-46.

Medina-Inojosa, J.R., Somers, V.K., Garcia, M., Thomas, R.J., Allison, T., Chaudry, R., Wood-Wentz, C.M., Bailey, K.R., Mulvagh, S.L., Lopez-Jimenez, F., 2023. Performance of the ACC/AHA Pooled Cohort Cardiovascular Risk Equations in Clinical Practice. *Journal of the American College of Cardiology.* Vol. 82(15), p. 1499-1508.

Pavlova, A.I., 2003. Application of machine learning algoritms for heart disease prediction. *Siberian Journal of Life Sciences and Agriculture*. Vol. 15(3), p. 475-496.

Perng,W., Aris, I.M., Slopen, N., Younoszai, N., Swanson, V., Mueller, N.T., Sauder, K.A., Dabelea, D., 2023. Application of Life's Essential 8 to assess cardiovascular health during early childhood. *Annals of Epidemiology*. Vol. 80, p. 16-24..

Potemkina, R.A., Mylnikova, L.A., Kamynina, N.N., Pivovarova, O.A., 2021 Prevention of noncommunicable diseases: from risk factors to the national programs. *Zdravookhranenie Rossiiskoi Federatsii (Health Care of the Russian Federation, Russian journal).* Vol. 65(5): 44-446.

Ramos, R., Solanas, P., Cordón, F., Rohlfs, I., Elosua, R., Sala, J., Masiá, R., Faixedas, M.T., Marrugat, J., 2003. Comparación de la función de Framingham original y la calibrada del REGICOR en la predicción del riesgo coronario poblacional. *Medicinaclínica (Barcelona).* Vol. 121(14), p. 521-526.

Ramsay, S.E., Morris, R.W., Whincup, P.H., Papacosta, A.O., Thomas, M.C., Wannamethee, S.G., 2011. Prediction of coronary heart disease risk by Framingham and SCORE risk assessments varies by socioeconomic position: results from a study in British men. *The European Journal of Cardiovascular Prevention & Rehabilitation.* Vol. 18(2), p. 186-193.

Shlyakhto, E.V., Zvartau, N.E., Villevalde, S.V., Yakovlev, A.N., Soloveva, A.E., Alieva, A.S., Avdonina, N.G., Medvedeva, E.A., Fedorenko, A.A., Kulakov, V.V., Karlina, V.A., Endubaeva, G.V., Zaitsev, V.V., Soloviev, A.E, 2019. Cardiovascular risk management system: prerequisites for developing, organization principles, target groups. *Rossijskij kardiologicheskij zhurnal (Russian journal of cardiology)*. Vol.24(11),p.

Shalnova, S.A., Kalinina, A.M., Deev, A.D., Pustelenin, A.V., 2013. Russian expert system ORISKON – Assessment of the major non-communicable disease risk. *Kardiovaskulyarnaya terapiya i profilaktika (Cardiovascular Therapy and Prevention).* Vol. 12(4), p. 51-55.

Shastin, A.S., Gazimova, V.G., Guselnikov, S.R., Stamikov, N.I., Bakhtereva, E.V., 2022. Morbidity among metallurgists by the results of periodic health checkups and the analysis of temporary disability. *Medicina truda i ekologiya cheloveka (Occupational Health and Human Ecology).* Vol.4, p. 46-64.

Tsareva, G.R., Ignatieva ,T.V., Shpak, N.A., 2019. Application of risk-oriented approach to the organization of public control (supervision) on the example of Rospotrebnadzor of the Republic of Mari El. *Vek kachestva (The Age of Quality).* Vol. 2, p. 19-31.

WHO CVD Risk Chart Working Group, 2019. World Health Organization cardiovascular disease risk charts: revised models to estimate risk in 21 global regions. *Lancet Global Health.* Vol. 7(10), p. e1332-e1345.

Zaitseva, N.V., Onishchenko, G.G., May, I.V., Shur P.Z., 2022. Developing the methodology for health risk assessment within public management of sanitary-epidemiological welfare of the population. *Analiz riska zdorov'yu (Health Risk Analysis)*. Vol. 3, p. 4-20.

Zhiting, G., Jiaying, T., Haiying, H., Yuping, Z., Qunfei, Y., Jingfen, J., 2022. Cardiovascular disease risk prediction models in the Chinese population ¬ a systematic review and meta-analysis. *BMC Public Health*. Vol.22, p.1608.