Applying the FAIR Principles to Accelerate Health Research in Europe in the Post COVID-19 Era
J. Delgado et al. (Eds.)

doi:10.3233/SHTI210835

Influence of Healthcare Organization Factors on Cardiovascular Diseases Mortality

Oleg METSKER[a,1] and Georgy KOPANITSA[b]
[a] *Almazov National Medical Research Centre, Saint Petersburg, Russia*
[b] *ITMO University, Saint Petersburg, Russia*

Abstract. One serious pandemic can nullify years of efforts to extend life expectancy and reduce disability. The coronavirus pandemic has been a perturbing factor that has provided an opportunity to assess not only the effectiveness of health systems for cardio-vascular diseases (CVD), but also their sustainability. The goal of our research is to analyze the influence of public health factors on the mortality from circulatory diseases using machine learning methods. We analysed a very large dataset that consisted of the information collected from the national registers in Russia. We included data from 2015 to 2021. It included 340 factors that characterize organization of healthcare in Russia. The resulting area under receiver operating characteristic curve (AUC of ROC) of the Random Forest based regression model was 92% with a testing dataset. The models allow for automated retraining as time passes and epidemiological and other situations change. They also allow additional characteristics of regions and health care organizations to be added to existing training datasets depending on the target. The developed models allow the calculation of the probability of the target for 6-12 months with an error of 8%. Moreover, the models allow to calculate scenarios and the value of the target indicator when other indicators of the region change.

Keywords. Cardiovascular disease, machine learning, public health, prediction, keyword

1. Introduction

One serious pandemic can nullify years of efforts to extend life expectancy and reduce disability [1]. The coronavirus pandemic has been a perturbing factor that has provided an opportunity to assess not only the effectiveness of health systems for cardio-vascular diseases (CVD), but also their sustainability [2,3]. The dynamics of total and CVD mortality can be used as a measure of health system resilience. Efficiency and sustainability are different and, in many ways, mutually exclusive, but sustainability is as necessary for sustained positive dynamics as efficiency is for achieving goals [2]. Analyzing the situation with COVID-19 we understand the need to assess the sustainability of health systems in relation to CVD care, a sustainable system will have different characteristics compared to an effective one [4].

[1] Corresponding Author, Georgy Kopanitsa, ITMO University, Saint-Petersburg, Russia; E-mail: georgy.koapnitsa@gmail.com.

In the post-COVID era, we should strive to build balanced, rather than efficient, systems with sufficient resilience[1]. So, when solving the inverse problem of forecasting an indicator/indicator of quality of treatment of a region it is possible to calculate what the characteristics of the region should be today and tomorrow (values quantitatively qualitatively) to get the required indicator the day after tomorrow.

Models and algorithms for the analysis of risk factors and prognosis of mortality in the acute phase of the disease have been developed. Many works apply methods of artificial intelligence. For example [5], considers the identification of ischemic stroke risk factors in conditions of data shortage. Many studies, such as [6–8] use large population databases, including up to 800,000 patients, to predict stroke incidence over 5 years. Despite rather high accuracy: up to 87% correct prediction of ischemic stroke and up to 82% prediction of hemorrhagic stroke, the developed methods based on neural networks and machine of reference vectors do not allow to work in conditions of uncertainty and data gaps in electronic medical histories.

Much attention is paid to treatment planning and prognosis of recovery in the acute phase of the disease. For ischemic stroke, models based on neural networks and support vector method show the best performance [9,10]. The correctness of the models reaches 74% in the best cases, which cannot be considered a satisfactory result. However, the influence of organizational factors on population mortality from CVD has not been considered in detail in the scientific literature.

1.1. Objectives

The goal of our research is to analyze the influence of public health factors on the mortality from circulatory diseases using machine learning methods.

2. Methods

2.1. Dataset

The dataset consisted of the following information collected from the national registers in Russia. We included data from 2015 to 2021. Information about the activities of the organization providing medical care; Information on the number of diseases registered in patients residing in the service area of the medical organization; information on the movement of patients; Information on confirmed cases of death in the following nosologies and their International Statistical Classification of Diseases and Related Health Problems (ICD 10) codess: Diseases of the circulatory system (I00-I99), acute coronary syndrome (I20- I22), Cerebrovascular diseases - Subarachnoid hemorrhage, Intracerebral hemorrhage, Brain infarction, Stroke not specified as hemorrhage or infarction, Congestion and stenosis of the precerebral arteries, Embolisms, Consequences of cerebrovascular disease (I60-I69), Novoplasms (C00-D48) including oncohematological patients with C90, Delivery O80-O84, Endocrine diseases, eating disorders and metabolic disorders (E00-E90) including diabetes and obesity, as well as death from Sepsis (A40-41), Anemias (D50-D64), Selected disorders involving the immune mechanism (D80-D89), Obesity (E66), Chronic rheumatic heart disease (I05-I09), Influenza (J09-J11), Acute respiratory upper respiratory tract infections (J00-J06, line 11. 1)); Information on the staff of medical organizations, information on surgical

work, information on resources of clinics. Indicators of socio-economic development of regions.

2.2. Machine learning

The regression task for predicting the CVD mortality was solved using the scikit-learn library. In total 340 indicators were used as predictors. Each experiment ran in the setting of stratified 5-fold cross-validation (i.e., random 80% of records were used for training and 20% for testing, target class ratios in the folds were preserved).

For the performance assessment, we ran it 100 times; and 100 x 5-fold cross-validation with total of 500 predictions. As an additional performance assessment score, we used the AUC of ROC. The AUC was calculated based on an average of 5 curves (one curve per fold in the setting of 5-fold cross-validation). Features importance was calculated using a random-forest model.

3. Results

The resulting AUC of ROC of the Random forest based regression model was 92% with a testing dataset. Figure 1 shows the result of calculating the significance of predictors using a machine learning model in solving the regression problem. Training was performed on the data set of 340 indicators of RF regions from 2015 to 2021, including both dynamic indicators (spread of coronavirus infection, mortality from other nosologies including cerebrovascular diseases, coverage of vaccination campaign, population movement, etc.), intensity and coverage of measures to reduce mortality in the region. Examples of such activities were the number of publications in the media

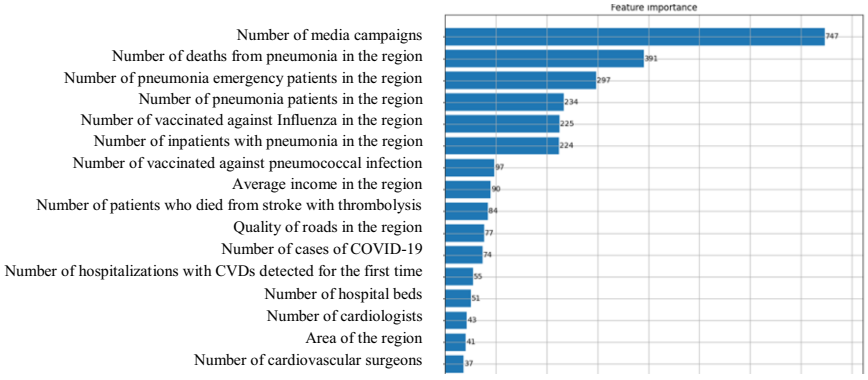

Figure 1. Example of calculating the contribution of regional indicators to CVD mortality using machine learning methods

4. Discussion

The models allow for automated retraining as time passes and epidemiological and other situations change. They also allow additional characteristics of regions and health care organizations to be added to existing training datasets depending on the target. The

developed models allow the calculation of the probability of the target for 6-12 months with an error of 8%. Moreover, the models allow to calculate scenarios and the value of the target indicator when other indicators of the region change.

5. Conclusion

The development and implementation of medical information technologies based on machine-learning methods contributes to the development of a unified accessible methodology for analyzing the processes of providing medical care for quality management at all levels of the healthcare system, while maintaining the success achieved in informatization and the existing infrastructure without significant additional costs.

Acknowledgements

This work was financially supported by the Ministry of Science and Higher Education of the Russian Federation (Agreement No. 075-15-2020-901). This work financially supported by the government of the Russian Federation through the ITMO fellowship and professorship program.

References

[1] Palmer K, Monaco A, et al. The potential long-term impact of the COVID-19 outbreak on patients with non-communicable diseases in Europe: consequences for healthy ageing. Aging clinical and experimental research. 2020 Jul;32:1189-94.
[2] Antony J, Sreedharan R, Chakraborty A, Gunasekaran A. A systematic review of Lean in healthcare: a global prospective. International Journal of Quality & Reliability Management. 2019 Sep 2.
[3] Gasmi A, Peana M, Pivina L, Srinath S, Benahmed AG, Semenova Y, Menzel A, Dadar M, Bjørklund G. Interrelations between COVID-19 and other disorders. Clinical Immunology. 2020 Dec 14:108651.
[4] Huang Y, Cai X, Mai W, Li M, Hu Y. Association between prediabetes and risk of cardiovascular disease and all cause mortality: systematic review and meta-analysis. Bmj. 2016 Nov 23;355.
[5] Reberg K, et al. Chronic subdural hematoma, atrial fibrillation and ishemic stroke, considerations about treatment options, a caserepport. Eur Stroke J.2018;3.
[6] Glymour MM, Maselko J, Gilman SE, Patton KK, Avendano M. Depressive symptoms predict incident stroke independently of memory impairments. Neurology. 2010 Dec 7;75(23):2063-70.
[7] Li T, Li G, Guo X, Li Z, Yang J, Sun Y. Predictive value of echocardiographic left atrial size for incident stoke and stroke cause mortality: a population-based study. BMJ open. 2021 Mar 1;11(3):e043595.
[8] Watanabe J, Kakehi E, Kotani K, Kayaba K, Nakamura Y, Ishikawa S. Isolated low levels of high‐density lipoprotein cholesterol and stroke incidence: JMS Cohort Study. Journal of clinical laboratory analysis. 2020 Mar;34(3):e23087.
[9] Wu G, Chen X, Lin J, Wang Y, Yu J. Identification of invisible ischemic stroke in noncontrast CT based on novel two‐stage convolutional neural network model. Medical Physics. 2021 Mar;48(3):1262-75.
[10] Liu Y, Yin B, Cong Y. The Probability of Ischaemic Stroke Prediction with a Multi-Neural-Network Model. Sensors. 2020 Jan;20(17):4995.