e-Cardiology / Digital Health, Public Health, Health Economics, Research Methodology –
Research Methodology: Big Data Analysis

3557

Adapting neural network models to predict 10-year CVD development based on regional data calibration

I. Korsakov, D. Gavrilov, L. Serova, A. Gusev, R. Novitskiy, T. Kuznetsova

K-SkAI, Petrozavodsk, Russian Federation

Background: The used tools for prediction the individual risk of developing cardiovascular diseases and their complications using machine learning methods have proven better prognostic value in comparison with commonly used scales (e.g., Framingham, SCORE). To create such methods, the long-term accumulation of large amount of qualitative data are required. Moreover, to improve the accuracy of models, it is necessary to take into account regional characteristics that affect health: ethnic, nutritional characteristics, climatic conditions, living standards and medical care. These regional characteristics could significantly affect the development and outcomes of CVDs. However, the amount of regional data is not enough to build a qualitative model. Therefore, it is proposed to create models based on publicly available data and validate them on regional medical data sufficient for validation and calibration.

Methods: Two models were trained using data from the Framingham study. Model 1 was trained on 2 588 patient data and predicts a 10-year CVD probability according to the following risk factors: age, gender, cholesterol, HDL, smoking, SBP, and BP medications. Model 2 was trained on 4,363 patient data and predicts a 10-year death probability from CVD according to the following criteria: age, gender, cholesterol, smoking, SBP, BMI, heart rate. To retrain the obtained models, we used dataset created from data from patients in the northwestern part of Russia. The dataset consists of 438 patients, including the signs used in the trained models. This dataset includes CVD and death from it during a 10-year follow-up

Evaluation: We used randomized data splitting: divided the dataset into a training and a test set with an 80/20 proportion. The models was implement with keras convolution neural network (CNN) using 3 hidden layers. For data validation was used a 10 K-fold method.

Results: We compared the initial model metrics and those obtained after local data retraining. The accuracy of model 1 before retraining is 78%, after – 81.3%, the area under the ROC curve (AUC) before retraining: 0.77 (at 95% CI: 0.72–0.82C), after – 0.803. The accuracy of model 2 before retraining is 79%, after – 85.6%, the area under the ROC-curve (AUC) before retraining: 0.78 (at 95% CI: 0.72–0.82), after – 0.828.

Conclusion: Using this method of retraining predictive models, we can take into account local characteristics of the population and significantly increase the accuracy of predicting events. Expand the population to use the model according to local characteristics.