Machine learning in medicine: an algorithm for automatic 12-lead ECG classification.

Mateusz Soliński[1], Michał Łepek[1], Dorota Kokosińska[1], Judyta Salamon[1], Katarzyna Muter[2], Antonina Pater[1], Zuzanna Puzio[1], Przemysław Wiszniewski[3]

[1]Faculty of Physics, Warsaw University of Technology

[2]Faculty of Electronics and Information Technology, Warsaw University of Technology

[3]Faculty of Electrical Engineering, Warsaw University of Technology

Cardiovascular diseases (CVDs) are the leading cause of death worldwide, killing about 17.9 million lives each year (WHO, 2016). The growing demand for diagnostic testing is leading to the development of new solutions for automatic classification of recorded signals. The electrocardiogram (ECG) is a standard tool in the diagnosis of many cardiovascular diseases. Creating a robust and fast algorithm for automatic classification of ECG signal is crucial to improve the quality of healthcare, especially in countries where there is a lack of experienced specialists and where the healthcare system is overloaded. The aim of PhysioNet Challenge 2020 is to create an algorithm for classification of 12-lead ECGs including nine classes: AF, I-AVB, LBBB, Normal, PAC, PVC, RBBB, STD, STE. The initial training set consisted of 6,877 ECG recordings (sampled at 500 Hz) lasting from 6 to 60 seconds. We propose the machine learning algorithm based on neural networks. The ECG signal was pre-processed using moving median filters to remove high-frequency noise and baseline wandering. Next, we removed all signal segments with a low quality and distorted morphology of the complexes using a simple threshold method based on standard deviation of ECG signal. We used an alternative hybrid approach to QRS detection in order to obtain RR time intervals. It consists of two complementary methods in a hierarchical order: one based on nonlinear transformation and first-order Gaussian differentiator as superior and the other based on wavelet transform. We selected 57 features: 24 features were proposed in sample entry (written in MATLAB) and 33 created based on international criteria for diagnosing abnormal classes featured in the Challenge. The best overall score (geometric mean of F_2 and G_2) we achieved in the unofficial phase of the PhysioNet Challenge 2020 is 0.558 with F_2 0.707 and G_2 0.442.