"Science Stays True Here"
Journal of Mathematics and Statistical Science (ISSN 2411-2518), Vol.5, 81-84 | Science Signpost Publishing

Diagnosis of Heart Diseases with Machine Learning

Bogdanov M.R.[1,2], Dumchikov A.A.[3], Akhmerova A.Z[3], Nasyrov D.I[1]

1. Ufa State Aviation Technical University, Ufa City, Russia.

2. M.Akmullah named after Bashkir State Pedagogical University, Ufa City, Russia.

3. B.N.Elcin Ural Federal University, Ekaterinburg, Russia.

Abstract

Under the research, the effectiveness of 14 machine learning algorithms for the diagnosis of cardiovascular diseases was evaluated. A PTB database of digitized electrocardiograms was used. It was found that the most preferred are a Label Propagation classifier (accuracy of recognition is 0.94), An extremely randomized tree classifier (accuracy is 0.92) and a Classifier implementing the k-nearest neighbors vote (accuracy is 0.90).

Keywords: Heart diseases, Diagnosis, Machine Learning, Electrocardiogram.

Previous work

Machine learning it is getting more popular in medicine. In particular, thanks to methods of machine learning, it is possible to carry out remote and automatic diagnostics of diseases, identify risk factors, substantiate optimal treatment strategies. There are supervised learning and unsupervised learning methods of Machine Learning. Supervised methods tries to predict of outcome, make classification of observation and estimation of a parameter [1]. Johnson et al. reported on using of Regularized regression [2], Ensembles of decision trees [3] and Support vector machines in diagnosis of heart diseases [4]. Unsupervised Machine Learning methods discovers of hidden structure in a data and tries to research the relationships between variables. There are a few papers on using of unsupervised methods in cardiology. Kiranyaz et al. used a Tensor factorization for Real-Time Patient-specific ECG classification [5], Abdolmanafi et al. reported on using on Topological data analysis for automatic tissue classification of coronary artery [6], Choi et al. used a recurrent neural network models for early detection of heart failure onset [7].

Motivation and Aim

The goal of the researching is founding the more suitable method of machine learning for diagnosing of heart diseases.

Corresponding Author: Bogdanov M.R., Associate Professor at Computation Mathematics and Cybernetics Department of Ufa State Aviation Technical University, Ufa City, Russia, +7(917)734-93-67, bogdanov_marat@mail.

Materials and Methods

We used a Physikalisch-Technische Bundesanstalt (PTB) database of digitized electrocardiograms presented by professor Michael Oeff to physionet.org project [8]. The database contains 549 records from 290 subjects. Each subject is represented by one to five records. Each record includes 15 simultaneously measured signals: the conventional 12 leads together with the 3 Frank lead ECGs. Each signal is digitized at 1000 samples per second, with 16 bit resolution over a range of ± 16.384 mV [9].

Preprocessing and Feature Extraction

We used a python biosppy library for preprocessing and feature extraction [10]. While preprocessing we extracted the first lead from ECG signals, reduced a noises and extracted the QRS complexes from signals (Fig. 1). We used temporal and amplitude characteristics of P,Q,S and T regions of cardio cycles and amplitute values of R-peaks (in total 9 features) (Fig. 1).

Cross-validation

A PTB is a good annotated ECG database. There are following class labels: Myocardial infarction, Cardiomyopathy/Heart failure, Bundle branch block, Dysrhythmia, Myocardial hypertrophy, Valvular heart disease, Myocarditis, Miscellaneous, Healthy controls (in total 9 class labels). We composed 6 datasets (feature matrix plus class label vector) for ECG signals of 5, 10, 15, 20, 25 and 30 seconds length. Then each dataset was spitted into training set and testing set with ratio of 75:25.

Classification

We used 14 methods of Machine Learning for classification (Naive Bayes classifier for multivariate Bernoulli models, A decision tree classifier, An extremely randomized tree classifier, Classifier implementing the k-nearest neighbors vote, Label Propagation classifier, Linear Discriminant Analysis, Linear Support Vector Classification, Logistic Regression (aka logit, MaxEnt) classifier, Nearest centroid classifier, A random forest classifier, Classifier using Ridge regression, Ridge classifier with built-in cross-validation, Gaussian Mixture Models, SVM).

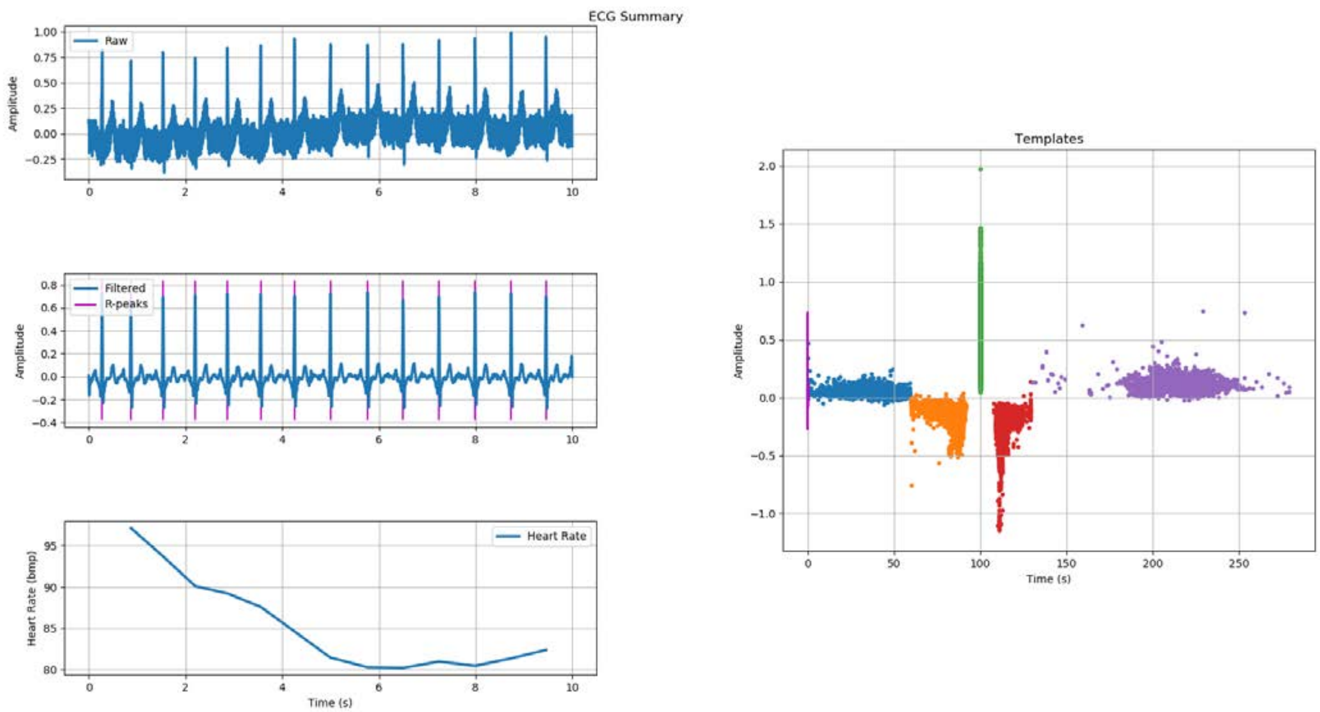

Fig. 1. Feature preprocessing with biosppy python library.

Results and Discussion

Table 1 shows effectiveness of various methods of Machine Learning for heart diseases classification. As we can see the most accurate methods of classification are Label Propagation classifier (accuracy of recognition is 0.94), An extremely randomized tree classifier (accuracy is 0.92) and a Classifier implementing the k-nearest neighbors vote (accuracy is 0.90). Second, signal duration of 5 seconds is enough for good recognition of cardiovascular diseases.

Table 1. Effectiveness of various methods of Machine Learning for heart diseases classification

	5 sec	10 sec	15 sec	20 sec	25 sec	30 sec
Naive Bayes classifier for multivariate Bernoulli models	0.70	0.69	0.68	0.35	0.67	0.67
A decision tree classifier	0.83	0.89	0.90	0.88	0.92	0.92
An extremely randomized tree classifier	0.92	0.96	0.97	0.97	0.98	0.98
Classifier implementing the k-nearest neighbors vote	0.90	0.96	0.95	0.95	0.97	0.97
Label Propagation classifier	0.94	0.97	0.97	0.96	0.98	0.98
Linear_Discriminant_Analysis	0.72	0.69	0.68	0.73	0.68	0.68
Linear Support Vector Classification	0.72	0.70	0.68	0.62	0.68	0.68
Logistic Regression (aka logit, MaxEnt) classifier	0.71	0.70	0.68	0.80	0.68	0.68

Nearest centroid classifier	0.18	0.21	0.21	0.59	0.22	0.22
A random forest classifier	0.73	0.70	0.69	0.17	0.69	0.69
Classifier using Ridge regression	0.72	0.70	0.68	0.20	0.68	0.68
Ridge classifier with built-in cross-validation	0.72	0.70	0.68	0.20	0.68	0.68
Gaussian Mixture Models	0.65	0.66	0.65	0.88	0.65	0.65
SVM	0.79	0.85	0.86	0.92	0.87	0.87

Acknowledges

The reported study was funded by RFBR according to the research project № 19-07-00780.

Literature

[1]. Kipp W. Johnson, Jessica Torres Soto, Benjamin S. Glicksberg, Khader Shameer, Riccardo Miotto, Mohsin Ali, Euan Ashley, Joel T. Dudley. Artificial Intelligence in Cardiology. Journal of the American College of Cardiology. vol. 71, no. 23, 2018.

[2]. Kolek MJ, Graves AJ, Xu M, et al. Evaluation of a prediction model for the development of atrial fibrillation in a repository of electronic medical records. JAMA Cardiol 2016;1:1007-13.

[3]. Pavlou M, Ambler G, Seaman SR, et al. How to develop a more accurate risk prediction model when there are few events. BMJ 2015;351:h3868.

[4]. Narula S, Shameer K, Salem Omar AM, Dudley JT, Sengupta PP. Machine-learning algorithms to automate morphological and functional assessments in 2D echocardiography. J Am Coll Cardiol 2016;68:2287-95.

[5]. Kiranyaz S, Ince T, Gabbouj M. Real-Time Patient-specific ECG classification by 1-D convolutional neural networks. IEEE Trans Biomed Eng 2016;63:664-75.

[6]. Abdolmanafi A, Duong L, Dahdah N, Cheriet F. Deep feature learning for automatic tissue classification of coronary artery using optical coherence tomography. Biomed Opt Express 2017;8: 1203-20.

[7]. Choi E, Schuetz A, Stewart WF, Sun J. Using recurrent neural network models for early detection of heart failure onset. J Am Med Inform Assoc 2017;24:361-70.

[8]. Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 101(23):e215-e220 [Circulation Electronic Pages; http://circ.ahajournals.org/content/101/23/e215.full]; 2000 (June 13).

[9]. Bousseljot R, Kreiseler D, Schnabel, A. Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet. Biomedizinische Technik, Band 40, Ergänzungsband 1 (1995) S 317

[10]. Official web site of biosppy project. https://pypi.org/project/biosppy/

Published: Volume 2019, Issue 4 / April 25, 2019