A ADVANCED JOURNAL IN ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING TRENDS IN BIOLOGICAL SCIENCES
Open Journal

Review

Using Machine Learning Algorithms for Heart Rate Time Series Classification

Zahra Akbari*

Department of Computer Science Karaj Azad University, Iran

****Corresponding author***
Zahra Akbari
Department of Computer Science Karaj Azad University, Iran

Article information
Received: February 17th, 2024; **Revised:** April 18th, 2024; **Accepted:** May 22nd, 2024; **Published:** June 14th, 2024

Cite this article
Akbari Z. Using machine learning algorithms for heart rate time series classification. 2024; 3(1). doi: https://doi.org/10.70705/ppp.bioai.2024.v03.i01.pp27-30

ABSTRACT
The electrocardiogram (ECG) is a vital tool for the detection of cardiac problems. Doctors are busier than usual due to the high volume of cardiac patients. An automated computer detection method is required to lessen their workload. This research presents a computer system that can categorize electrocardiogram (ECG) data. Analyses are conducted using the MIT-BIH ECG arrhythmia database. Data feature extraction follows the preprocessing step of making the ECG signal noisy. The feature extraction process involves building a support vector machine (SVM) and using a decision tree to divide the ECG signal into two groups. We may call it normal or abnormal. The system successfully categorizes the input ECG signal with a sensitivity level of 90%, according to the findings.

Keywords
Classification; Machine learning; Time series; Decision tree algorithm; SVM algorithm.

INTRODUCTION

Time series modeling is a dynamic research area that has attracted the attention of the research community over the past decades. The The primary goal of time series modeling is to construct an appropriate model that explains the series' underlying structure by precisely collecting observations of the series in the past. In order to make predictions about the future based on what we know about the past, we apply this model for time-series predictions [1].

By collecting data on physical, chemical, and biological factors, medical equipment may track the features of a patient's clinical state. As an example, electrocardiograms (ECGs) are crucial for the diagnosis of many heart ailments and other problems [2]. These ECGs entail the monitoring of changes in the electrical potential generated by the heart over time.

Cardiovascular Variability

Extrinsic rules of heart rate (HR) are the primary determinants of heart rate variability (HRV), which is defined as the time-dependent variation between consecutive heartbeats. By differentiating and responding rapidly to arbitrary incentives, HRV is meant to mimic the heart's adaptability. HRV testing may assess the state of the heart and the autonomic nerve system (ANS) that controls flexible heart rate variability. When it comes to sympathetic nervous system positioning, HRV is an appropriate indication. When discussing beat intermissions or HRV's prompt HR, HRV makes note of discrepancies. The usual

The circulatory system and the heart's autonomic neuronal regulation produce HR erraticism [3].

Despite their limitations, many neural network and randomization approaches to time series modeling and prediction perform well in forecasting applications [4, 5]. More and more people are looking to Support Vector Machines (SVMs), a novel statistical learning theory, for classification and prediction purposes. Although support vector machines (SVMs) were initially developed to address pattern classification issues in fields like text classification, face recognition, and optical character recognition, they quickly found widespread use in other domains, such as function approximation, regression estimation, and time-series predictions. Citation: Nima Hatami, 2017 [6].

Methods for Classifying Time Series

One rapidly expanding area of machine learning is time series classification (TSC) [7, 8]. Classifying a time series may be done in two ways. Using a certain approach is one of them. Long short-term memory (LSTM) and repeating neural networks are two cases in point. An additional option is to use the series' capabilities for regular supervised learning [9].

Vector Support Systems

When it comes to accurately classifying data that is either linearly divided or linearly related, support vector machines (SVMs) are a core formed family of methods in machine learning. Converting data that isn't linearly divisible to a developed dimensional space is the foundational skill when dealing with such data. kernel purpose. In this new space, the samples can typically be categorized with higher precision. Many kinds of kernel purposes have been advanced, with the greatest used ones being polynomial and circular-based.

Decision Tree

A decision tree is a flowchart-like tree building, where each interior node signifies a test on an aspect, apiece branch signifies an ending of the examination, class label is characterized by each leaf node. Assumed a tuple X, the characteristic values of the tuple are verified contrary to the decision tree. A route is drawing from the root to a leaf node which grips the class forecast for the tuple. It is easy to change decision trees into classification rubrics. Decision tree learning uses a decision tree as a prognostic model which maps explanations about an entry to deductions about the item's mark value. It is one of the prognostics demonstrating methods used in numbers, data mining and machine learning. Tree copies where the board flexible can take a limited set of values are called classification trees, in this tree construction, leaves signify class labels and branches characterize combinations of topographies that lead to those class labels. Decision tree can be built comparatively fast associated to other approaches of cataloguing. SQL announcements can be built from tree that can be used to access databases professionally. Decision tree classifiers gain alike or better precision when associated with other classification approaches. An amount of data mining methods has previously been done on instructive data mining to advance the presentation of pupils like Regression, Genetic algorithm, Bays classification, k-means gathering, subordinate guidelines, forecast etc. Data mining methods can be used in instructive arena to improve our sympathetic of learning procedure to emphasis on classifying, mining and assessing variables connected to the learning procedure of pupils. Classification is one of the greatest commonly [10].

Materials and Methods

Time characteristics are important for data processing data as the data under review provides this feature. This process can be used to support the decision-making process to extract relevant and interesting knowledge from a large data set, including knowledge [10]. Learning Machine (ML) helps support data mining. However, most ML methods do not deal directly with the time feature because they assume that the data are distributed independently and uniformly. However, since the data set is time-oriented, the occurrence of observation at a particular point in time usually depends on the values already observed [7, 11].

The proposed approach jointly applies two strategies to construct a feature display for time series attributes. On the one hand, some of the extracted features are typically related to descriptive statistics such as mean, standard deviation, and maximum and minimum, which provide information about the global behavior of a time series. After that, this representation will be the attribute value as input for ML algorithms [12].

Step 1: Time Series Preprocessing

In the first phase, the time series for processing some common problems in time data, such as differences in scale and time interval, data with noise; And the presence of missing values are preprocessed.

Step 2: Extraction Properties

In this step, the properties are identified using global feature extraction features and local descriptions from time-series data. Two independent stages make up this stage [13].

In this research first, the ECG signal data is prepared from the database and their preprocessing will be done to select the appropriate signals. Then the appropriate properties are extracted and based on these properties, the classification operation is performed.

Build a categorized model based on the labeled time series and then use that model to predict unlabeled time-series labels. To classify time series in python, you must first extract the properties from the time series data, which is done with the decision tree algorithm, and then use existing classification techniques such as SVM on that set of features.

Time series to anticipate some common problems in time data, such as differences in scale and time interval, data with noise; and the existence of missing values. Is used. Properties are then identified by extracting global properties and local descriptions from time-series data. Next, machine tree decision and learning algorithms are used to construct a prediction model. The choice of an algorithm should be according to the ultimate goal of the extraction patterns. The selected algorithms are then applied and the generated models can be evaluated using objective as well as qualitative methods.

In this research, the MIT-BIH natural cytosine rhythm database, MIT-BIH arrhythmia database, MIT-BIH atrial fibrillation database, and MIT-BIH malignant ventricular arrhythmia database were used. The MIT-BIH Arrhythmia Database contains 48 pieces of two-way ambulance ECG recordings, each recording lasting about 30 minutes at a sampling frequency of 360 Hz, the MIT-BIH Normal Rhythm Database includes 18 long-term ECG recordings, MIT-BIH Database consists of 25 long ECG cycles with a sampling frequency 250 HZ and the MIT-BIH malignant ventricular arrhythmia database containing 25 minutes recording at 250 Hz. A total of 53 ECG files includes 18 normal rhythms. Figure 1 shows the diagram of the proposed classification algorithm:

Datasets and Evaluation Specifics

In this study, two different datasets will be examined, including records obtained from healthy young and elderly people. The second data set includes people who are naturally at risk for cardiovascular disease.

An electrocardiogram was recorded from 25 men aged 32 to 89 years and 22 women aged 23 to 89 years, about 60% of whom were hospitalized. The signals are recorded from two channels, and due to the differences in the anatomical features of the individuals, lead II and V1 are used in most of the recordings. The portfolio frequency is 360 Hz. The database includes 48 half-hour electrocardiograms recorded in 24 hours out of 47 people.

Pulses are monitored and marked using a QRS tilt detector. Each signal is then interpreted by two cardiologists. Approximately 110,000 beats of this path have been examined and their type has been identified.

We first receive the data from the following:
Address https://www.physionet.org/physiobank/database/mitdb/
Discussion and Results
The third dataset is attained by mining HRV features from MIT- BIH databases, which contain: MIT-BIH Normal Sinus Rhythm, Normal Sinus Rhythm RR Interval, MIT-BIH Arrhythmia, MIT- BIH Supraventricular Arrhythmia, BIDMC Congestive Heart Failure, and Congestive Heart Failure RR-interval. The normal and abnormal sample signals in the MIT-BIH database are shown in the following figures, respectively. Before extracting the signal properties, all ECG signals are destroyed using a simple medium filter. The output of the feature extraction using the decision tree can be seen in the figure below.

The version of normal signals is shown in Figure 4a and the abnormal signals are shown in Figure 4b. The noisy version of normal signals is shown in Figure 4c.

Evaluation Specifics
For assessment of the classifiers on each dataset, we use 10x10- fold cross-validation. Assessment measures used are normal in BTS examines: sensitivity (SENS), and specificity (SPEC):

Where TP, TN, FP, and FN are the records of: true positives, true negatives, false positives, and false negatives, correspondingly. For multiclass event, these measures can be gained from the misperception matrix by associating amounts of examples for each class in the matrix beside cases of all the other classes. The described values have been weighted and averaged amongst classes.

Factors of the algorithms were altered in order to gain the best possible consequence using methodical method on the first 10-fold iteration.

In summary, for evaluation purposes, we use standard criteria in the field of medicine, namely General Classification Accuracy (ACC), Sensitivity (SENS), Specificity (SPEC), and Positive Predictive Value (PPV).

Finally, the classification of a signal is analyzed using the SVM classifier with the help of statistical features extracted from the recorded version of the input ECG. From k-fold, cross-validation is used. Figure 5 ECG signal classification shows the analysis of the ROC curve and the tangled matrix. It can be seen from Figure 5 that the ECG system also offers 90% in terms of sensitivity, specificity, and accuracy.

In this research, ECG signal classification has been analyzed using feature extraction. MIT-BIH arrhythmia database records are used for classification work. First, the processing is performed using a medium filter and then statistical properties are extracted. Finally, DT-SVM-based classification was used to classify the signals. Experimental results show the performance of the ECG signal classification system with promising results. The DT-SVM classifier accurately classifies 90% of the given ECG signal with simple statistical properties.

Comparison of DT-SVM with models used in other studies Neural Networks
These are models of information processing, as the name implies,

in this way the information of the human nervous system is processed. An important aspect of this information model is its unique structure. Many highly interconnected processing paradigms (commonly referred to as neurons) work together to solve very specific problems.

The neural network model has been trained several times to change the number of hidden layers as well as the decay factor, and the light adjustment for maximum accuracy is best achieved with 1 hidden layer and 13 neurons. The results of neural network classification are shown in Table 1 [14].

On average, prediction (in case of regression) of individual trees works. This is a simulation of prediction trees (decision trees) in which the result of each tree depends on the value of a randomly sampled vector independently with the same distribution for all forest trees. Random forests try to reduce the high variance problems often seen in single-decision trees by reducing the average balance between the two boundaries.

In The case of classification problems, according to a set of random predictor variables and simple trees, the Random Forest algorithm determines a margin function, which determines the degree of an average number of votes for the actual class that is higher than the average number of votes for each other class many classification trees grow in the dependent variable of random forests. If we want to classify a new object, we place the input vector under each tree and then classify each tree. So, each tree belongs to a specific class. The class with the highest number of votes is selected as the result of a random forest. Table 2 shows the results obtained from the stochastic forest classification, Anish batra [15-17].

Conclusion
For the purpose of analyzing and classifying ECGs, a MATLAB application is made available [18]. The project's goal of developing a software that can extract features and classify them as normal or abnormal has been accomplished. In the feature identification part of this research, we have used split samples collected from the MIT-BIH site to classify using the MATLAB programming language. using the decision tree function to choose the ECG signal time series [19, 20].

Future works

For future work, the weight of each selected feature will be determined in the ECG signal classifier and will be expanded on future findings in future work by evaluating information amplification techniques performed by time series analysis of different heart rate types. Comparisons will also be made with other machine learning methods, which include assessment of random forests, cannulation neural networks, and other methods.

REFERENCES

Circuits and Systems, 2003. ISCAS ‘03. Proceedings of the 2003 International Symposium 2003: 741–744. 1. “Application of critical support vector machine to time series prediction” by T. Raicharoen, C. Lursinsap, and P. Sanguanbhoki.

According to Saul JP (1990), the regulation of cardiac autonomic outflow is reflected in the beat-to-beat changes in heart rate. Notices in Physiology and Experimental Biology 5: 32–37.

Alves, Myrela, Garner, David M., Fontes, Anne M., Sousa, Luiz Vinicius de, and Valenti, Vitor E. (2018) Heart Rate Variability Measurements in Healthy Women: Linear and Complex (WILEY 2018: 14) in Response to Traffic Noise.

Fourthly, in 2018, Fazle Karim, Somshubra Majumdar, Houshang Darabi, and Shun Chen developed Lstm fully convolutional networks to classify time data. 1662–1669. (IEEE Access, 2006).

5. A framework for time series categorization based on distance, by Kaya and Gunduz-oguducu (2015). Journal of Information Systems, 51(1), 27–42.

Time-series classification using a bag of recurrence patterns was presented by Nima Hatami in 2017.

Research on the dynamics of interbeat interval time series in healthy and congestive heart failure participants using a scale-based symbolic entropy analysis was conducted by Imtiaz Awan, Wajid Aziz, and Imran Hussain Shah (2018). published in 2018 by PLOS One, pages 1–18.

“A neural network ensemble method with jittered training data for time series forecasting” (ISBN: 1773-5329) by G.P. Zhang is the source for this information.

9. Faizal Mahananto (2017) A Basic Thematic Model of In 2017, patients with congestive heart failure were identified using heart rate variability. Its namesake

10. Sunil Kumar and Himani Sharma (2013) Information Mining Decision Tree Algorithms: A Comprehensive Review, 2013 International Journal of Science and Research (IJSR).

Classification of healthy and diseased individuals’ heart rate data using threshold-based symbolic entropy, 2014, by Wajid Aziz 11.

IEEE Transactions on Biomedical Engineering, vol. 32, no. 2, 1985, pp. 230–236, by J. Pan and W. J. Tompkins.

13. M Adnane, Z Jiang, S Choi (2000) (Comp. Meth. Prog. Biomed 93: 20-31), Developed a QRS detection algorithm for a wearable cardiorespiratory device.

The interaction between heart rate and respiration: part II - nonlinear analysis based on computer modeling was published in Automatica in 1982 by R. Kitney, D. Linkens, A. Selman, and A. McDonald, and can be found in volume 4, pages 141–153.

15. In the 2016 International Journal of Biology and Biomedicine, Anish Batra wrote on how to classify arrhythmias using a combination of machine learning algorithms and electrocardiogram diagnostic criteria.

16. “Real-time patient-specific ECG classification by 1-D convolutional neural networks” (IEEE Transactions on Biomedical Engineering 63: 664-675, 2015). Authors: S Kiranyaz, T Ince, and M Gabbouj.

IEEE Communication, Control, and Intelligent Systems 2015: 190-193, “On the aspect of feature extraction and classification of the ECG signal” by S Basu and YU Khan (2015).

18. “Classification of ECG signals using machine learning techniques: A survey” was published in the 2015 IEEE International Conference on Advances in Computer Engineering and Applications and was authored by SH Jambukia, VK Dabhi, and HB Prajapati. The paper can be found on pages 714-721.

19. In the 2015 IEEE International Conference on Automation, Cognitive Science, Optics, Micro-Electro-Mechanical System, and Information Technology, A. Rizal and S. Hadiyoso presented a paper titled “ECG signal classification using Hjorth Descriptor” (pp. 87–90).

Time series classification utilizing motif and characteristic extraction: a case study on ECG databases was published in 2013 by André G. Maletzke, Huei D. Lee, and Gustavo E.A.P.A. Batista. Proceedings of the Fourth International Workshop, Eureka-2013.