
Amin Janghorbani *, Abdullah Arasteh **, Mohammad Hassan Moradi ***
*Amirkabir University of Technology, a.janghorbani@aut.ac.ir
**Sharif University of Technology, arasteh@ee.sharif.ir
***Amirkabir University of Technology, nhmoradi@aut.ac.ir

Abstract: Acute hypotension episodes are one of the hemodynamic instabilities with high mortality rate that is frequent among many groups of patients. Prediction of acute hypotension episodes can help clinicians to diagnose the cause of this physiological disorder and select proper treatment based on this diagnosis. In this study new physiological time series are generated based on heart rate, systolic blood pressure, diastolic blood pressure and mean blood pressure time series. Statistical features of these time series are extracted and patients whom are exposed to acute hypotension episodes in future 1 hour time interval and whom are not, are classified based on these features and with the aid of Logistic Regression (LR) model and Support Vector Machine (SVM) classifiers. The best accuracy of classification was 88% with applying SVM classifier and based on selected features which were selected with genetic algorithm.

Keywords: Acute hypotension Episodes, Prediction, Logistic Regression Model, Support Vector Machine, Genetic Algorithm

1. Introduction

Hypotension is a clinical condition in which patient's blood pressure has abnormally low values below 60 mmHg. If the blood pressure stays down for a long time, as a result of inadequate tissue perfusion a clinical syndrome occurs this is named “shock”. If shock persists for a long time inadequate oxygen delivery leads to irreversable cell injury and even can cause patient’s death [1].

Several groups of patients are exposed to hypotension. 22% of patients in MIMIC II database experienced hypotension episodes and the mortality rate of this group is more than twice of MIMICII population as a whole [2]. In addition, hypotension is one of the most frequent side effects of spinal anesthesia with incidence of 33% [3] and hemodialysis with incidence of 15-25% [4].

Delaying in treatment of acute hypotension episodes may result in organ damage and death. Timely and rapid intervention can help to save patient’s life. Determining proper treatment depends on diagnosing the cause of acute hypotension which might be sepsis, myocardial infarction, cardiac arrhythmia, pulmonary embolism, hemorrhage, dehydration, anaphylaxis and also anything which causes hypovolemia. To this end, some clinical examinations are required which normally are very time consuming. Therefore, prediction of acute hypotension at the right time and before onset of the disease can help clinicians to select an effective and suitable treatment without exposing the patient to the risks of delay in receiving treatment.

Chiarugi [5] extracted significant features from arterial blood pressure and heart rate time series and used decision tree and SVM for classification of these features in order to predict acute hypotension episodes. However, Jousset [6] applied the SVM to classify the statistical features from biological time series and then could predict which patient will experience acute hypotension episodes within a forecast window of 1 hour.

Mneimneh [7], though, used a rule based approach for prediction of acute hypotension episodes. This paper demonstrates how acute hypotension episodes can be predicted in the next 1 hour time interval. In order to achieve this aim, several new physiological parameters based on initial set of four primary physiological time series (heart rate (HR), systolic arterial pressure (SAP), diastolic arterial pressure (DAP) and mean arterial pressure (MAP)) were generated. Then statistical features of them were classified with the aid of LR model and the SVM. The result of classification facilitated predicting which patient is exposed to acute hypotension episodes in the next 1 hour time interval from predefined time origin (T0). Finally the classification accuracy of the two methods was compared together and with previous studies.

In the following section the database and methods of the application are described in detail. In section 3, the results of methods are presented and discussed. Finally section 4 concludes the paper with a summary of the study and a comparison between the findings of this research and the results of previous studies.

2. Methods

2.1 Database:
Database which was used in this study is from the 10th annual physionet/computer in cardiology challenge on predicting acute hypotension episode in the intensive care unit (ICU). This database is comprised of two groups of patients, group H (patients with acute hypotension episode in forecast window) and group C (patients without acute hypotension episode in forecast window) [8]. 95 patients record of training set and test set B of the challenge were investigated in this study in order to predict acute hypotension episodes.  The records include at least the data of HR, SAP, DAP and MAP time series. Train and test set B of the challenge were used in this study for prediction of acute hypotension episodes.

2.2 Feature Extraction
Some additional physiological time series were generated based on initial physiological time series of database records (HR, SAP, MAP and DAP). These new time series are [9]:

Some additional physiological time series were generated based on initial physiological time series of database records (HR, SAP, MAP and DAP). These new time series are [9]:

1. \( |dHR| \): the absolute successive difference of HR data sample
2. \( HR \times BP \): heart rate blood pressure product which is an index for cardiac oxygen consumption
3. \( HR / BP \): heart rate to blood pressure ratio which is known as an early indicator of cardiogenic, septic and hypovolemic shock
4. ECO: Estimated cardiac output that is calculated from Eq.1
   \[
   \frac{SAP - DAP}{MAP} \times HR
   \]  
5. Slope: slope of linear least square regression line to a moving window (30 min) for ECO, SAP, MAP and DAP.

Statistical features of these new and the primary physiological time series such as minimum, maximum, mean, standard deviation , skewness, (5th, 10th, 50th, 90th, 95th) percentiles and inter percentile range(IPR) - which is difference between two percentiles values such as 10th percentile and 5th percentile - are extracted from physiological time series and classification was performed based on these features.

2.3 Feature Selection and Classification with LR
Logistic Regression model is one of the regression methods which are used to find the relationship between an outcome variable and a set of independent variables, called predictor variables. Specific form of Logistic Regression Model is indicated in Eq.2, in which each \( x \) is a predictor variable and \( y \) is the response variable in interval \([0,1]\).

\[
y = \frac{e^{\beta_0 + \beta_1 x_1 + ... + \beta_n x_n}}{1 + e^{\beta_0 + \beta_1 x_1 + ... + \beta_n x_n}}
\]  

The parameters of the model (\( \hat{\beta} \)) are estimated in training process of the model using Maximum Likelihood Estimation (MLE).

First a single variable LR model for each feature was developed. It means every feature was assumed as a predictor or explanatory variable and the output of the model was the label of patient which was 1 for H group and 0 for C group. In classification output of the model for each feature was calculated. 0.5 was intended as the threshold for occurrence of hypotension episodes in forecast window, values greater than 0.5 were labeled as H patients and values smaller than 0.5 were labeled as C patients. T-student statistical test was used to examine the significance of each coefficient in the model and only the features with statistically significant model coefficients (\( P < 0.05 \)) were selected for classification.

In addition, multivariable LR models were developed for various subsets of the selected features in order to evaluate the performance of combination of statistically significant features. Also, a genetic algorithm was used for feature selection in order to ensure that the entire extracted features were searched rigorously. Genetic algorithms are a class of techniques for the solution of the optimization problems based on population of solutions which are developed through successive generations by means of the application of three genetic operators: selection, cross over and mutation [10]. There is a straightforward method for using genetic algorithm in feature selection. In this method, chromosomes are coded with binary codes that each bit indicates existence or absence of a feature in the selected feature vector. For example if \( i^{th} \) bit is 1, \( i^{th} \) feature would be selected and if \( i^{th} \) bit is 0, \( i^{th} \) feature would be omitted. It is obvious that the most fitted feature would be selected and if \( i^{th} \) bit is 0, \( i^{th} \) feature would be omitted. It is obvious that the most fitted chromosome is the one that maximizes the accuracy of classification based on its associated feature vector. So in each generation, based on binary bits of each chromosome, the selected features of all feature vectors are given to the classifier and the overall accuracy of the classification is determined as the output of the fitness function of each chromosome and is used for production of the next generation.

Fig.1 shows the improvement of classification accuracy with LR model during the generations. A summary of genetic algorithm parameters that was used for feature selection in this study are given in Table 1.

2.4 Feature selection and classification with SVM
SVMs originally were introduced by Vapnik within the
area of statistical theory and structural risk minimization. SVMs were used for many applications such as regression and pattern recognition [11].

![Accuracy of classification with the LR during various generations of genetic algorithm](image)

**Fig.1: Accuracy of classification with the LR during various generations of genetic algorithm**

<table>
<thead>
<tr>
<th>TABLE I: Parameters of Genetic Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coding of chromosome</td>
</tr>
<tr>
<td>Size of population</td>
</tr>
<tr>
<td>Cross over method</td>
</tr>
<tr>
<td>Cross over rate</td>
</tr>
<tr>
<td>Mutation method</td>
</tr>
<tr>
<td>Mutation rate</td>
</tr>
<tr>
<td>Reproduction</td>
</tr>
<tr>
<td>Reproduction rate</td>
</tr>
</tbody>
</table>

In order to improve the accuracy of classification, a SVM was used to classify the feature vectors. The SVM generalization performance depends highly on good tuning of kernel function parameters and the soft margin parameter. So, if the parameters of the SVM are tuned properly, generalization error would be minimized substantially. In this work a SVM with RBF kernel was used for classification. Optimization of classifier parameters and feature selection was done simultaneously with genetic algorithm. Feature selection process was the same as described method in 2.3 section, but for tuning the SVM parameters (kernel function parameter and soft margin parameter), 40 bit binary codes were added to each chromosome (20 bit for each one) and these parameters were coded as 20 bit binary codes. Finally for each chromosome the classification was done based on its selected features and the selected parameters for SVM. Fig.2 clarifies improvement of classification accuracy with SVM during various generations of the genetic algorithm.

3. **Results**

Evaluating the accuracy of the classification was performed according to Leave One Out (LOO) cross validation method. The LOO is one of the most reliable cross validation methods in which a patient is selected as test set while all the others are used as training set. Then the single test patient is classified with the trained classifier. All of the patients would be tested in the same process. Ultimately, the overall accuracy for all of the data set was counted by calculating the classification error of the classification. The best results for LR classifier and feature selection according to the statistical significance of model coefficients, was 80%.

![Accuracy of classification with the SVM during various generations of genetic algorithm](image)

**Fig.2: Accuracy of classification with the SVM during various generations of genetic algorithm**

The selected features of the methods are listed in table 2. Applying genetic algorithm for feature selection improved accuracy of classification with LR. Among all of the features, 10 were selected in this method which produced 86% accuracy. The best result for prediction of acute hypotension episodes was achieved when SVM with RBF kernel was used as classifier and the genetic algorithm was applied for optimization of kernel function and learning parameters and feature selection. The accuracy of this method was 88% with selection of 7 features with genetic algorithm.

4. **Conclusion**

In this study statistical features of physiological time series were extracted and were classified with LR model and SVM classifier. Moreover, a genetic algorithm was used for feature selection and optimization of SVM parameters. The best accuracy of classification by this method with LOO cross validation and SVM classifier was 88% while in the research which was conducted by Mneimneh [7] the classification accuracy of training set was 68.3% and for the test set B was 90%. Meanwhile, the classification accuracy of training set with LOO cross
validation method which was tested by Chiarugi [5] was 83% and for the test set was 75%. However, Jousset [6] concluded 82% accuracy in classification in training set and 75% in test set B. Considering the results of the similar studies confirms that applying the SVM as classifier and genetic algorithm for feature selection can facilitate the prediction of acute hypotension episodes in the next one hour time interval with a higher rate of accuracy.

TABLE II: Results of Various Methods and the Selected Features in each method

<table>
<thead>
<tr>
<th>Method</th>
<th>Selected feature</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR model &amp; feature selection based on statistical significance of model coefficients</td>
<td>10th percentile of DAP mean of MAP maximum of ECO</td>
<td>80%</td>
</tr>
<tr>
<td>LR model &amp; GA feature selection</td>
<td>10th percentile of DAP maximum of ECO 95th percentile of ECO skewness of dHR mean of MAP mean of ECO slope 5th percentile of ECO slope mean of MAP slope mean of DAP slope</td>
<td>86%</td>
</tr>
<tr>
<td>SVM &amp;GA feature selection</td>
<td>HR/SAP IPR10-5* 50th percentile of HR/MAP 95th percentile of dHR skewness of dHR mean of MAP slope 5th percentile of MAP slope 5th percentile of SAP slope</td>
<td>88%</td>
</tr>
</tbody>
</table>

References

[10] Vahid Aboualebi, Mohammad Hassan Moradi, Mohammad Ali Khalilzadeh " A new approach for EEG feature extraction in P300-based lie detection" Elsevier computer methods and programs in biomedicine 9 4 ( 2 0 0 9 ) 48–57

All in-text references underlined in blue are linked to publications on ResearchGate, letting you access and read them immediately.