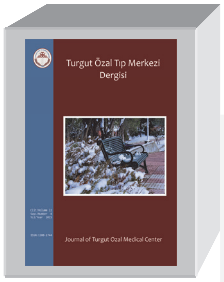

ORİJİNAL MAKALE/ORIGINAL ARTICLE
J Turgut Ozal Med Cent 2015;22(4):221-4

Journal of Turgut Ozal Medical Center
www.jtomc.org

Myocardial Infarction Classification with Support Vector Machine Models

Emek Güldoğan[1], Jülide Yağmur[2], Saim Yoloğlu[1], Musa Hakan Asyalı[3], Cemil Çolak[1]

[1]İnönü University, Faculty of Medicine, Department of Biostatistics and Medical Informatics, Malatya, Turkey
[2]İnönü University, Faculty of Medicine, Department of Cardiology, Malatya, Turkey
[3]International Antalya University, Faculty of Engineering, Department of Biomedical Engineering, Antalya, Turkey

Abstract
Aim: Support vector machines (SVM) is one of the classification methods that aims to find the best hyper-plane separating a space into two parts with known positive and negative samples. The goal of this study is to classify myocardial infarction (MI) using SVM models.
Material and Methods: The data used in the MI classification contains information related to 184 individuals which is randomly taken from the database created for the Department of Cardiology, Faculty of Medicine, Inonu University. Estimated SVMs are models generated from the SVM-linear and SVM-Radial Based kernel functions.
Results: In this study, 90 individuals of the study group (48.9%) are MI patients, while 94 (51.1%) patients are not. The classification success rate is 83.70% for SVM-linear model and 90.76% for the SVM-Radial Based model.
Conclusion: In this study, it is observed that SVM-Radial based model presented a better classification performance than the linear SVM model. The use of SVM models based on various kernel type functions can improve disease classification performance.
Keywords: Support Vector Machines; Myocardial Infarction; Classification.

Destek Vektör Makinesi Modelleri ile Miyokard İnfarktüsün Sınıflandırılması

Özet
Amaç: Destek vektör makinesi (DVM), pozitif ve negatif örnekleri bilinen bir uzayı ikiye ayıran en iyi hiper-düzlemi bulmaya çalışan sınıflandırma yöntemlerinden biridir. Bu çalışmada, Miyokard İnfarktüsün (Mİ) DVM modelleri ile sınıflandırılması amaçlanmaktadır.
Gereç ve Yöntemler: Sınıflandırmada kullanılacak Mİ verileri, İnönü Üniversitesi Tıp Fakültesi Kardiyoloji anabilim dalı için oluşturulan veritabanından rastgele alınan 184 bireye ilişkin bilgileri içermektedir. Tahmin edilen DVM'ler, DVM-Doğrusal ve DVM-Radyal Tabanlı kernel fonksiyonlarından oluşturulan modellerdir.
Bulgular: Çalışmada incelenen grubun 90'ı (%48.9) Mİ hastası iken, 94'ü (%51.1) Mİ hastası değildi. Mİ'nin doğru sınıflama başarısı, DVM-Doğrusal modeli için %83.70 ve DVM-Radyal Tabanlı modeli için %90.76 olarak elde edilmiştir.
Sonuç: Bu araştırmada Mİ'nin sınıflanmasında radyal tabanlı DVM modelinin, doğrusal DVM modelinden daha iyi sınıflama performansı gösterdiği belirlenmiştir. Farklı kernel tipi fonksiyonlara dayalı DVM modellerinin kullanımı, hastalıkların sınıflama performansını artırılabilir.
Anahtar Kelimeler: Destek Vektör Makineleri; Miyokard İnfarktüsü; Sınıflama.

INTRODUCTION

The ischemic necrosis of the myocardia is called myocardial infarction (MI). Despite developments in medical and interventional treatment in the last 40 years, acute myocardial infarction is still an important cause of mortality and morbidity. Acute myocardial infarction is often caused by the interruption of coronary blood flow in narrowed arteries due to previous atherosclerosis. Rupture and erosion of vulnerable atherosclerotic plaque, effective transformation of platelets, and clot formation are pathophysiological processes leading to acute coronary syndrome (1).

Decision support systems, which are also called machine learning methods, help estimate corresponding output for the given input by using previous knowledge and experience in cases that cannot be clearly defined. Support vector machine (SVM) is a machine learning method based on powerful statistical theories. It was first offered by Vapnik in 1995 for classification and regression type of problem solutions (2).

Following successful applications of machine learning (data mining), support vector machines have become a successful method of classification in the field of remote sensing. This classification method is based on the principle of dividing two classes of data from each other on a hyperplane by converting data into larger sizes. In the high size conversion stage, functions with different features are employed. These functions are called kernel functions. For the use of the kernel functions, some parameters in the mathematical expression of these functions must be set by the user (3).

Nowadays, SVM is adaptable to many world problems, and this increases the interest in the SVM method along with the number of studies applying this method in all areas. SVM is widely used in several pattern recognition applications like image and text classification, object recognition, handwriting recognition, voice recognition, and face recognition. SVM also shows a rising success for biological applications. In medicine, it is used in the morphology of cancer, success of therapy, determining related genes, and diagnosis of various diseases (4).

In this study, we aim to present a classification of myocardial infarction with support vector machine models.

MATERIALS and METHODS

The data of 184 patients with and without MI were randomly retrieved from the Department of Cardiology (Inonu University, Faculty of Medicine) database. The study was based on the following variables: age, gender, height, weight, body mass index (BMI), hypertension (HT), diabetes mellitus (DM), hyperlipidemia (HPL), family history (FH), smoking, glucose, cholesterol, high density lipoprotein (HDL) cholesterol, low density lipoprotein (LDL), and triglycerides. Estimated SVMs made up the the SVM-linear-models and SVM-radial basis kernel functions. The kernel functions employed in this study are as follows (5):

SVM-linear: $$K(x,z) = (x \cdot z)$$

SVM-Radial-based: $$K(x,z) = \exp\left(-\frac{\|x - z\|^2}{2\sigma^2}\right)$$

σ is the user-defined positive real number and refers to the width. x and z are the variables which define the kernel functions. Support vector machine models were created by using IBM SPSS Modeler Professional 14.2 for Windows. The success of the established support machine models in classifying MI data was evaluated with the correct classification rates.

RESULTS

90 patients (48.9%) in the study group were MI patients, the remaining 94 (51.1%) patients were not MI patients. Descriptive statistics for the data used in this study are shown in Tables 1 and 2. The significance of the variables generated from the results of models based on SVM-linear and SVM-radial based kernel functions is presented in Table 3.

Table 1. Descriptive statistics of quantitative variables.

Variables	MI Patients (n=94)	Non-MI Patients (n=90)
BMI [Mean±SD]	27.58±3.96	26.29±3.69
Height [Mean±SD]	1.67±0.07	1.70±0.07
Glucose [Mean±SD]	105.02±27.31	145.19±65.25
HDL [Mean±SD]	41.23±9.72	40.72±5.41
Weight [Mean±SD]	76.63±13.90	75.51±10.74
Cholesterol [Mean±SD]	194.56±36.74	196.24±31.38
LDL [Mean±SD]	122.26±29.67	128.98±28.40
Triglycerides [Mean±SD]	166.25±79.45	161.91±50.13
Age [Mean±SD]	52.73±10.08	59.29±12.44

BMI: Body mass index; HDL: High density lipoprotein ; LDL: Low density lipoprotein

Table 2. Descriptive statistics of qualitative variables.

Variables	MI Patients (n=94)	Non-MI Patients (n=90)
Gender (Male), {n(%)}	35(66.0)	18(34.0)
HT (Present), {n(%)}	29(48.3)	31(51.7)
DM (Present), {n(%)}	12(46.2)	14(53.8)
HPL (Present), {n(%)}	39(79.6)	10(20.4)
FH (Present), {n(%)}	35(46.1)	41(53.9)
Smoking (Present), {n(%)}	26(35.1)	48(64.9)

HT: Hypertension; DM: Diabetes mellitus; HPL: Hyperlipidemia; FH: Family history.

Table 3. Significance values of the variables.

Variables	Model	
	SVM-linear	SVM-Radial-based
HPL	0.21	0.24
Smoking	0.20	0.14
Glucose	0.19	0.13
Gender	0.08	0.09
Height	0.04	0.07
Age	0.09	0.06
DM	-	0.05
HT	-	0.05
Weight	-	0.05
Triglycerides	-	0.03
LDL	0.05	-
FH	0.04	-
HDL	0.01	-
Cholesterol	0.01	-

DISCUSSIONS

Assessment of the findings showed that correct MI classification success rate was 83.70% for the SVM-linear model and 90.76% for the SVM-radial basis model, respectively. The analysis of the significance rates of the SVM-linear and SVM-radial based models has shown that the most significant variable values were that of hyperlipidemia with 0,21 and 0,24, respectively. With 0,20 and 0,14, smoking took the second place. The third important variable was glucose with 0,19 and 0,13. Other significant variables in the SVM-linear model, from more important to less important, were age, sex, LDL, height, family history, HDL, and cholesterol, respectively. Similarly, other important variables for the SVM-radial model were gender, height, age, diabetes mellitus, hypertension, weight, and triglycerides, respectively.

SVM is successful in large applications with less data. Because of this, SVM, or data mining, has been used in many application areas such as customer fraud detection and image classification (6).

Cardiovascular disease risk is the result of an interaction of multiple risk factors. There are evaluation systems that estimate whether the patient is more likely to develop coronary artery disease or myocardial infarction by using these risk factors and allow risk classification for cardiovascular events as low, medium, and high (7). High serum total and LDL cholesterol as well as low serum HDL cholesterol are among major independent risk factors for CADs. High total and LDL cholesterol levels mean higher risk of atherosclerotic events. In societies with relatively high levels of average cholesterol, low HDL cholesterol is a strong indication suggestion CADs (8).

Hypertension is responsible for 35% of all atherosclerotic cardiovascular events. Coronary heart disease is 2-3 times higher in hypertensive patients than normotensive patients. Hypertension increases the risk of acute myocardial infarction 2-3 times both in men and women. A-15-mmHg increase in the diastolic blood pressure or a-25-mmHg rise in systolic blood pressure increases the risk of renewed infarction by 40% and 37%, respectively (9). In this study, an association between MI and hypertension has been determined; however, the fact that it was not among the effective risk factors can be explained by the different results due to the different SVM model employed.

Age is a strong and independent risk factor for developing atherosclerosis and myocardial infarction. In men, risk increases with each decade of life. Compared to premenopausal women, men are more prone to CADs by approximately 10 years earlier (8). As long as atherosclerosis-related events are concerned in societies with both high and low risks, smoking is a major but intervenable risk factor. In patients who have reached the level of atherosclerosis, it increases the risk of thrombotic events; it is also a strong predictor of myocardial infarction development. Giving up smoking is probably the most effective of all secondary prevention measures (10).

Among all the risk factors, family history has been an important independent risk factor at all stages of atherosclerosis for CAD development. To evaluate family history is the most important method in identifying genetics suspected coronary artery diseases. The most common family history bringing about risk of coronary artery disease is the presence of coronary heart disease at an early age in first-degree relatives. Development of CAD before the age of 55 in first-degree male relatives and before 65 in first-degree female relatives is an indication of approximately 2-fold risk of atherosclerosis. Therefore, along with traditional risk factors, family history should be considered in risk analysis (8). Many clinical study have shown that diabetes is a crucial risk factor in the development of atherosclerosis and myocardial infarction. Compared to non-diabetic patients, diabetic patients have a 2-6-fold risk of cardiovascular event induced death (11).

To conclude, radial based SVM has shown a better classification performance than the linear SVM model in the classification of MI. The use of SVM models based on different kernel type functions may improve disease classification performance.

This article was presented at the 14th National Biostatistics Conference (including international presenters) that was held between 4-7 October 2012 in Kayseri, Turkey, and was published in the Book of Abstracts.

REFERENCES

1. Kurtul A. Akut miyokard infarktüsünde patofizyoloji nedir? klinik bulguları nelerdir? Turkiye Klinikleri Journal of Cardiology Special Topics 2009;2(2):1-6.
2. Çolak MC, Çolak C, Kocatürk H, Sağıroglu Ş, Barutçu İ. Predicting Coronary Artery Disease Using Different Artificial Neural Network Models. Anadolu Kardiyoloji Dergisi 2008;8:249-54.
3. Gülkesen KH, Köksal IT, Bilge U, Saka O. Comparison of methods for prediction of prostate cancer in Turkish men with PSA levels of 0-10 ng/mL. J Buon 2010;15(3):537-42.
4. Akşehirli ÖY, Ankaralı H, Aydın D, Saraçlı Ö. Tıbbi Tahminde Alternatif Bir Yaklaşım: Destek Vektör Makineleri. Türkiye Klinikleri Biyoistatistik Dergisi 2013;5(1):19-28.

5. Kavzoglu T, Colkesen I. A Kernel Function Analysis for Support Vector Machines for Land Cover Classification. International Journal of Applied Earth Observation and Geoinformation 2009; 11(5):352-9.
6. Güner N, Çomak E. Mühendislik Öğrencilerinin Matematik I Derslerindeki Başarısının Destek Vektör Makineleri Kullanılarak Tahmin Edilmesi. Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi 2011;17(2):87-96.
7. Conroy RM, Pyorala K, Fitzgerald AP, Sans S, Menotti A, De Backer G, et al. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. Eur Heart J 2003;24:987–1003.
8. Hamm CW, Möllmann, H, Bassand JP, Van de Werf F. Acute Coronary Syndrom.In: Camm AJ, Lücher TF, Serruys PW,editors. The ESC Textbook of Cardiovascular Medicine. 2nd ed. New York: Oxford Universty Press; 2009. p535-97.
9. Kannel WP. Blood pressure as a cardiovascular risk factor:prevention and treatment. JAMA 1996;275:1571-76.
10. Thomson CC, Rigotti NA. Hospital- and clinic-based smoking cessation interventions for smokers with cardiovascular disease. Prog Cardiovasc Dis 2003;45:459-79.
11. Gaede P, Vedel P, Larsen N, Jensen GV, Parving HH, Pedersen O. Multifactorial intervention and cardiovascular disease in patients with type 2 diabetes. N Engl J Med 2003;348(5):383-93.

Received/Başvuru: 15.12.2014, Accepted/Kabul: 30.04.2015

Correspondence/İletişim

Emek GÜLDOĞAN
İnönü Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, MALATYA, TÜRKİYE
E-mail: emek.guldogan@inonu.edu.tr

For citing/Atıf için

Guldogan E, Yagmur J, Yologlu S, Asyali MH, Çolak C. Myocardial infarction classification with support vector machine models. J Turgut Ozal Med Cent 2015;22:221-4
DOI: 10.7247/jtomc.2014.2671