Machine Learning Techniques in Diagnosis of Pulmonary Embolism

Pulmoner Emboli Tanısında Makine Öğrenmesi Teknikleri

Acilde Yapay Zeka / Artificial Intelligence in Emergency

Göksu Berıkol[1], Serdar Kula[2], Oktay Yıldız[3]
[1]Acil Tıp Kliniği, Karaman Devlet Hastanesi, Karaman,
[2]Çocuk Sağlığı ve Hastalıkları A.D. Pediatrik Kardiyoloji B.D. Gazi Üniversitesi Tıp Fakültesi, Ankara,
[3]Bilgisayar Mühendisliği A.D. Gazi Üniversitesi Mühendislik Fakültesi, Ankara,Türkiye

Özet

Amaç: Pulmoner emboli (PE), tanıda klinik şüphe ve tanısal laboratuvar ve görüntüleme sonuçlarının yüksek öneme sahip olduğu, yüksek mortalitesi olan bir hastalıktır. Bazı vakalarda antikoagülasyon ve fibrinolitik tedavilere kara vermek zor olmakta ve bu nedenle erken tanı acil tıp açısından önem arz etmektedir.Gereç ve Yöntem: Çalışma Ocak 2010 ve Ekim 2013 arasında acil servise nefes darlığı ve göğüs ağrısı da dahil olmak üzere akciğer şikayetleri ile başvuran 201 hastanın retrospektif kayıtlarına dayanarak tasarlanmıştır. Makine öğrenmesi teknikleri PE tanısında başarı hesaplanması için kullanıldı. Bulgular: PE tespiti için sınıflandırma ağacı yönteminin başarı oranının (%95), KNN sınıflaması (% 75) ve Naive Bayes Sınıflandırmasına (% 88.5) göre daha yüksek olduğu saptandı. Tartışma: Özellikle tanı konmasının zor olduğu hastalarda ve tetkiklerin kısıtlı olduğu acil servislerde, Sınıflandırma ağacı ve Bayes yöntemi gibi makine öğrenmesi teknikleri teşhis veya pulmoner emboli olasılığını tanımlamak için seçilebilir.

Anahtar Kelimeler

Pulmoner Emboli; Makine Öğrenmesi; Yapay Zeka

Abstract

Aim: Pulmonary embolism (PE), is a high mortality disease which clinical suspicion and a variety of diagnostic laboratory and imagingresults have a high importance in diagnose. Anticoagulation and fybrinolytic treatments are hard to decide in some cases therefore early diagnose is important in emergency medicine.Material and Method: The study was designed retrospectively based on the records of the 201 patients who were presenting to Emergency Department with pulmonary complaints including dyspnea and chest pain between January 2010 and October 2013. Results: Machine learning techniques were used for calculating the success in diagnosing PE. The success rate of the classification tree method for detection of PE was 95%, which was higher than that of KNN classification (75%) and Naive Bayes Classification (88.5%). Discussion: Classification tree and Bayesian method can be selected ones to diagnose or define possibility of pulmonary embolism in emergency centers with limited study tests and for the patients difficultly diagnosed.

Keywords

Artificial Intelligence; Machine Learning; Pulmonary Embolism

DOI: 10.4328/JCAM.3404 Received: 20.03.2015 Accepted: 24.04.2015 Printed: 01.12.2015 J Clin Anal Med 2015;6(suppl 6): 729-32
Corresponding Author: Göksu Berikol, Acil Servis Dept. Karaman Devlet Hastanesi, Gevher Hatun Mah. 1823 Sok. 1/5 Merkez, Karaman, Türkiye.
GSM: +905534803384 E-Mail: geuqsou@gmail.com

Introduction

Pulmonary embolism (PE), is a high mortality disease as blockage of pulmonary artery blood flow of the main branches or sub-branches by a thrombus. It is an urgent cardiovascular disease, and the most common cause is venous thromboembolism (VTE)[1,2]. Mortality of pulmonary embolism in patients with timely diagnosis and have appropriate prophylaxis and treatment, 2-10%, which of 2/3 the incorrect diagnosis or concomitant disease is present mortality up to 30% [1, 3]. The most commonly used laboratory methods to assist in establishing the diagnosis of pulmonary embolism is d-dimer but reliability is low. In radiological imaging, electrocardiography, echocardiography, computed tomography, ventilation-perfusion scintigraphy and pulmonary angiography can be used at specific standard. Pulmonary angiography is of limited use in emergency conditions. Clinical suspicion a variety of diagnostic laboratory and imaging results have a high importance. Anticoagulation and fybrinolytic treatments are hard to decide in some cases therefore early diagnose is important in emergency medicine.
Machine learning is a scientific discipline that deals with design and development processes of command indices to be used for transforming inputs into outputs and enables learning based on data types such as sensor data or database of the computers. It plays a role in gaining skills of perceiving complex patterns and making rational decisions based on data. Learning strategies include supervised learning, unsupervised learning, learning through problem solving, learning via neural networks, and genetic algorithms [4]. Medical use of machine learning is recently in define possibility and diagnosis.
In our study, it is planned to creation of artificial intelligence technique in diagnosing pulmonary embolism using machine learning methods. Medical history, physical examination, biochemical markers and imaging techniques of 201 patients with pulmonary complaints admitted to Mersin University Medical Faculty Hospital Emergency Department between January 2010 - October 2013 were recorded.

Material and Method

Our study was approved by the ethics committee of Gazi University. The study was designed retrospectively based on the records of the 201 patients who were presenting to Emergency Department with pulmonary complaints including dyspnea and chest pain between January 2010 and October 2013.. Based on the tests and examinations performed in the Emergency Department, the patients were grouped into two categories as “pulmonary embolism present” and “pulmonary embolism absent”. Age, sex, risk factors, d-dimer level, echocardiography findings, Doppler findings, and thoracic CT findings of the patients were recorded. The subjects were randomly assigned to education and test groups and the diagnostic accuracy (success) was assessed using MATLAB R2012b.

Results

Classification methods and success rating

Data of a total of 201 patients were used to diagnose PE. To construct a decision tree, 161 (80%) samples were assigned to education group and 40 (20%) to the test group. These samples were evaluated with success rating by using 3 separate machine learning method classification tree (CTREE), Naïve Bayes Theorem (NB), K nearest neighbor (KNN) in order to make PE diagnosis [5-8]. Accuracy (the number of correctly classified positive or negative samples), error ratio (the number of incorrectly classified positive or negative samples) as well as sensitivity (True positive ratio) and specifity ratio (False positive ratio) were calculated with the following formula :

True positive ratio = True positive / True positive + False negative

False positive ratio= False positive / True negative + False positive

Experimental Study

CTREE

In this part, presence of PE was classified using the classification tree. The algorithm formed by the decision tree operates with 95% success rate for detection of PE. The algorithm is shown on Figure 1, Figure 2.

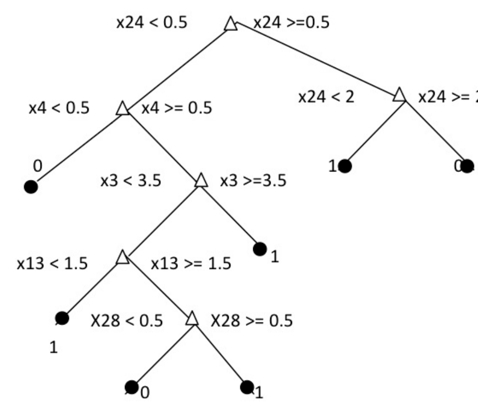

Figure 1. Algorithm of classification tree

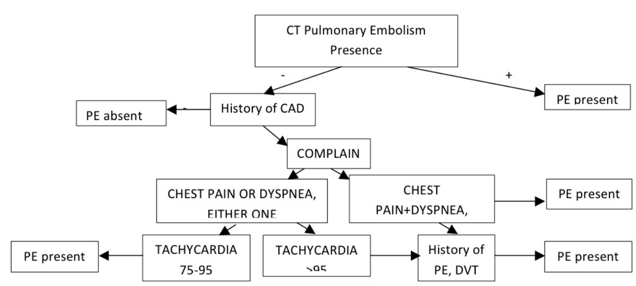

Figure 2. Algortihm of classification tree symptoms and findings with definitions (CT: Computer tomography, CAD: Coronary artery disease, DVT: Deep venous thrombosis)

In the classification study performed with the decision tree on 40 test data the complexity matrix was obtained as shown on Table 1. CTREE correctly identified 28 of 29 PE patients in the test data while it misclassified 1 of them as negative for PE. It also correctly identified 10 of 11 healthy subjects in the test data. Its sensitivity was found 96.5% and its specifity was found 90.9%.

Table 1. Complexity matrix for classification according to Classification Tree Method

		Pulmonary Embolism		
		Present	Absent	Total
CTREE	Present	28	1	40
	Absent	1	10	

KNN METHOD

The success rate of the K nearest neighbor method for detection of PE was 75%, which was lower than that of the decision tree (K=6)(Table 2). Its sensitivity was 77.1% and specifity was 60%. It was determined that the success rate was reduced to 72.5% with K=5.

Table 2. Complexity matrix for classification according to KNN Method

		Pulmonary Embolism		
		Present	Absent	Total
KNN	Present	27	2	40
	Absent	8	3	

NAİVE BAYES CLASSIFICATION METHOD

The success rate of the Naive Bayes Classification Method for detection of PE was 88.5%, which was higher than that of KNN classification. The classification was done on 201 test data. The complexity matrix of this classification was shown on Table 3. With the classification NBO, 134 cases in the test data were true positive and 11 cases were true negative. However, 12 of patients were false negative while 44 of them were false negative. The sensitivity and specifity of the method were found 91.7% and 80%, respectively.

Table 3. Complexity matrix for classification according to Naive Bayes Method

		Pulmonary Embolism		
		Present	Absent	Total
NBO	Present	134	11	201
	Absent	12	44	

Conclusion

The average of annual incidence in United States of America (USA is 1/1000 approximately. The European Union of Cardiology issued a study report of annual number of new cases of PE for Italy, Wales and France as 60000, 65,000 and 100,000 respectively [2]. In USA, annual insidance of pulmonary embolism is estimated as 600,000[9]. These results clearly show its inadequacy of PE in our country. But in recent years due to advances in the imaging and diagnosis of PE is increasing [10,11]. Rapid diagnosis and treatment of pulmonary embolism in the emergency department substantially reduce mortality and morbidity rates. Specific examination of pulmonary embolism, due to lack of laboratory and imaging findings, risk classification is determined. Purpose of determining the probability scores based on risk stratification, treatment approaches for these patients is formed.

This score is most frequently known Wells scoring, “Canadian” scores, also known as the modified Geneva score of patients with clinical probability low, medium and high risk group classification and laboratory investigations along with the evaluation[12-14].

Diagnostic aids that hasten diagnosis and increase diagnostic accuracy have been developed in the form of artificial neural networks and machine learning algorithms. Particularly, studies on imaging methods corroborated by artificial intelligence have been increased in recent years. Bouma et al found in a study that a system educated by 38 data sets created by contrast-enhanced computed tomography images had a sensitivity of 63%[8]. By forming knowledge-based artificial neural networks (KBANN), Serpen et al classified pulmonary embolism diagnosis with the modified criteria of prospective investigation of pulmonary embolism diagnosis (PIOPED) in an attempt to compare the performance of this method with those of other methods such as Bayes and decision Tree, and reported successful results [15]. Differently from PIOPED study(54.5%) , our study has a success rate of %88.5 with Naive Bayes method. Blackmon et al in another study (CAD) for pulmonary embolism diagnosis reported that computer-assisted tomography interpretation help inexperienced healthcare personnel make diagnosis and increased false positivity, albeit to a lesser extent[16]. Falsetti et al compared Wells and Geneva scores with artificial neural networks and found better results for diagnosis with artificial neural networks[17].

Similar with literature, we found same accuracy with the study of Luciani et al (88.6%)[18]. Sensitivity of 91.6% and specifity of 86.6% was similar to our study 91.7% and 80%, respectively. As in this study, no pulmonary angiography used to diagnose in the first place, Luciani et al also found 83.6%, sensitivity 88.8% and specificity 79.6% in the subgroup that pulmonary angiography was not used[18].

Pulmonary embolism has a high mortality rate and its symptoms are occasionally diagnosed in emergency services. Machine learning systems can be used especially in diseases diagnosed with clinical scores. These methods can help physicians that have no advanced diagnosing techniques as V/P sintigraphy, angiography in emergency medicine Classification tree and Bayesian method can be selected ones to diagnose or define possibility of pulmonary embolism.

Competing interests

The authors declare that they have no competing interests.

References

1. Tresoldi S, Kim YH, Baker SP, Kandarpa K. MDCT of 220 consecutive patients with suspected acute pulmonary embolism: incidence of pulmonary embolism and of other acute or non-acute thoracic findings. Radiol Med 2008;113(3):373-84.
2. Tsai AW, Cushman M, Rosamond WD. Cardiovascular risk factors and venous thromboembolism incidence: the longitudinal investigation of thromboembolism etiology. Arch Intern Med 2002;162(10):1182-9.
3. Riedel M. Venous thromboembolic disease, acute pulmonary embolism: pathophysiology, clinical presentation, and diagnosis. Heart 2001;85(2):229-40.
4. Mitchell M, editor. An introduction to genetic algorithms. London: MIT press;1998.p.66-70.
5. Anderson JR. Knowledge compilation: the general learning mechanism. In: Michalski RS, Carbonell JG, Mitchell TM, editors. Machine learning: An artificial intelligence approach. California: Morgan Kaufmann Press; 1986.p.27-42.
6. Alpaydin E, editor. Introduction to machine learning. London: MIT Press; 2004.p.168-70.
7. Pawlak Z. A rough set view on Bayes' theorem. Int J Intell Syst 2003;18:487-98.
8. Sugumaran V, Muralidharan V, Ramachandran KI. Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing. Mech Syst Sıgnal Pr 2007;21(2):930-42.
9. Torbicki A, Perrier A, Konstantinides S, Agnelli G, Galiè N, Pruszczyk P et al. ESC committee for practice guidelines (CPG). Guidelines on the diagnosis and management of acute pulmonary embolism: the task force for the Diagnosis and Management of Acute Pulmonary Embolism of the European Society ofCardiology (ESC). Eur Heart J 2008;29(18):2276-315.
10. Burge AJ, Freeman KD, Klapper PJ, Haramati LB. Increased diagnosis of pulmonary embolism without a corresponding decline in mortality during the CT era. Clin Radiol 2008;63(4):381-6.
11. Demonaco NA, Dang Q, Kapoor WN, Ragni MV. Pulmonary embolism incidence is increasing with use of spiral computed tomography. Am J Med 2008;121(7):611-7.
12. Sema U, Sevgi BS. Pulmoner tromboembolizm tanı tedavi uzlaşı raporu. Turk Toraks Derg 2009;10:20-1.
13. Ekim N. Pulmoner Tromboembolizm. In: Barıs I, editor. Akciğer Hastalıkları Cep

Kitabı. Ankara: Atlas Kitapevi 1998.p.309-28
14. Tapson VF. Acute pulmonary embolism. N Engl J Med 2008; 358:1037-52.
15. Serpen G, Tekkedil DK, Orra M. A knowledge-based artificial neural network classifier for pulmonary embolism diagnosis. Comput Biol Med 2008;38(2): 204-20.
16. Blackmon KN, Florin C, Bogoni L, McCain JW, Koonce JD, Lee H, et al. Computer-aided detection of pulmonary embolism at CT pulmonary angiography: can it improve performance of inexperienced readers. Eur Radiol 2011;21(6):1214-23.
17. Falsetti LG, Merelli E, Rucco M, Nitti C, Gentili T, Pennacchioni M, et al. A data-driven clinical prediction rule for pulmonary embolism. Eur Heart J 2013;34(Suppl.1):24-5.
18. Luciani D, Cavuto S, Antiga L, Miniati M, Monti S, Pistolesi M, et al. Bayes pulmonary embolism assisted diagnosis: a new expert system for clinical use. Emerg Med J 2007;24(3):157-64.

How to cite this article:
Berıkol G, Kula S, Yıldız O. Machine Learning Techniques in Diagnosis of Pulmonary Embolism. J Clin Anal Med 2015;6(suppl 6): 729-32.