aurum

AURUM MÜHENDİSLİK SİSTEMLERİ VE MİMARLIK DERGİSİ
AURUM JOURNAL OF ENGINEERING SYSTEMS AND ARCHITECTURE

Cilt 2, Sayı 2 | Kış 2018
Volume 2, No. 2 | Winter 2018, 1-8

REVIEW/DERLEME

APPLICATION AREA OF CLASSIFICATION TECHNIQUES IN MEDICINE

Oğuz ATA[1]
[1]Altınbaş University, School of Engineering and Natural Sciences, Software Engineering, Istanbul.
oguz.ata@altinbas.edu.tr ORCID No: 0000-0003-4511-7694

Mustafa FAYEZ[2]
[2]Altınbaş University, School of Engineering and Natural Sciences, Software Engineering, Istanbul.
bhawk606@ogr.altinbas.edu.tr ORCID No: 0000-0002-8123-8963

Received Date/Geliş Tarihi: 05/11/2018 **Accepted Date/Kabul Tarihi:** 30/11/2018

Abstract

The health care industry produces a huge amount of data that collects complex patient and medical information. Data mining is popular in various fields of research because of its applications and methodologies to extract information correctly. Data mining techniques have the capabilities to find out veiled forms or relationships among the objects in the medical data. In addition the most data mining algorithms that had used in medical industry until this time are neural network including deep learning, SVM, Bayesian and fizzy logic. The main reason of use these algorithms that because they are gave best results with high accuracy with different type of medicine datasets. Finally, data mining continues with medicine industry to help people with or solve different clinical problems.

Keywords: Data mining algorithms, Medicine, Classification.

TIPTA SINIFLANDIRMA YÖNTEMLERİNİN UYGULAMA ALANLARI

Özet

Sağlık endüstrisi çok büyük miktarda veri üretmekte ve bu veriler karmaşık hasta ve sağlık bilgileri içermektedir. Veri madenciliği, veriden bilgi çıkarma uygulaması olduğundan pek çok alanda çok popülerdir. Veri madenciliği yöntemleri gerekli bilgi ve nesneleri medikal-tıbbi veriden çıkarılmak için de kullanılmaktadır. Bugüne dek tıp alanında kullanılan veri madenciliği algoritmaları derin öğrenme, SVM, Bayes ve bulanık mantıktır. Bunların kullanılmasındaki ana neden, farklı türdeki tıp verilerine çok doğru sonuçlar verebilme yetenekleridir. Veri madenciliği tıp alanında insanlara yardım etmeye ve çeşitli klinik sorunları çözmeye devam edecektir.

Anahtar Kelimeler: Veri madenciliği, Algoritmaları, Tıp, Sınıflandırma.

1. INTRODUCTION

Machine learning (ML) podia for medical purposes are also scarab the headlines the senses are a hole to the heart we've all heard it before[1]. While popular sayings are not intended to be booked literally, recent research suggests there may be some truth to this one. Collaboration with Stanford Medical College, Google and its sister establishment, Verily Life Sciences, recently informed a deep learning model that can identify the risks of cardiovascular disease from high photos of the retinal fundus At about the similar time a group of experts from the University of California, San Diego, and the University of Guangzhou described an AI platform to examine and diagnose the common causes of severe blindness at a stage where diseases are still treatable[2]. Further, the authors demonstrated the general applicability of their automated learning system by demonstrating their potential to diagnose pneumonia in children using X-ray computer vision. Elevation can used to interpret echocardiography and to do so more accurately than those of trained experts While these developments a good illustration of the potential of artificial intelligence in imaging-based medical diagnosis is not complete unanticipated[3]. It is passable that machineries can supply great quantities of data and be trained to identify patterns that are better than persons [4]. What is surprising is the speed with which this possibility is unleashed now. Medical regulators have also opened doors to automated learning algorithms. The authors wrote that what doctors pick up from data wants to be weighed heavily alongside what they distinguish from their own medical skill. Overreliance On steering the machine may lead to self-fulfilling prophecies. For example, if doctors always withdraw upkeep in patients with certain diagnoses, such as severe prematurity or brain injury, they may learn automated learning systems that such diagnoses are always fatal.

Conversely, ML learning systems can help resolve disparities in health care delivery by compensating for known biases or by locating where additional study is required balance the underlying data. However, the challenge increases as the interest in data mining grows rapidly. In order to grip these problems without using the traditional statistical methods, soft computing has emerged to be one of the encouraging data mining solutions in this area.

2. METHODOLOGY

Solutions resulted from classification algorithm are commendable but as of now, none is diverse and flexible to be accepted generally in the medical data classification community. Categorical variables in medical data are occasionally useful to arrive at decisions and to generalize information. Categorical data classification of disease and no disease sets is handy for data mining technique and also easy to extract medical information. Data mining challenge are Heterogeneity of Healthcare Data, Voluminous Data, Change Capture, Noisy, Redundant, Inconsistent, Incomplete Data, Incorporation of Constraints, Expert Knowledge and Background Knowledge in Data Mining, Imbalanced Data and usefulness of Interesting Patterns. There are list of the most focus of future machine learning classification techniques in medicine field.

2.1 Rule based Classification

In rule based classification learned model is represented as a set of IF-THEN rules[4]. The rules are collected of two parts namely rule antecedent which is the If part and rule consequent - which is the else part. An IF-THEN rule is of the form.

2.2 Bayesian Belief Networks (BBN)

Naive Bayes classifier assumes that the attributes are independent of each other but in real world scenario, the attributes may be correlated as in the medical domain where a patient's symptoms and health state are correlated[5]. In practice dependencies can exist and hence Bayesian belief networks are used to model the dependencies between attributes using joint conditional probability distributions. A Bayesian Belief network consists of a) directed acyclic graph which shows dependencies among attributes and b) Conditional Probability Table (CPT) associated with each attribute.

2.3 Artificial Neural Network (ANN)

An artificial neural network is a computational model based on biological neural systems. It consists of interconnected Treatment elements are called nodes or neurons that work together to harvest an outcomes task[6]. It is an adaptive system that changes its configuration based on exterior or interior information that movements through the network in the learning time.

2.4 Fuzzy Logic

Fuzzy logic is a multivalued logic and is an effective tool to handle problems of uncertainty[7]. In crisp logic the fact values of the predicates take only to two values either a (1) or a (0) whereas in fuzzy logic, the truth values are multivalued and hence take values in the range (0 – 1). Fuzzy logic is being used to deal with the conception of incomplete reality where the value of truth may range between perfectly right and totally wrong.

2.5 Genetic Algorithm

Genetic algorithms (GA) are search algorithms based on normal choice and normal heredity. Provides powerful examine abilities in difficult spaces. GA is an iterative process that runs on a population. An initial population consisting of randomly generated rules is created, where each rule is signified as string of bits[8]. Every single in the populace is assigned a fitness value by means of a fitness function. Based on the theory of existence of the rightest a new population (off springs) is formed consisting of the rightest rules in the existing populace.

3. RESULTS

In the results of previous study until this time the most data mining algorithm which used in medicine with different purpose like regression, clustering and classification we saw that the deep learning neural network with logitboost, fuzzy logic, j84 and Bayesian techniques gave the better results with best accuracy.

In this figure below we shows that the used of machine learning techniques in medicine fields.

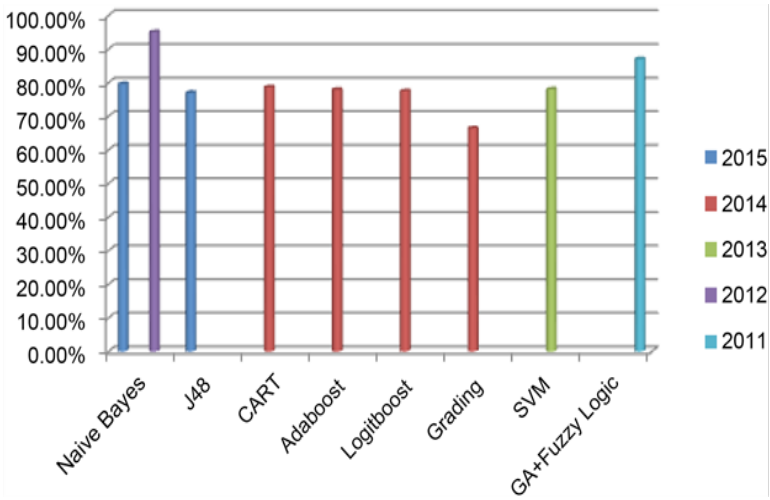

Figure 1. The distribution of techniques with years[9].

Also when lock to the previous study we saw that the most data mining algorithm used with medical sides or healthcare production datasets. We get that the support victor machine (SVM) and neural network are the most data mining algorithms used in healthcare field and the other percentages divide between other algorithms. In the figures below we show the different percentages for each algorithm used in clinical field.

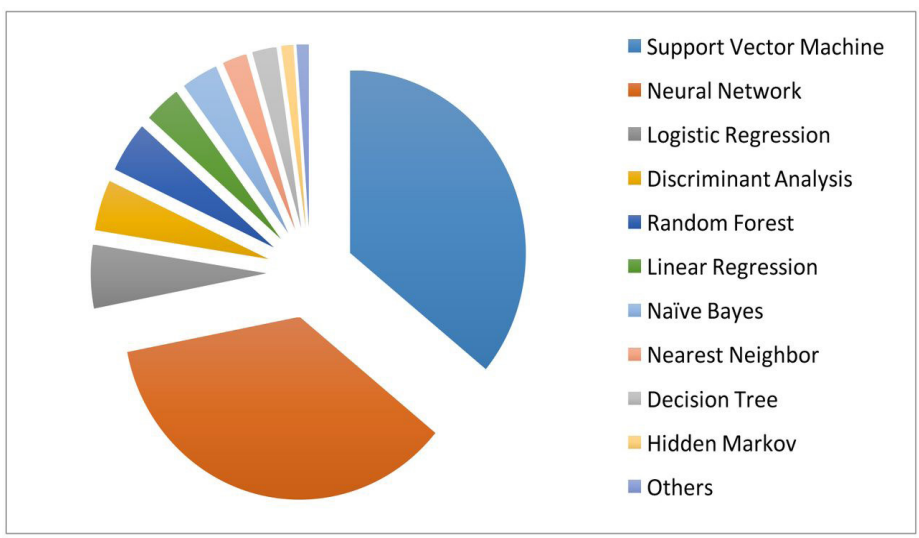

Figure 2. The percentages of algorithms in healthcare[9].

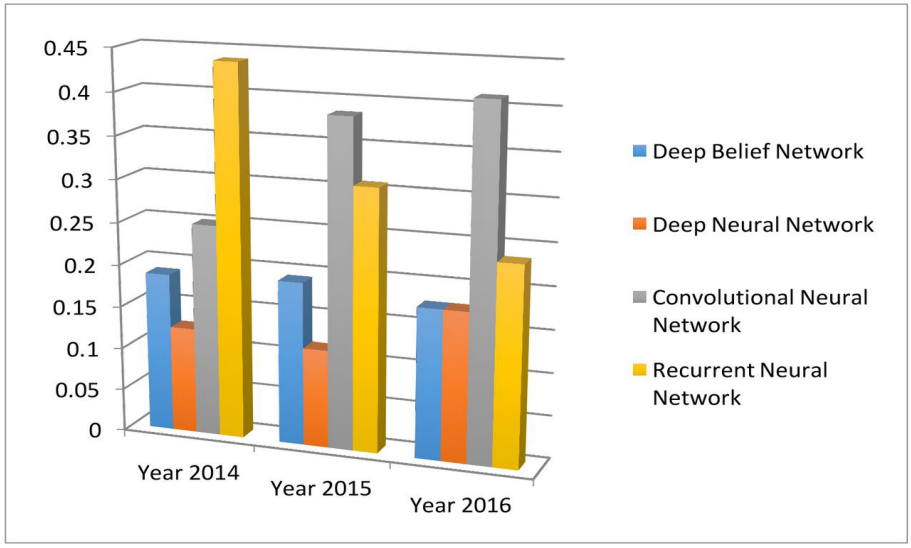

Figure 3. The percentages of most used algorithms in medical study[9].

In the figure above we showed that the most data mining algorithms used in medical study between years (2014-2016) are deep learning, deep neural network, convolutional neural network and recurrent neural network.

In addition we show the distribution of the most side in healthcare which take attract of researchers for classification or clustering and regression between years (2013-2016) in the figure below we shows that different distribution between most healthcare studies.

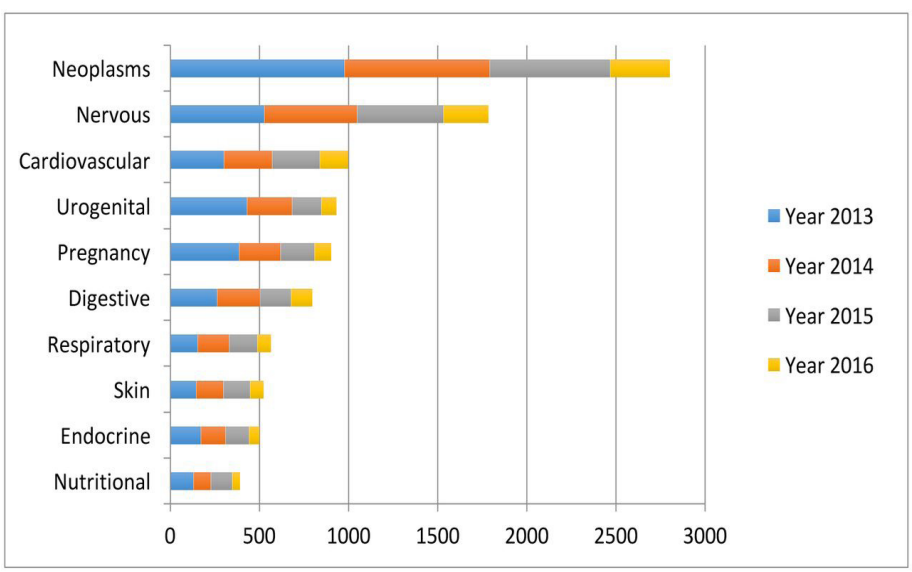

Figure 4. The distribution of most healthcare topics used[9]**.**

In the figure above we showed that the most medicine topic problem used with data mining algorithms we saw that the most is neoplasms then nervous and cardiovascular which deals with heart disease so heart disease problems take the most topic or attract of peoples scientific study around the wide world.

4. CONCLUSION

With the recent studies of medicine data that is supported by medical industry many researchers had started, or are eager to start, exploring these data. In this paper we observe some data mining techniques that have been employed for medical data and the most medicine topic used with data mining. As there are voluminous records in this industry and because of this, it has become requisite to use data mining techniques to help in decision support and prediction in the field of Healthcare to identify the kind of disease. The medical data mining produces business intelligence which is useful for diagnosing of the disease. Table 1 summarizes the medical data mining, its techniques used and for the related diseases.

Disease	Technique Used
Conventional Pathology Data	Extracting patterns & detecting trends using neural networks [10].
Coronary heart disease	Prediction models using decision tree algorithms such as ID3,C4.5, C5, and CART [11].
Lymphoma Disease and Lung Cancer	Distinguish disease subtypes using Ensemble approach [12].
Psychiatric Diseases	Predicate the probability of a psychiatric patient on the basis detected symptoms using BBN Bayesian networks [12].
Frequent Disease	Identify frequency of diseases in particular geographical area using Apriori algorithm [13].
Liver diseases	Classification using Bayesian Ying Yang (BYY) [13].
Skin Disease	Categorization of skin disease using integrated decision tree model with neural network classification methods[14].
Diabetes	Classification of Medical Data using Genetic Algorithm [15].
Chest Disease	Constructed a model using (ANN) [16].
Coronary Heart Disease	Improving classification accuracy using naive Bayesian [7].
Chronic Disease	Prediction of Diseases Using Apriori Algorithm [17].
Breast Cancer	Accurate Classification of medical data using K- means, self-Organizing map (SOM) and naive Bayes [18].
Cardio Vascular Diseases	Diagnose (CVD) using classification algorithm [19].
Parkinson Disease	Familiarized an adaptive Fuzzy K-NN approach for diagnosing the disease [20].

Table 1. Summary of medical data mining techniques.

Conflict of Interests/Çıkar Çatışması

Authors declare no conflict of interests/Yazarlar çıkar çatışması olmadığını belirtmişlerdir.

5. REFERENCES

A. Fernández Hilario, A. Altalhi, S. Alshomrani, and F. Herrera. 2017. "Why Linguistic Fuzzy Rule Based Classification Systems perform well in Big Data Applications?, *International Journal of Computational Intelligence Systems,* vol. 10(1), p. 1211-1225.

A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau*, et al.* 2017. "Dermatologist-level classification of skin cancer with deep neural networks," *Nature,* vol. 542, p. 115, 2017.

C.-L. Chang and C.-H. Chen. 2009. "Applying decision tree and neural network to increase quality of dermatologic diagnosis," *Expert Systems with Applications,* vol. 36, pp. 4035-4041.

C. S. Dangare and S. S. Apte. 2012. "Improved study of heart disease prediction system using data mining classification techniques," *International Journal of Computer Applications,* vol. 47, pp. 44-48.

D. Landuyt, A. Grêt-Regamey, and R. Haines-Young. 2017. "Bayesian belief networks," in *Mapping Ecosystem Services*, ed: Pensoft Publishers, pp. 138-143.

D. S. Kermany, M. Goldbaum, W. Cai, C. C. Valentim, H. Liang, S. L. Baxter, *et al.* 2018. "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell,* vol. 172, pp. 1122-1131. e9.

D. Shalvi and N. DeClaris. 1998. "An unsupervised neural network approach to medical data mining techniques," in *Neural Networks Proceedings, 1998. IEEE World Congress on Computational Intelligence. The 1998 IEEE International Joint Conference on*, pp. 171-176.

G. A. Longo, C. Zilio, L. Ortombina, and M. Zigliotto. 2017. "Application of Artificial Neural Network (ANN) for modeling oxide-based nanofluids dynamic viscosity," *International Communications in Heat and Mass Transfer,* vol. 83, pp. 8-14.

H. Ahuja and U. Batra. 2018. "Performance Enhancement in Requirement Prioritization by Using Least-Squares-Based Random Genetic Algorithm," in *Innovations in Computational Intelligence*, ed: Springer, pp. 251-263.

M. L. Raymer, W. F. Punch, E. D. Goodman, L. A. Kuhn, and A. K. Jain. 2000. "Dimensionality reduction using genetic algorithms," *IEEE transactions on evolutionary computation,* vol. 4, pp. 164-171.

M. Durairaj and V. Ranjani. 2013. "Data mining applications in healthcare sector: a study," *International journal of scientific & technology research,* vol. 2, pp. 29-35.

M. Koch. 2018. "Artificial intelligence is becoming natural," *Cell,* vol. 173, pp. 531-533.

O. Er, N. Yumusak, and F. Temurtas. 2010. "Chest diseases diagnosis using artificial neural networks," *Expert Systems with Applications,* vol. 37, pp. 7648-7655.

R. Das, I. Turkoglu, and A. Sengur. 2009. "Effective diagnosis of heart disease through neural networks ensembles," *Expert systems with applications,* vol. 36, pp. 7675-7680.

S. Patel and H. Patel. 2016. "Survey of data mining techniques used in healthcare domain," *International Journal of Information,* vol. 6.

S. ZahidHassan and B. Verma. 2007. "A hybrid data mining approach for knowledge extraction and classification in medical databases," in *Intelligent Systems Design and Applications, 2007. ISDA 2007. Seventh International Conference on*, pp. 503-510.

V. Venkaiah. 2017. "Study Of Data Mining Techniques Used In Medicinal Services Domain," *International Journal For Research In Advanced Computer Science And Engineering (ISSN: 2208-2107),* vol. 3, pp. 12-21.

W.-L. Zuo, Z.-Y. Wang, T. Liu, and H.-L. Chen. 2013. "Effective detection of Parkinson's disease using an adaptive fuzzy k-nearest neighbor approach," *Biomedical Signal Processing and Control,* vol. 8, pp. 364-373.

Y. J. Fei Jiang, Hui Zhi, Yi Dong, Hao Li, Sufeng Ma, Yilong Wang, Qiang Dong, Haipeng Shen, Yongjun Wang. 2017. "Artificial intelligence in healthcare: past, present and future " *Stroke Vasc Neurol.,* 2.

Y. Xing, J. Wang, and Z. Zhao. 2007. "Combination data mining methods with new medical data to predicting outcome of coronary heart disease," in *Convergence Information Technology, 2007. International Conference on*, pp. 868-872.