Türkiye Klinikleri Biyoistatistik Dergisi
Turkiye Klinikleri Journal of Biostatistics
Turkiye Klinikleri J Biostat. 2023;15(1):13-8

ORİJİNAL ARAŞTIRMA ORIGINAL RESEARCH

DOI: 10.5336/biostatic.2022-93499

A Simulation Study to Compare Classification Performance of Fuzzy Inference System with Machine Learning: Methodological Research

Bulanık Çıkarsama Sistemi ile Makine Öğrenmesi Yöntemlerinin Sınıflandırma Performansını Karşılaştıran Bir Simülasyon Çalışması: Metodolojik Araştırma

İrem KAR[a], Batuhan BAKIRARAR[a], Derya GÖKMEN[a], Serdal Kenan KÖSE[a]

[a]Department of Biostatistics, Ankara University Faculty of Medicine, Ankara, Türkiye

ABSTRACT Objective: Classification is one of the most important research topics of machine learning that aims to correctly predict the target class for each case in the data. In this study, classification performances of fuzzy inference systems that learn from experts and machine learning methods that learn from the data were compared in different sample sizes. **Material and Methods:** This study was planned as a methodological research. The machine learning algorithms used in the comparison are Multilayer Perceptron, Random Forest, Support Vector Machine, which are frequently encountered classifiers in the literature. The dataset were generated for 6 (sex, chest pain type, max heart rate, exercise induced, oldpeak and major vessels) independent variables determined by variable importance, preserving the characteristics of heart disease data in the University of California Irvine database. Sample sizes were determined in four different sizes as 100, 250, 500 and 1,000. The datasets were divided into two separate sets randomly, with 70% training set and 30% test set. In fuzzy inference systems, the fuzzy rules were automatically generated and Chi's technique was used to create the rules. Accuracy, precision, sensitivity and F-measure were used as performance metrics for classification to compare four methods. **Results:** As a result of the study, it had been observed that fuzzy inference systems are affected by the sample size, and the classification performance is better than other methods as the sample size increases. **Conclusion:** In general, it has been observed that as the sample size increased, the classification performance of the methods increased.

Keywords: Fuzzy inference system; classification; machine learning

ÖZET Amaç: Sınıflandırma, veri setinde her gözlem için hedef sınıfı doğru tahmin etmeyi amaçlayan makine öğrenmesinin en önemli araştırma konularından biridir. Bu çalışmada, uzmandan öğrenen bulanık çıkarsama sistemleri ile veriden öğrenen makine öğrenmesi yöntemlerinin farklı örnek büyüklüklerinde sınıflandırma performansları karşılaştırılmıştır. **Gereç ve Yöntemler:** Bu çalışma metodolojik bir araştırma olarak planlanmıştır. Karşılaştırmada kullanılan makine öğrenmesi algoritmaları literatürde sıklıkla karşılaşılan sınıflandırıcılar olan; Çok Katmanlı Algılayıcı, Rastgele Orman ve Destek Vektör Makinesi yöntemleridir. Bu çalışmada, University of California Irvine (UCI) makine öğrenmesi veritabanında yer alan kalp hastalığı verisinin özellikleri korunarak değişken önemine göre belirlenen 6 bağımsız değişken (cinsiyet, göğüs ağrı tipi, maksimum kalp atım hızı, egzersizin neden olduğu anjin, oldpeak ve ana damar sayısı) için türetilen veriler kullanılmıştır. Örneklem büyüklükleri 100, 250, 500 ve 1000 olmak üzere dört farklı büyüklükte belirlenmiştir. Veri seti %70 eğitim ve %30 test seti olmak üzere ikiye ayrılmıştır. Bulanık çıkarsama sistemlerinde kurallar otomatik olarak oluşturulmuş olup, kuralların oluşturulmasında Chi'nin tekniğinden yararlanılmıştır. Dört yöntemin karşılaştırılmasında, sınıflandırma performans ölçütleri olarak doğru sınıflama oranı, kesinlik, duyarlılık ve F-ölçütü kullanılmıştır. **Bulgular:** Çalışma sonucunda, bulanık çıkarsama sistemlerinin örneklem sayısına daha duyarlı olduğu, örneklem büyüklüğü arttıkça sınıflama performansının da diğer yöntemlere göre iyi olduğu gözlenmiştir. **Sonuç:** Genel olarak ise örneklem büyüklüğü arttıkça, yöntemlerin sınıflama performanslarının arttığı gözlemlenmiştir.

Anahtar kelimeler: Bulanık çıkarsama sistemleri; sınıflama; makine öğrenmesi

Correspondence: İrem KAR
Department of Biostatistics, Ankara University Faculty of Medicine, Ankara, Türkiye
E-mail: ikar@ankara.edu.tr

Peer review under responsibility of Turkiye Klinikleri Journal of Biostatistics.

Received: 26 Sep 2022 ***Received in revised form:*** 22 Nov 2022 ***Accepted:*** 27 Nov 2022 ***Available online:*** 12 Dec 2022

In computer systems, learning can be achieved in 2 different ways. One is learning from the data which is accepted as based on machine learning science and the other is learning based on the expert opinion. When data is available, it is possible to make the required inferences with machine learning methods, otherwise methods that benefit from expert opinion should be used.[1]

Fuzzy logic provides a broader, multidimensional framework for classical logic approach which can be inadequate to solve uncertainty problems. Accordingly, sharp lines in classical logic turn into clusters whose boundaries are not crisp. Fuzzy inference systems (FIS) can be defined as systems that are used for modeling the fuzzy logic approach and that contain rules and associations with all of the inputs.[2,3] They need expert opinion and not an emphasis on learning process, and they are considered as an intuitive solution with these features.[1-3]

In the present study, classification performances of FIS that learn from experts and machine learning methods that learn from the data were compared in different sample sizes using the heart disease dataset. Machine learning algorithms used in the comparison are multilayer perceptron (MLP), support vector machine (SVM) and random forest (RF), which are applied in the literature frequently. Recently, these methods are used in a variety of heart disease applications.

BACKGROUND

FIS

FIS can be defined as systems that contain rules and sets, which are used to model the fuzzy logic approach, which relate all of the inputs to the output. Fuzzy systems are successfully implemented in many areas such as expert systems, decision support systems, automatic control systems, image recognition and robotics.[1,2]

IF-THEN rule structure is one of the most important term in fuzzy systems. The system based on fuzzy rules is quite useful in modeling some complex systems based on human perception. Ultimately, it can be reduced to a simple set of rules as shown in Table 1.[4,5]

TABLE 1: The canonical form for fuzzy rule-based system.

Rule 1:	IF condition C^1, THEN restriction R^1
Rule 2:	IF condition C^2, THEN restriction R^2
.	
.	
.	
Rule r:	IF condition C^r, THEN restriction R^r

The fuzzy rule-based system is most convenient in the process of modeling some complex systems. Based on the input variable, a restriction is made on the output variable within certain conditions. Restrictions are usually expressed with "and, or, else" and are applied to the results of the rules.[4,5]

A FIS is basically composed of 4 steps: fuzzification, rule fitting, aggregation and defuzzification, respectively. As a result of the classification of the consequent parts of the rules formed in the FIS based on linguistic expressions, 3 basic inference methods which are widely used are constructed: Mamdani, Takagi-Sugeno and Tsukamoto.[4-9]

MLP

MLP is the most used supervised learning model among artificial neural network models which is developed by Rumelhart et al.[10] MLP model has an input layer, a single or more hidden layer(s) and an output layer.

These networks are instructed both inputs and outputs which should be produced in return for those inputs during training.[11]

SVM

SVM are used in both classification and regression analysis in the literature, but it is generally used for classification. This method is based on a supervised learning model. It can be used in core functions depending on the type of data in the process of the algorithm. In this way, both linear and nonlinear classifications can be performed.[11,12]

RF

RF is a method that consists of more than one decision tree and performs depending on the decision trees. In this method, each tree with a similar distribution processes the data independently. In the classification process, the most popular class is determined by voting from each tree.[11,13]

MATERIAL AND METHODS

DATASET

The dataset used in our study is the Heart disease dataset from Statlog (270 patterns) from the University of California Irvine (UCI) database. The dataset consists of a total of seven variables, 6 independent variables and a dependent variable which refers to the presence or absence of heart disease in the patient. Detailed information on these variables is presented in Table 2.

In this study, which was planned as a methodological study, data were generated for 6 (sex, chest pain type, max heart rate, exercise induced, oldpeak and major vessels) independent variables determined by both clinical importance and variable importance, preserving the characteristics of heart disease data in the UCI database.[14] While 44.4% of the patients in the study had heart disease, 55.6% did not, and this ratio was preserved when simulated the data for all sample sizes.

TABLE 2: Input variables in dataset.

Input variables	Description
Sex	Two categories: male, female
Chest pain type	Four categories: typical angina, atypical angina, non-anginal pain, asymptomatic
Max heart rate	Maximum heart rate achieved
Exercise induced	Two categories: yes, no
Oldpeak	ST depression induced by exercise relative to rest
Major vessels	Number of major vessels (0-3) colored by flourosopy

PERFORMANCE MEASURES

In order to compare the classification performances of the methods, accuracy, precision, recall and F-measure were calculated as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$F - \text{measure} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

where TP = true positives, TN = true negatives, FP = false positives and FN = false negatives.

EVALUATION PROCESS

The simulated data were generated in different sample sizes such that n=100, 250, 500, 1,000. The datasets were divided into 2 separate sets randomly, with 70% training set and 30% test set. Accuracy, precision, sensitivity and F-measure were used as performance metrics for classification to compare four methods. Each scenario for machine learning methods were repeated 1,000 times.

Hyperparameters for all methods were as follows. In MLP, number of units in the hidden layers was 5, maximum of iterations to learn was 50, learning rate was 0.3 and momentum was 0.7. In SVM, type was C-classification, kernel was radial and gamma was 0.1 whi. In RF, the number of trees was 100 and the number of randomly selected variables as candidates to split a node was determined as p/3 for regression, where p equals the number of variables. In FIS, the fuzzy rules were automatically generated and Chi's technique was used to create the rules.[15] The rules were created separately for the 4 different sample size settings previously mentioned. The number of rules established for 100, 250, 500 and 1,000 sample sizes were 47, 67, 97 and 108, respectively. In this study, Mamdani inference method was used as the fuzzy inference types and weighted average method was used as the defuzzification method in the analysis of the FIS.

The fuzzy rules were automatically generated in the open source program called as KEEL v3.0 (Knowledge Extraction based on Evalutionary Learning).[16,17] R 3.3.2 programming language was utilized to analyze the data.[18]

R packages used in the analysis of MLP, RF, SVM and FIS algorithms were "RSNNS", "party", "randomForest", "e1071" and "frbs", respectively.[19-25]

RESULTS

The simulation results for n=100, 250, 500, 1,000 were given in Table 3.

TABLE 3: The performance metrics of classifiers with changing of sample size.

Sample size (n)		Accuracy	Precision	Recall	F-measure
n=100	MLP	0.767	0.846	0.687	0.758
	RF	0.746	0.832	0.711	0.762
	SVM	0.780	0.761	0.807	0.776
	FIS	0.760	0.871	0.760	0.772
n=250	MLP	0.816	0.817	0.809	0.812
	RF	0.794	0.750	0.776	0.756
	SVM	0.793	0.757	0.743	0.747
	FIS	0.746	0.745	0.746	0.745
n=500	MLP	0.818	0.815	0.817	0.815
	RF	0.792	0.763	0.755	0.757
	SVM	0.804	0.787	0.821	0.801
	FIS	0.781	0.780	0.781	0.780
n=1,000	MLP	0.873	0.888	0.879	0.873
	RF	0.881	0.883	0.873	0.856
	SVM	0.888	0.902	0.897	0.894
	FIS	0.899	0.904	0.899	0.900

MLP: Multilayer perceptron; RF: Random forest; SVM: Support vector machines; FIS: Fuzzy information systems.

Instead of interpreting the precision and recall criteria separately, it is preferable to make an interpretation according to the F-measure, which is the harmonic mean of the precision and recall. In addition to the F-measure, the most commonly used classification performance metric is accuracy. These 2 criteria were used in the study as an evaluation metrics in interpreting the results.

When sample size was 100, SVM had the highest accuracy value with 0.780 and was followed by MLP, FIS and RF algorithms. According to F-measure, SVM obtained the best result with 0.776, the following algorithms are FIS, RF and MLP respectively. When sample size was 250, MLP performed better than the other algorithms used in comparison with the values of 0.816 accuracy and 0.812 F-measure. When sample size was 500, MLP performed better than the other algorithms used in comparison with the values of 0.818 accuracy and 0.815 F-measure. The MLP algorithm was followed by SVM, RF and FIS, respectively. Finally, when the sample size was 1,000, all algorithms performed fairly well. However, FIS algorithm performed better than other algorithms with 0.889 of accuracy and 0.900 of F-measure.

DISCUSSION

There are many examples of using machine learning methods in the literature related to healthcare field. However, the number of studies using fuzzy logic is much less. Adeli and Neshat designed a system based on fuzzy rules for heart disease diagnosis.[26] They used V.A. Medical Center database which has 11 input variables and one output variable. The output variable refers to the presence or absence of heart disease. The fuzzy system predicted 94% of patients correctly.

Barman and Choudhury presented a fuzzy rule based system for the diagnosis of the heart disease.[27] The proposed system has seven input variables and one output variable. They created three membership functions for prediction process of heart disease and selected one of them which has the minimum value of absolute residual.

Bhatla and Jyoti designed a system to predict heart disease. In their dataset, there were many variables related to the heart disease, so they aimed to generate a system which used to reduce number of attributes automatically and then predict heart disease. The aim of the created system was to obtain the best predictive model with less variables. In the study, the prediction rate was 99.6% with the 15-variable model, 96.6% with the 13-variable model, 99.2% with the 6-variable model and 100% with the 4-variable (recommended) model.[28]

Kumari and Sunita used Artifical Neural Networks, Fuzzy Logic and Neuro-Fuzzy Integrated Approach to diagnose heart disease and compared their results. The used ten input variables and one output variable. They achieved the best result with neuro-fuzzy integrated approach.[29]

In our study, datasets consisting of 6 input and one output variables with 100, 250, 500 and 1,000 sample sizes were used. Classification performances were tested on these datasets by using MLP, RF, SVM and FIS methods. The best performance was achieved with the FIS method in a 1,000-sample size dataset. The accuracy of the model was found to be 89.9% and the F-measure as 90.0%.

CONCLUSION

Machine learning methods and FIS are widely used methods in healthcare field. In this study, classification performances of the most frequently used machine learning methods and FIS were compared in different sample sizes.

In datasets with 100, 250 and 500 sample sizes according to the performance criteria of this study, accuracy and F-measure, SVM and MLP were better than the other 2 algorithms. The FIS algorithm yielded better results in large datasets. In this study, it was seen that FIS is the best method of giving results in a dataset with 1,000 sample sizes. FIS was followed by SVM, MLP, RF respectively.

There are many methods and algorithms used in machine learning. The results of these algorithms vary depending on the dataset, data preprocessing, and selection of parameters. In this study, FIS showed the best result in the large sample size but the number of independent variables and the sample size can be increased in future studies in order to obtain better classification performance. Furthermore, the selected algorithm parameters can be changed and compared. In addition, it can be aimed to increase the classification success of the model by creating more rules in the analyses using FIS.

Source of Finance

During this study, no financial or spiritual support was received neither from any pharmaceutical company that has a direct connection with the research subject, nor from a company that provides or produces medical instruments and materials which may negatively affect the evaluation process of this study.

Conflict of Interest

No conflicts of interest between the authors and/or family members of the scientific and medical committee members or members of the potential conflicts of interest, counseling, expertise, working conditions, share holding and similar situations in any firm.

Authorship Contributions

Idea/Concept: *İrem Kar;* ***Design:*** *İrem Kar;* ***Control/Supervision:*** *İrem Kar, Serdal Kenan Köse;* ***Data Collection and/or Processing:*** *İrem Kar, Batuhan Bakırarar;* ***Analysis and/or Interpretation:*** *İrem Kar;* ***Literature Review:*** *İrem Kar, Derya Gökmen;* ***Writing the Article:*** *İrem Kar;* ***Critical Review:*** *Derya Gökmen, Serdal Kenan Köse.*

REFERENCES

1. Mutlu B, Sezer EA, Nefeslioglu HA. A defuzzification-free hierarchical fuzzy system (DF-HFS): Rock mass rating prediction. Fuzzy Sets Syst. 2017;307:50-66. [Crossref]
2. Zadeh LA. Fuzzy sets. Information and Control. 1965;8(3):338-53. [Crossref]
3. Zadeh LA. Outline of a new approach to the analysis of complex systems and decision processes. IEEE Trans Syst Man Cybern. 1973;3(1):28-44. [Crossref]
4. Ross TJ. Fuzzy Logic with Engineering Applications. 3rd ed. Chichester, UK: John Wiley & Sons; 2010. [Crossref]
5. Jang, JSR, Sun CT, Mizutani E. Neuro-Fuzzy and Soft Computing-A Computational Approach to Learning and Machine Intelligence. 1st ed. Upper Saddle River, NJ: Prentice Hall; 1996.
6. Mamdani EH, Assilian S. An experiment in linguistic synthesis with a fuzzy logic controller. Int J Man Mach Stud. 1975;7(1):1-13. [Crossref]
7. Takagi T, Sugeno M. Fuzzy identification of systems and its applications to modeling and control. IEEE Trans Syst Man Cybern. 1985;(1):116-32. [Crossref]
8. Sugeno M, Kang GT. Structure identification of fuzzy model. Fuzzy Sets Syst. 1988;28(1):15-33. [Crossref]
9. Tsukamoto Y. An approach to fuzzy reasoning method. In: Gupta MM, Ragade RK, Yager RR, eds. Advances in Fuzzy Set Theory and Applications. 1st ed. Amsterdam: Elsevier North-Holland; 1979. p.137-49.
10. Rumelhart DE, Durbin R, Golden R, Chauvin Y. Backpropagation: The basic theory. Backpropagation: Theory, Architectures and Applications. 1st ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1995. p.1-34.
11. Han J, Pei J, Kamber M. Data Mining: Concepts and Techniques. 3rd ed. Waltham, MA: Elsevier; 2011.
12. Cortes C, Vapnik V. Support-vector networks. Mach Learn. 1995;20(3):273-97. [Crossref]
13. Breiman L. Random forests. Mach Learn. 2001;45(1):5-32. [Crossref]
14. Dua D, Karra Taniskidou E. UCI Machine Learning Repository. Irvine, CA: University of California, School of Information and Computer Science; 2017. Access date: [10 March 2022]. Access link: [Link]
15. Chi Z, Yan H, Pham T. Fuzzy Algorithms: with Applications to Image Processing and Pattern Recognition. Vol. 10. 1st ed. Singapore: World Scientific; 1996. [Crossref] [PubMed]
16. Alcalá-Fdez J, Sánchez L, Garcia S, del Jesus MJ, Ventura S, Garrell JM, et al. KEEL: a software tool to assess evolutionary algorithms for data mining problems. Soft Comput. 2009;13(3):307-18. [Crossref]
17. Alcalá-Fdez J, Fernández A, Luengo J, Derrac J, García S, Sánchez L, et al. Keel data-mining software tool: data set repository, integration of algorithms and experimental analysis framework. Journal of Multiple-Valued Logic & Soft Computing. 2011;17:255-87. [Link]
18. Team RC. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2016. [Link]
19. Bergmeir CN, Benítez Sánchez JM. Neural networks in R using the Stuttgart neural network simulator: RSNNS. American Statistical Association. Journal of Statistical Software. 2012;46(7). [Crossref]
20. Hothorn T, Bühlmann P, Dudoit S, Molinaro A, Van Der Laan MJ. Survival ensembles. Biostatistics. 2005;7(3):355-73. [Crossref] [PubMed]
21. Liaw A, Wiener M. Classification and regression by random Forest. R News. 2002;2(3):18-22. [Link]
22. Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F, Chang CC, et al. e1071: Misc Functions of the Department of Statistics, Probability Theory Group. Vienna: R Foundation for Statistical Computing; 2017. [Link]
23. Strobl C, Boulesteix AL, Zeileis A, Hothorn T. Bias in random forest variable importance measures: Illustrations, sources and a solution. BMC Bioinformatics. 2007;8(1):25. [Crossref] [PubMed] [PMC]
24. Strobl C, Boulesteix AL, Kneib T, Augustin T, Zeileis A. Conditional variable importance for random forests. BMC Bioinformatics. 2008;9(1):307. [Crossref] [PubMed] [PMC]
25. Riza LS, Bergmeir CN, Herrera F, Benítez Sánchez JM. frbs: Fuzzy rule-based systems for classification and regression in R. American Statistical Association. Journal of Statisticak Software. 2015;65(6). [Crossref]
26. Adeli A, Neshat M. A fuzzy expert system for heart disease diagnosis. Proceedings of International Multi Conference of Engineers and Computer Scientists. Hong Kong: IMECS 2010, 17-19 March, 2010. [Link]
27. Barman M, Choudhury JP. A fuzzy rule base system for the diagnosis of heart disease. Int J Comput Appl. 2012;57(7). [Link]
28. Bhatla N, Jyoti K. A novel approach for heart disease diagnosis using data mining and fuzzy logic. Int J Comput Appl. 2012;54(17):16-21. [Crossref]
29. Kumari N, Sunita S. Comparison of ANNs, fuzzy logic and neuro-fuzzy integrated approach for diagnosis of coronary heart disease: a sur-vey. International Journal of Computer Science and Mobile Computing, IJCSMC. 2013;2(6):216-24. [Link]