THE JOURNAL OF COGNITIVE SYSTEMS
Journal Homepage: https://dergipark.org.tr/en/pub/jcs
Vol.6, No.2, 2021

Heart Disease Classification Based on Performance Measures Using a Deep Learning Model

[1]Ipek Balikci Cicek, [2]Zeynep Kucukakcali

[1,2]Inonu University Department of Biostatistics and Medical Informatics, Faculty of Medicine, Malatya, Turkey. (e-mail: {ipek.balikci, zeynep.tunc}@inonu.edu.tr).

ARTICLE INFO

Received: Sep.,23.2021
Revised: Nov,27.2021
Accepted: Nov.,30.2021

Keywords:
Artificial intelligence
Heart disease
Machine learning
Deep learning
Classifiction

Corresponding author: Zeyep Küçükakçalı
✉ zeynep.tunc@inonu.edu.tr
☎ +90 422 3410660/1337

ISSN: 2548-0650

DOI: https://doi.org/10.52876/jcs.1015210

ABSTRACT

Heart disease, which is one of the most common diseases in the world, is expected to remain the leading cause of mortality on a global scale. Therefore the aim of this study is to classify heart disease using a deep learning approach in an open-access dataset that includes data from patients with and without heart disease.

In this study, a deep learning model was applied to an open-access data set containing the data of patients with and without heart disease. The performance of the method used was evaluated with the performance criteria of specificity, sensitivity, accuracy, positive predictive value, and negative predictive value. Specificity, sensitivity, accuracy, positive predictive value and negative predictive value from the performance criteria obtained from the model were calculated as 0.946, 0.903, 0.9245, 0.9436 and 0.907, respectively.

As a result of the findings obtained from the study, it was seen that the data set we discussed was successfully classified with the deep learning model used. With this obtained high classification performance, the factors associated with the disease can be revealed.

1. INTRODUCTION

HEART disease refers to a set of conditions involving the heart, its vessels, muscles, valves, or internal electrical pathways responsible for muscle contraction. With a more comprehensive definition, heart diseases are complex clinical syndromes that can affect the endocrine, hematological, musculoskeletal, renal, respiratory, peripheral vascular, hepatic, and gastrointestinal systems. According to the Centers for Disease Control, heart disease is one of the leading causes of death worldwide. In addition, despite the advances in modern medicine, it continues to affect millions of people around the world and has high mortality rates [1, 2]. Usually one in four deaths occur as a result of heart disease. Heart disease is common among both men and women in most countries around the world. Therefore, people should consider heart disease risk factors. Although genetics may play a role, certain lifestyle factors significantly influence heart disease. It is predicted that heart disease will continue to be the leading cause of death for a long time globally. For this reason, it is of great importance to estimate the factors associated with the disease by using data mining algorithms with the data of patients diagnosed with heart disease, to take precautions related to the disease, and for physicians to benefit from it [1, 3].

Artificial intelligence is a machine learning that can act like a human, imitate human behavior, make rational decisions, and respond in a meaningful way. While doing these, the basic requirement is education. For years, many scientists thought that computer intelligence was useless without learning.However, this idea has changed and developed rapidly with the advent of machine learning and later deep learning [4]. Machine learning; it is one of the sub-branches of artificial intelligence and has become one of the most important areas to obtain useful, meaningful information from large data sets. Machine learning areas are increasing with the development of today's technologies. This is because the size and complexity of data is increasing day by day. Analyzing large and complex data have becomes even more difficult. In such a case, the use of machine learning has become a necessity for analysis [4, 5].

Deep learning, it contains multiple artificial neural networks and is a sub-branch of machine learning used to obtain new data by analyzing the properties of data such as existing images, audio, with many algorithms including machine learning algorithms [6].

The aim of this study is to classify heart disease using a deep learning approach in an open-access dataset that includes data from patients with and without heart disease. According to the results of this classification model, the factors causing heart

disease can be determined and medical professionals will be able to benefit from these results and can be used in preventive medicine.

2. MATERIAL and METHODS

In this study, a data set containing information about patients with and without heart disease obtained from the address "https://www.kaggle.com/asaumya/healthcare-problem-prediction-stroke-patients" was used. The data set used in the study is unbalanced. For this reason, SMOTE, an oversampling method, was used to balance the data set.

It is very difficult to analyze data stacks containing very large and different numerical data using classical algorithms and techniques [7]. In order to solve this problem, data science has made a rapid development in recent years. The importance of data science comes to the fore one more step, especially when considering the ever-growing digital data. As a result of such rapid development of data science, the concepts of data mining, artificial intelligence, machine learning and deep learning have emerged. In order to obtain meaningful patterns from large amounts of data, different algorithms are constantly being developed with data mining. In this sense, the concepts of artificial intelligence, machine learning and deep learning continue to develop in parallel with data mining [8].

Artificial intelligence in literature, transferring the working structure of human intelligence to computers in general; it is defined as the ability of computers to perform tasks that require logic, such as drawing conclusions from human-specific behaviors, finding solutions, making generalizations, understanding the problem, and learning by making use of past experiences [9].

Deep learning is a newer type of artificial neural network algorithm compared to others and is one of the sub-branches of machine learning. Deep learning is an algorithm that has one or more inputs, containing many layers, and one or more outputs at the end. In each layer, it combines the previous information and generates values with complex and meaningful results from this information. In this respect, it is more consistent and powerful than other neural network algorithms. Deep learning models consist of different data transformation stages by considering the properties of the data in the sources and learning them in their hidden layers [10].

The deep learning method learns the distinguishing features itself from a large number of given inputs. This feature learning stages consists of a number of layers. The lower-level layers have less distinctive features, while the higher-level layers, which are a combination of these layers, have more distinctive features. The lower-level layers form the basis of the higher-levels and enable more meaningful features to be produced. Unlike traditional machine learning, it does the learning process on its own instead of calculating the basic features determined by the human [11].

2.1. Data Analysis

Quantitative data were expressed as median (minimum-maximum), and qualitative data as number (percentage). Conformity to normal distribution was evaluated using the Kolmogorov-Smirnov test. Whether there is a statistically significant difference between the "No heart disease" and "Suffering from heart disease" groups, which are the categories of dependent / target variable (heart disease) in terms of independent variables, was examined using the Mann-Whitney U test and Pearson chi-square test. Values of p <0.05 were considered statistically significant. IBM SPSS Statistics 26.0 package program was used for all analyzes.

3. RESULTS

The table showing the distribution of the dependent variable in the data set used in this study is given below.

TABLE I
TABLE SHOWING THE DISTRIBUTION OF THE DEPENDENT VARIABLE

No heart disease		Suffering from heart disease	
Count	Percentage (%)	Count	Percentage (%)
1409	83.6	276	16.4

Descriptive statistics of the independent variables in this study are given in Table 2. According to this table; there is a statistically significant difference between the groups of the dependent variable (Heart Disease) in terms of age, avg glucose level and BMI variables (p<0.05).

TABLE II
DESCRIPTIVE STATISTICS TABLE OF QUANTITATIVE INDEPENDENT VARIABLES

Variables	Heart Disease		p-value*
	No heart disease	Suffering from heart disease	
	Median (min-max)	Median (min-max)	
Age	47 (0.08-82)	71 (2-82)	<0.001
Avg Glucose Level	93.05 (55.22-267.76)	106.55 (56.31-271.74)	<0.001
BMI	28.3 (10.3-78)	29.8 (19.1-54.7)	0.001

***: Mann Whitney U test**

Table 3 shows that; there is a statistically significant relationship between the gender, hypertension, ever married, work type and smoking status variables and the dependent variable (Heart Disease) groups (p<0.05).

TABLE III
DESCRIPTIVE STATISTICS TTABLE OF QUANTITATIVE INDEPENDENT VARIABLES

Variables	Categories of Variables	Heart Disease		p-value*
		No heart disease	Suffering from heart disease	
		Number (%)	Number (%)	
Gender	Male	570 (40.5)	163 (59.1)	<0.001
	Female	838 (59.5)	113 (40.9)	
Hypertension	No Hypertension	1258 (89.3)	212 (76.8)	<0.001
	Suffering from Hypertension	151 (10.7)	64 (23.2)	
Ever Married	No	457 (32.4)	32 (11.6)	<0.001
	Yes	952 (67.6)	244 (88.4)	
Work Type	Private	848 (60.2)	158 (57.2)	<0.001
	Self Employed	216 (15.3)	81 (29.3)	
	Govt Job	169 (12.0)	36 (13.0)	
	Children	170 (12.1)	1 (0.4)	
	Never Worked	6 (0.4)	0 (0.0)	
Residence Type	Urban	709 (50.3)	142 (51.4)	0.731
	Rural	700 (49.7)	134 (48.6)	
Smoking Status	Never Smoked	553 (39.2)	90 (32.6)	<0.001
	Formerly Smoked	249 (17.7)	77 (27.9)	
	Unknown	395 (28.0)	48 (17.4)	
	Smokes	212 (15.0)	61 (22.1)	
Stroke	No Stroke	1207 (85.7)	229 (83.0)	0.249
	Suffered Stroke	202 (14.3)	47 (17.0)	

***: Pearson chi-square test**

Table 4 shows the classification matrix for the associative classification model that was used to classify the Heart Disease Dataset in this study.

TABLE IV
DESCRIPTIVE STATISTICS TABLE OF QUANTITATIVE INDEPENDENT VARIABLES

Prediction	Reference		
	No heart disease	Suffering from heart disease	Total
No heart disease	4573	469	5042
Suffering from heart disease	261	4365	4626
Total	4834	4834	9668

Table 5 shows the results of the classification performance criterion for the associative classification model. The model's specificity was calculated to be 0.946, the sensitivity to be 0. 903, the accuracy to be 0.9245, the positive predictive value to be 0.9436 and the negative predictive value to be 0.907.

TABLE V
THE MODEL'S CLASSIFICATION PERFORMANCE CRITERIA'S VALUES

Metric	Value
Specificity	0.946
Sensitivity	0.903
Accuracy	0.9245
Positive predictive value	0.9436
Negative predictive value	0.907

4. DISCUSSION

It is estimated that heart disease, which is one of the most common diseases in the world, will continue to be the most common cause of death on a global scale. It is reported that these deaths tend to decrease in developed countries compared to developing countries. The most effective factor in the reduction of these deaths in the world is the preventability of heart and cardiovascular diseases. Studies such as identifying the factors causing the disease and preventing their emergence have been made possible by developing health services. Recent research has been able to identify risk factors for heart disease, but many researchers agree that more research is needed to use this information to reduce the incidence of heart disease. Heart diseases may be due to different characteristics. Some literature studies have shown that reducing these risk factors for heart disease may actually help prevent heart disease. There are many studies and studies on the prevention of heart disease risk. More studies on heart disease will offer more opportunities to prevent heart disease [1, 12, 13].

With the use of new application areas obtained as a result of developments in computer science in health services, it has been possible to predict disease-related risk factors from patient data. Thus, by determining the factors that cause diseases, important steps are taken in terms of preventability.

Computer and human is one of the interdisciplinary fields of study that deals with the design, implementation and evaluation of interactive technologies. The field of human and computer interaction started to gain momentum in the 2000s and studies in this field are among the areas that have priority today. One of the human and computer interaction fields of study is deep learning, which is a method of machine learning [14]. Deep learning methods are a set of algorithms involved in machine learning; and attempts to model high-level abstractions of data using model architectures that result from multiple nonlinear transformations [15].

Deep learning methods, with the increase in processing power and the advancement of graphics processors, many more applications such as big data analysis, image classification, voice identification, generic visual recognition, face recognition, pedestrian detection, natural language processing, handwriting recognition, multiclassification, regression problems, time series estimation, etc. started to take place in a wide variety of fields [15, 16].

In this study, a deep learning model was applied using an open-source dataset containing information about patients with and without heart disease, and it was aimed to classify the output variable and to obtain information about the factors associated with the disease. Specificity, sensitivity, accuracy, positive predictive value, and negative predictive value from the performance criteria obtained from the model as a result of classification were calculated as 0.946, 0.903, 0.9245, 0.9436, and 0.907, respectively.

As a result of the findings obtained from the study, it was seen that the data set we discussed was successfully classified with the deep learning model used. This successful classification performance will give medical professionals an idea about the risk factors that may be associated with heart disease and will be decisive for preventive medicine practices.

REFERENCES

[1] Özmen, Ö., Ahmad, K. H. D. R., & Engin, A. V. C. I. (2018). Sınıflandırıcıların Kalp Hastalığı Verileri Üzerine Performans Karşılaştırması. Fırat Üniversitesi Mühendislik Bilimleri Dergisi, 30(3), 153-159.

[2] Felman, A. (2018). Everything you need to know about heart disease. Medical News Today.

[3] Yaşar, B. Kalp Hastalıkları Sindirim Sistemini Etkiler mi?.

[4] Zencirli, K. (2020). Bipolar parsiyel protez uygulanmış kalça kırıklı hastalarda makine öğrenme yöntemleri ile perioperatif prognoz ve maliyet analizi.

[5] Perçin, İ., Yağin, F. H., Arslan, A. K., & Çolak, C. (2019, October). An Interactive Web Tool for Classification Problems Based on Machine Learning Algorithms Using Java Programming Language: Data Classification Software. In 2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT) (pp. 1-7). IEEE.

[6] Doğan, F., & Türkoğlu, İ. (2019). Derin öğrenme modelleri ve uygulama alanlarına ilişkin bir derleme. Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi, 10(2), 409-445.

[7] Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. Journal of big data, 2(1), 1-21.

[8] Hajkowicz, S., Karimi, S., Wark, T., Chen, C., Evans, M., Rens, N., ... & Tong, K. J. (2019). Artificial Intelligence: Solving problems, growing the economy and improving our quality of life.

[9] Nabiyev, V. V., & Zeka, Y. (2010). Seçkin Yayıncılık, 3. baskı.

[10] SAKARYA, Ş., & YILMAZ, Ü. (2019). Derin öğrenme mimarisi kullanarak bist30 indeksinin tahmini. European Journal of Educational and Social Sciences, 4(2), 106-121.

[11] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," nature, vol. 521, pp. 436-444, 2015.

[12] M. E. TAŞÇI and R. ŞAMLI, "Veri Madenciliği İle Kalp Hastalığı Teşhisi," Avrupa Bilim ve Teknoloji Dergisi, pp. 88-95, 2020.

[13] Kolaç, N. K. Vardiyalı Çalışanlarda Uykusuzluk ve Kalp Hastalıkları Riskleri: Sistematik Derleme. Arşiv Kaynak Tarama Dergisi, 30(1), 13-21.

[14] Tüfekçi, M., & Karpat, F. Derin Öğrenme Mimarilerinden Konvolüsyonel Sinir Ağları (CNN) Üzerinde Görüntü İşleme-Sınıflandırma Kabiliyetininin Arttırılmasına Yönelik Yapılan Çalışmaların İncelenmesi.

[15] Bılgıç, A., Kurban, O. C., & Yildirim, T. (2017, May). Face recognition classifier based on dimension reduction in deep learning properties. In 2017 25th Signal Processing and Communications Applications Conference (SIU) (pp. 1-4). IEEE.

[16] Yağin, F. H., Güldoğan, E., Ucuzal, H., & Çolak, C. A Computer-Assisted Diagnosis Tool for Classifying COVID-19 based on Chest X-Ray Images. Konuralp Medical Journal, 13(S1), 438-445.

 ISSN: 2548-0650 http://dergipark.gov.tr/jcs

BIOGRAPHIES

İpek Balıkçı Çiçek obtained her BSc. degree in mathematics from Çukurova University in 2010. She received MSc. degree in biostatistics and medical informatics from the Inonu University in 2018. She currently continues Ph.D. degrees in biostatistics and medical informatics from the Inonu University. In 2014, she joined the Department of Biostatistics and Medical Informatics at Inonu University as a researcher assistant. Her research interests are cognitive systems, data mining, machine learning, deep learning.

Zeynep Küçükakçalı obtained her BSc. degree in mathematics from Çukurova University in 2010. She received MSc. degree in biostatistics and medical informatics from the Inonu University in 2018. She currently continues Ph.D. degrees in biostatistics and medical informatics from the Inonu University. In 2014, she joined the Department of Biostatistics and Medical Informatics at Inonu University as a researcher assistant. Her research interests are cognitive systems, data mining, machine learning, deep learning.