THE JOURNAL OF COGNITIVE SYSTEMS
Journal Homepage: https://dergipark.org.tr/en/pub/jcs
Vol.7, No.2, 2022

Risk Prediction Model for Dementia by Deep Learning Using Clinical Data

[1]Onural ÖZHAN, [2]Zeynep Küçükakçalı, [3]İpek Balıkçı Çiçek

[1] Inonu University, Faculty of Medicine, Department of Medical Pharmacology, Malatya, Turkey. (e-mail: onural.ozhan@inonu.edu.tr)

[2,3]Inonu University, Faculty of Medicine, Department of Biostatistics and Medical Informatics, Malatya, Turkey. (e-mail: { zeynep.tunc, ipek.balikci }@inonu.edu.tr).

ARTICLE INFO

Received: July.,09.2022
Revised: Aug.,28.2022
Accepted: Sep.,17.2022

Keywords:
Dementia,
Classification,
Machine learning,
Deep learning,
Variable importance.

Corresponding author: Onural Özhan
✉ onural.ozhan@inonu.edu.tr
☎ +90 535 932 33 44

ISSN: 2548-0650

DOI: https://doi.org/10.52876/jcs. 118828

ABSTRACT

It is estimated that dementia, which is the most important public health problem in the elderly, will increase day by day. It is stated that this situation will create great challenges for public health and aged care systems in all countries of the world. For this reason, it has become very important to determine the management and treatment procedures of dementia, to reduce the level of progression of the disease and to increase the quality of life of individuals exposed to the disease. The purpose of this study is to predict dementia and reveal the factors related to the disease with the deep learning approach.

In the current study, open-access dementia data, which includes the information of 376 patients, was used. Dementia prediction was made using the deep learning method. Model results were evaluated with accuracy, balanced accuracy, sensitivity, selectivity, positive predictive value, negative predictive value, and F1-score performance metrics. In addition, 10-fold cross-validation method was used in the modeling phase. Finally, variable importance values were obtained by modeling.

When the results are examined The highest metric values among the performance criteria achieved for group variable types were calculated for Demented; and were found that Accuracy, Sensitivity, Specificity, Positive predictive value, Negative predictive Value, and F1-score were 0.947, 0.946, 0.978, 0.966, 0.965 and 0.956 respectively.

As a result, when the findings obtained from this study were examined, the dementia dataset, which consisted of imaging data and information about patients with clinical data, was classified with high accuracy using the deep learning method. The risk factors for dementia were determined with the variable importance values obtained as a result of the model.

1. INTRODUCTION

DEMENTIA is one of the most important public health problems frequently seen in the elderly, and it occurs with chronic and usually progressive decrease in memory and intellectual capacity. It is characterized by impairment in at least two cognitive functions such as memory, speech, perception, calculation, orientation, judgment, abstract thinking, and problem solving [1,2].

Dementia is very common nowadays. In 2015, it was determined that there were 47.47 million people diagnosed with dementia in the world. It is predicted that the number of people diagnosed with dementia will increase day by day, and it is expected that there will be 75.63 million people diagnosed with dementia in 2030 and 135.46 million in 2050 [3].

In parallel with the rapid increase in the elderly population, diseases that are common in old age are becoming a problem in the society. Dementia disease, which is encountered as a major health problem in aging societies and develops due to vascular-degenerative changes, follows an irreversible and progressive process. The main goals in the care of patients with dementia are to protect the patient's functionality, reduce his disability, regulate the environment and relationships to maintain stability, compensate for the losses caused by the disease, and provide a therapeutic environment that will protect the patient's individuality and sustain the quality of life. In order to achieve these goals, the risk factors of the disease should be known and the living conditions should be supported by stabilizing the patients. Therefore, studies are needed to determine the risk factors for the disease [4].

Deep learning is a form of machine learning that enables computers to learn from experience and understand the world in terms of a hierarchy of concepts. Deep learning is based on the principle that each sequential layer receives the output of the previous layer as its input [5]. It's a representation learning algorithm made up of a number of complex non-

linear transformation structures, as well as multiple intermediate processing layers, that can perform high-level abstraction on data automatically [6].

Deep neural networks, like shallow neural networks, are feed-forward networks with one or more hidden layers that can provide modeling for extremely complex non-linear problems [7]. Deep learning models are preferred to obtain results such as diagnosis, classification, estimation, and detection . The use of machine learning methods in the early diagnosis of diseases in the field of health has increased. Biomedical, health bioinformatics and medical imaging are among the areas where machine learning methods are most popular. It is seen that deep learning methods that are successful in analyzing big data with different network architectures and learning algorithms will help healthcare professionals both in the early diagnosis of diseases and in the early treatment of the disease [8].

The aim of this study is to use a deep learning approach on an open access dementia data set to identify dementia situations and evaluate relevant factors.

2. MATERIAL and METHODS

2.1. Data set

The open access "Dementia Classification" data set to be used in the study can be accessed at https://www.kaggle.com/deepak525/dementia-classification-compare-classifiers. There are 373 patients in the data set used. 146 of these patients are demented, 190 non-demented patients and 37 converted patients. The explanations of the variables are given in Table 1.

TABLE I
THE EXPLANATIONS OF THE VARIABLES

Variables	Explanations of Variable
CDR	Clinical Dementia Rating
MMSE	Mini Mental State Examination
VISIT	Number of visit
ETIV	Estimated Total Intracranial Volume
SES	Socioeconomic Status
MRDELAY	"-"
MF	Sex
EDUC	Years of Education
ASF	Atlas Scaling Factor
NWBV	Normalize Whole Brain Volume
Age	-

2.2. Deep Learning

Deep learning is a subfield of machine learning that includes methods for learning representations at various levels on data with complex relationships. Deep organized learning, or deep learning in its short form, has opened a new chapter in the field of machine learning since 2006 [5]. Deep learning, in its most general form, is a machine learning method that is used to solve problems and perform behavior such as analysis, inference, observation, and learning using large quantities of data, according to the descriptions in the literature. They can be in different hierarchical systems than conventional machine learning algorithms [9].

Artificial neural networks (ANNs) are a subfield of deep learning (ANN). It's a type of artificial neural network that uses nonlinear transformations to get a specific output value from raw data. The backpropagation algorithm is used in deep learning to investigate the dynamic structure of multidimensional data sets. This is accomplished by comparing the values of the parameters measured in each layer to the values obtained in the previous layer. Multi-layer neural networks have an input layer that represents the inputs, hidden layers that transform the information from the input layer into an output, and an output layer that converts the results from the last hidden layer into output values. The success rate in the fields of natural language processing, image processing, visual object identification, and drug discovery has increased significantly thanks to deep learning methods [10].

2.3. Data Analyses and Modelling

Qualitative data were expressed as number (percentage), while quantitative data were expressed as mean ± standard deviation, and median (minimum-maximum). The Kolmogorov-Smirnov test was used to assess conformity to the normal distribution. The Kruskal-Wallis, one-way analysis of variance, and Pearson chi-square tests were used to determine whether there is a statistically significant difference between the "Nondemented," "Demented," and "Converted" groups, which are the categories of the dependent/target variable (dementia) in terms of independent variables. Statistics were considered significant at $p<0.05$. All analyses used the IBM SPSS Statistics 26.0 package program.

A 10-fold cross-validation procedure was employed to determine the model's validity. The 10-fold cross-validation method divides the entire set of data into 10 equal portions. The method is performed ten times using one portion as a test set and the other nine parts as training data sets. In addition, accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and F1-score criteria were taken into account for the performance evaluation of the model.

3. RESULTS

Descriptive statistics for the independent variables examined in this study are given in Table 2. According to the findings in Table 2; there is a statistically significant difference between the dependent / target variable groups in terms of Mr delay, educ, mmse, cdr, nwbv, age variables ($p<0.05$).

According to the findings in Table 3; There is a statistically significant relationship between the mf, ses variables and the dependent / target variable (class) groups ($p<0.05$).

TABLE II
DESCRIPTIVE STATISTICS FOR QUANTITATIVE INDEPENDENT VARIABLES

Variables	Grup						*p-value*
	Nondemented		Demented		Converted		
	Median (min-max)	Mean± Standard deviation	Median (min-max)	Mean± Standard deviation	Median (min-max)	Mean ± Standard deviation	
MRDELAY	631[a] (0-2517)	-	491[b] (0-2508)	-	706 (0-2639)	-	**0.006 ****

EDUC	16[a] (8-23)	-	13[b] (6-20)	-	16 (12-20)	-	**<0.001****
MMSE	29[a] (26-30)	-	26[b] (4-30)	-	29 (24-30)	-	**<0.001****
CDR	0[a,b] (0-0.5)	-	0.5[b] (0.5-2)	-	0.5 (0-0.5)	-	**<0.001****
ETIV	1475 (1106-2004)	-	1477 (1143-1957)	-	1423 (1264-1722)	-	0.517**
NWBV		0.741[a,b] ±0.038		0.716± 0.032		0.724± 0.035	**<0.001*****
ASF		1.191±0.144		1.197± 0.0137		1.212± 0.109	0.5683***
Age		77±8		76[b]±7		80±7	**0.045*****

*: a: different according to the Demented group, b: different according to the Converted group
: Kruskal Wallis test, *: One way analysis of variance test

TABLE III
DESCRIPTIVE STATISTICS FOR QUALITATIVE INDEPENDENT VARIABLES

Variables		Group			p- value**
		Nondemented	Demented	Converted	
		Count (%)	Count (%)	Count (%)	
Visit	**1**	72 (37.9)	64 (43.8)	14 (37.8)	0.339
	2	70 (36.8)	62 (42.5)	12 (32.4)	
	3	34 (17.9)	16 (11.0)	8 (21.6)	
	4	10 (5.3)	3 (2.1)	2 (5.4)	
	5	4 (2.1)	1 (0.7)	1 (2.7)	
MF	**Male**	61 (32.1)	86 (58.9)	13 (35.1)	<0.001
	Female	129 (67.9)	60 (41.1)	24 (64.9)	
SES	**1**	41 (21.6)	26 (20.5)	21 (56.8)	<0.001
	2	71 (37.4)	25 (19.7)	7 (18.9)	
	3	42 (22.1)	33 (26.0)	7 (18.9)	
	4	34 (17.9)	38 (29.9)	2 (5.4)	
	5	2 (1.1)	5 (3.9)	0 (0.0)	

**: Pearson chi-square test

TABLE IV
DESCRIPTIVE STATISTICS FOR QUALITATIVE INDEPENDENT VARIABLES

Metrics \ **Tissue types**	Nondemented	Demented	Converted
Accuracy	0.909	0.947	0.893
Sensitivity	0.908	0.946	0.824
Specificity	0.988	0.978	0.935
Positive predictive Value	0.989	0.966	0.378
Negative predictive Value	0.896	0.965	0.991
F1-score	0.947	0.956	0.519

**: Pearson chi-square test

The highest metric values among the performance criteria achieved for group variable types were calculated for Demented; and we found that Accuracy, Sensitivity, Specificity, Positive predictive value, Negative predictive Value, F1-score were 0.947, 0.946, 0.978, 0.966, 0.965 and 0.956 respectively.

The values showing the significance of the variables are shown in Table 5. Thefive most important variables were found cdr (0.1184), mmse (0.1005) and visit (0.0990), etıv (0.0909), ses (0.0903).

TABLE V
DESCRIPTIVE STATISTICS FOR QUALITATIVE INDEPENDENT VARIABLES

EXPLANATORY VARİABLES	**Importance**
CDR	0.1184
MMSE	0.1005
VİSİT	0.0990
ETIV	0.0909
SES	0.0903
MRDELAY	0.0862
MF	0.0855
EDUC	0.0854
ASF	0.0846
NWBV	0.0807
Age	0.0779

4. DISCUSSION

The rapid aging in the world and the increase in the elderly population as a result of this is a global problem. By 2050, it is estimated that the population aged 60 and over will be around 2 billion, and the population aged 80 and over will be around 400 million [11,12]. It is stated that with this situation, the incidence of dementia will increase sharply and this disorder will create great challenges for public health and aged care systems in all countries of the world. For this reason, it has become very important to determine the management and treatment procedures of dementia, to reduce the level of progression of the disease and to increase the quality of life of individuals exposed to the disease [13]. In this study, it was aimed to determine the risk factors associated with dementia by using the data set that included the data of patients with dementia, non-dementia, and those who developed dementia during follow-up.

Machine learning is a branch of science concerned with the design and creation of algorithms that enable computers to learn from various types of data. Machine learning is both a database problem and a branch of artificial intelligence that enables current data to be used to predict and simulate future events[14]. Deep learning is a form of machine learning that is becoming increasingly common. The Deep learning approach has grown in popularity as graphics processing units (GPU) have improved [9]. The deep learning approach is designed inspired by the working principle of the brain. It is a process that was developed using the example of human learning systems and consists of several hidden layers and neurons. The deep learning method is used to process large data sets by contributing hardware features that evolve over time (especially the participation of graphics processors in the calculation). Motion detection, face recognition, health technologies, object recognition, and object detection are all fields where deep learning is used [15].

In the study, deep learning method was applied to the Dementia data set which is an open-source data set. When three different types of group variable were classified, the accuracy value was obtained as 91.96. In a study using the same data set, the highest accuracy obtained from

classification models made with different machine learning methods and voting method is 91.17% [16]. In the current study, higher accuracy was obtained in the classification made with the deep learning model. Demented is the one with the best performance metrics from three different categories for the group variable. Deep learning method gave successful predictive results in the classification of dementia according to the results of the performance criteria calculated in this study. In addition, the importance of risk factors related to dementia was obtained with the experimental findings. The most important 3 factors that may be associated with having a dementia were obtained as cdr, mmse, visit.

Detection and classification of dementia have a very important place in the diagnosis of neurogenerative disorders. As a result, when the findings obtained from this study were examined, the dementia dataset, which consisted of imaging data and information about patients with clinical data, was classified with high accuracy using the deep learning method. The risk factors for dementia were determined with the variable importance values obtained as a result of the model.

REFERENCES

[1] G. Demir, "Demans Ve Hemşirelik Bakımı," Black Sea Journal Of Health Science, Vol. 1, Pp. 35-39.

[2] M. Braun, U. Scholz, B. Bailey, S. Perren, R. Hornung, and M. Martin, "Dementia caregiving in spousal relationships: a dyadic perspective," Aging and mental health, vol. 13, pp. 426-436, 2009.

[3] M. Prince, M. Guerchet, and M. Prina, "Alzheimer's Disease International (2013b) Policy Brief for Heads Of Government: The Global Impact of Dementia 2013-2050," Alzheimer's Disease International, London, 2014.

[4] Ö. G. D. İ. Akyar, "Demanslı hasta bakımı ve bakım modelleri," Hacettepe Üniversitesi Hemşirelik Fakültesi Dergisi, vol. 18, pp. 79-88, 2011.

[5] L. Deng and D. Yu, "Deep learning: methods and applications," Foundations and trends in signal processing, vol. 7, pp. 197-387, 2014.

[6] Y. Bengio and Y. LeCun, "Scaling learning algorithms towards AI," Large-scale kernel machines, vol. 34, pp. 1-41, 2007.

[7] A. Şeker, B. Diri, and H. H. Balık, "Derin öğrenme yöntemleri ve uygulamaları hakkında bir inceleme," Gazi Mühendislik Bilimleri Dergisi (GMBD), vol. 3, pp. 47-64, 2017.

[8] K. Umut, A. YILMAZ, and Y. Dikmen, "Sağlık alanında kullanılan derin öğrenme yöntemleri," Avrupa Bilim ve Teknoloji Dergisi, pp. 792-808, 2019.

[9] K. Kayaalp and A. Süzen, "Derin Öğrenme ve Türkiye'deki Uygulamaları," Yayın Yeri: IKSAD International Publishing House, Basım sayısı, vol. 1, 2018.

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

[11] Y. Onadja, N. Atchessi, B. A. Soura, C. Rossier, and M.-V. Zunzunegui, "Gender differences in cognitive impairment and mobility disability in old age: a cross-sectional study in Ouagadougou, Burkina Faso," Archives of gerontology and geriatrics, vol. 57, pp. 311-318, 2013

[12] Ü. Görgülü and N. Akdemir, "İleri evre kanser hastalarına bakım verenlerin yorgunluk ve uyku kalitesinin değerlendirilmesi," Genel Tip Dergisi, vol. 20, 2010.

[13] M. N. Özata Değerli, "Orta Evre Alzheimer Tip Demans Tanılı Bireylerde Demansa Bağlı Davranışsal Ve Psikolojik Semptomlar İle Duyusal İşlemleme Süreci Arasındaki İlişkinin İncelenmesi," 2022.

[14] E. Alpaydin, Introduction to machine learning: MIT press, 2020.

[15] J. Schmidhuber, "Deep learning in neural networks: An overview," Neural networks, vol. 61, pp. 85-117, 2015.

[16] S. Bharati, P. Podder, D. N. H. Thanh, and V. Prasath, "Dementia classification using MR imaging and clinical data with voting based machine learning models," Multimedia Tools and Applications, pp. 1-22, 2022.

BIOGRAPHIES

Onural Özhan obtained her BSc. degree in pharmacy from Inonu University in 2009. He received MSc. degree in pharmacognosis from Ankara University in 2012. He received Ph.D. degree in pharmacology from Inonu University in 2019. In 2015, he joined the Department of pharmacology at Inonu University as a research assistant. Currently, he continues to work as an asst. prof. at Inonu University. His research interests are the cardiovascular system, oxidative stress, and ischemia-reperfusion.

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.

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.