Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi

Afyon Kocatepe University Journal of Science and Engineering

AKÜ FEMÜBİD **23** (2023) 045102 (955-961) AKU J. Sci. Eng. **23** (2023) 045102 (955-961)

DOI: 10.35414/akufemubid.1239360

Araştırma Makalesi / Research Article

Detection of Autistic Spectrum Disorder Using Artificial Neural Network

Şeyma Nur ÖZDEMİR[1], Kazım YILDIZ[2]

[1] *Marmara University, Institute of Pure and Applied Sciences, Computer Engineering, İstanbul, Türkiye*

[2] *Marmara University, Faculty of Technology, Computer Engineering, İstanbul, Türkiye*

Corresponding author e-mail:seymaozdemir22@marun.edu.tr ORCID ID: https://orcid.org/0000-0002-1033-4501*

e-mail : kazim.yildiz@marmara.edu.tr ORCID ID: https://orcid.org/0000-0001-6999-1410

Geliş Tarihi: 19.01.2023 Kabul Tarihi: 18.08.2023

Keywords
Autism Spectrum Disorder;Early Diagnosis System; Random Forest; Artificial Neural Network.

Abstract

Autistic Spectrum Disorder (ASD) is a neuro-developmental disorder that is congenital or manifests with a delay in social relations and physiological development at an early age, and also causes problems in communication. It is possible to reduce the effect of the disease on individuals with early diagnosis. However, detecting ASD at an early age requires time and cost. In the studies conducted in recent years, it is seen that there is a serious increase in ASD cases. In order to prevent this increase, decision support systems should be established for early diagnosis. It is important to develop decision support models to diagnose ASD, especially for children aged 12-36 months. In this study, a model was developed that can help in detecting ASD with high accuracy for 12-36 months old children. The data set used in the created model was collected from the mobile application named ASDTests developed by Thabtah. In the estimation phase, four different machine learning algorithms which are support vector machine, Naive Bayes,Random Forest and Artificial Neural Network were used. In the classification process, high success rate was obtained with artificial neural network, random forest classifier.

Otistik Spektrum Bozukluğunun Yapay Sinir Ağları ile Tespiti

Anahtar kelimeler
Otistik Spektrum Bozukluğu; Erken Tanı Sistemi; Rasgele Orman; Yapay Sinir Ağları.

Öz

Otistik Spektrum Bozukluğu (OSB), doğuştan gelen yada yaşamın ilk yaşlarında sosyal ilişkilerde ve fizyolojik gelişimde gecikme ile kendini gösteren ve aynı zamanda iletişimde sorunlara neden olan nöro-gelişimsel bir bozukluktur. Hastalığın bireyler üzerinde etkisinin erken tanı ile azaltılması mümkündür. Ancak OSB'yi erken yaşta tespit etmek zaman ve maliyet gerektirmektedir. Son yıllarda yapılan çalışmalarda OSB vakalarında ciddi bir artış olduğu görülmektedir. Bu artışı önlemek için erken tanı için karar destek sistemleri oluşturulmalıdır. Özellikle 12-36 aylık çocuklar için OSB tanısı koymak için karar destek modellerinin geliştirilmesi önem taşımaktadır. Bu çalışmada 12-36 aylık çocuklar için yüksek doğrulukta OSB tespitinde yardımcı olabilecek bir model geliştirilmiştir. Oluşturulan modelde kullanılan veri seti Thabtah tarafından geliştirilen ASDTests isimli mobil uygulamadan toplanmıştır. Tahminleme aşamasında destek vektör makinesi, Naive Bayes, rasgele orman , yapay sinir ağları olmak üzere dört farklı makine öğrenimi algoritması kullanılmıştır. Sınıflandırma sürecinde yapay sinir ağları, rasgele orman sınıflandırıcı ile yüksek başarı oranı elde edilmiştir.

1. Introduction

Autistic Spectrum Disorder (ASD) is a neuro-developmental difference that occurs at birth or at an early age. Research shows that early diagnosis and education will significantly reduce the problems to be experienced in the future.

The increase in ASD cases worldwide in recent years reveals the need for effective and easy-to-use diagnostic methods. Among these diagnostic methods, there is also the development of helpful decision support software that will assist specialists and reduce the time to clinical diagnosis for patients.

ASD is a developmental disability caused by differences in the brain (Int Kyn. 1). There are many situations that scientists have uncovered about individuals with ASD. The exact cause is not yet known. ASD begins to appear between the ages of 1-3. Over time, the disease begins to progress, and they gradually lose their acquired abilities.

Human genes can be adversely affected by environmental factors. There are some behaviors that people with ASD show: not being able to make eye contact, rocking movements, interest in rotating objects (Mohanty *et al*. 2021). Diagnosing autism takes a significant amount of time and cost. (Omar *et al*. 2019)

Machine learning methods in the field of health are used to detect many diseases due to their advantages such as cost and time. It is aimed to take faster results and early measures by using machine learning methods in ASD detection. Zhao *et al*. (2019) proposed a model using eye tracking data obtained from face-to-face interviews. In this study, four different machine learning techniques were used. These are SVM, Linear Discrimination Analysis, Decision Tree and Random Forest. As a result, 92.31% success rate was obtained with the SVM classifier.

Vaishali *et al*. (2018) aimed to develop a prediction model for ASD detection. In this study, a 21-feature ASD dataset from UCI was used. This study claims that machine learning models trained with minimal behavioral data are more successful. A success rate of 99.66% was obtained with the SVM and Multi Layer Perceptron method.

Thabtah and Fadi (2018) proposed a new mobile application by developing an autism spectrum disorder detection tool called ASDTests. With the application, 1452 samples were collected. It is aimed to predict the disease using logistic regression and naive bayes algorithm. In the study, it was shown that logistic regression was the most successful classification algorithm with a rate of 97.94% in the child category.

Erkan and Thanh (2019) developed a model for the diagnosis of ASD in children, adolescents and adults in their study. They used K-Nearest Neighbors (KNN), SVM and Random Forest (RF) algorithms. In terms of performance, the RF algorithm showed 100% success.

Baranwal *et al*. (2020) created a prediction model for adults, children and adolescents. In this study, they used the ASD scanning dataset. Artificial Neural Networks (ANN), RF, Logistic Regression, Decision Tree and SVM were used. ANN gave the best accuracy rate about 98.15%.

Mohanty *et al*. (2021) analyzed all individual categories of 2154 samples and 21 features from the UCI ML repository and Kaggle. It has been stated that using the Deep Neural Network (DNN) classifier algorithm performed acceptable for all object categories.

Metlek *et al.(*2020) used six different machine learning methods as supervised and unsupervised in their study. In the observed learning times used, 100% classification success rate was obtained in the operation process with support vector machines.

Al-Diabat (2018) used fuzzy data mining methods for ASD estimation. Fuzzy data mining classification algorithms FURIA (Dahe *et al*. 2021), JRIP (Catania 2021), RIDOR (Negin *et al.* 2021) and PRISM (Hadi *et al*. 2017) were used in the study. The FURIA classification model performed better. The aim of the study is to diagnose ASD in children aged 12-36 months.

Studies are aimed at diagnosing ASD in children. For this purpose, an important contribution has been made to the literature. First of all, there is a need for data sets containing behavioral characteristics of people diagnosed with ASD. Except for some European countries, there is no comprehensive dataset on the subject. Such datasets enable more detailed analyzes to be made by increasing the efficiency, sensitivity, specificity and performance ratio of the software to be developed to shorten the ASD diagnosis process. The data set provided by the Manukau Institute of Technology, which provides international source data, to researchers interested in ASD, was used in this study. The dataset has 18 features. There are 10 questions about ASD in the dataset. There is also a feature that includes the score in return for the answers to these questions. SVM, Naive Bayes, RF and ANN were used in the study. This paper is organized as follows: Section 2 describes material and method. Section 3 shows the experimental results. And finally conclusion part is given.

2. Material and Methods

2.1 İkinci dereceden başlık

In the application, 1054 records containing 18 different features presented by Manukau Institute of Technology for the detection of autistic spectrum disorder in infants (12-36 months) were used (Int Rfn. 3). A data set was created with the answers to the questions prepared according to the last parameters accepted in the literature for the diagnosis of ASD. Table 1 shows the attibutes of the dataset.

Table 1. The attributes of the dataset

No	Field Name	Description
1	Questions 1	The situation where the child looks at you when called by his/her name
2	Questions 2	The situation of making eye contact with the child
3	Questions 3	The state of the child to point out when he wants something
4	Questions 4	The state of the child pointing out something that he or she finds interesting
5	Questions 5	The child's state of playing games that require imitation
6	Questions 6	The state of the child to follow the place where the other person is looking with his eyes
7	Questions 7	If someone in the family is unhappy, the child's reaction to the person concerned
8	Questions 8	How to identify the first word the child speaks
9	Questions 9	The child's use of simple body movements to communicate
10	Questions 10	Child's focus on one place for no reason
11	Age(Months)	Age
12	Sex	Male or Female
13	Ethnicity	Ethnicities
14	Bom with jaundice	Have jaundice
15	Family member with ASD	Is there a family history of autistic spectrum disorder?
16	Who entered the data (Family/Another)	Family, health personel,etc.
17	Score	The score obtained according to the answers given to the questions in the test
18	Class	Ground truth

The answers were collected with the ASDTests app. Among the attributes used as input in the data set used, there are ten questions measuring the ASD as well as information containing demographic characteristics. The data set includes numeric and categorical data types. Some samples in the dataset have missing data. In the data set consisting of a total of 18 features; 17 features were used as input and 1 feature was used as output because it contained the result. First, all attributes were converted to numerical values In this study. These values are shown in Table-2. In order to be able to analyze the data set more effectively, data preprocessing steps such as completing the missing data and normalizing the data between 0 and 1 were carried out. The missing data on attributes such as ethnicity, kindship, score and age were excluded from the dataset.

Table 2. The numeric data

Gender		Have Jaoundice		Family member with ASD		Who entered the data	
female	0	No	0	No	0	Healthcare Professional	1
male	1	Yes	1	Yes	1	Others	2
-	-	-	-	-	-	Parent	3

2.2. Performance Metrics

The results obtained in the estimations made with the test data are expressed with the error matrix. Table-3 shows the confusion matrix. The key metrics used to evaluate the success of classifiers are F-Measure, precision, and accuracy(Nıshat *et al*, 2021).

Table 3. Confusion Matrix

		Actual	
		ASD	**Not ASD**
Predicted	**ASD**	TP(*True Positive*)	FP(*False Positive*)
	Not ASD	FN(*False Negative*)	TN(*True Negative*)

They are defined as equation1,2 and 3 respectively:

$$accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (1)$$

$$sensitivity = \frac{TP}{TP+FN} \quad (2)$$

$$F - measure = \frac{2*TP}{(2*TP+FP+FN)} \quad (3)$$

2.3. Classification Algorithms

In this section, information about machine learning methods will be given.

2.3.1 Support Vector Machine

SVM (Cortes *et al*. 1995) is a powerful machine learning algorithm based on statistical theories. Traditional AI algorithms require large amounts of training data for learning. In addition, overfitting problems are encountered. However, with the DVM algorithm, successful results are obtained with multidimensional and less data. The purpose of SVMs is to find the separator plane with the highest distance between classification algorithms, as shown in Figure 1. x and y axes indicate points of data. w; weight, x; input vector, b; represents the amount of deviation. If the result of a new value is less than 0, it will be closer to the dots on the right. Conversely, if the result is greater than or equal to 0, it will be closer to the left dots. In this study, linear SVM was used. The cost value for the SVM algorithm was determined as 1 and the gamma value as 0.1. While determining these values, cost and gamma values were obtained by using k-cross validation method, in which svm performed best.

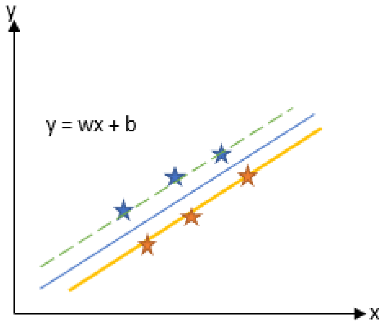

Figure 1. SVM Classifier (Linearly Separable Data) (Cortes *et al*. 1995)

2.3.2 Naive Bayes

It is a simplified version of Bayes' theorem with independent evaluation of properties. It works based on Bayes' Theorem, which provides a way tocalculate the probability of hypothesis based on our prior knowledge (Nayeem *et al*. 2021). The posterior probability is computed as equation 4:

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)} \quad (4)$$

Laplace smoothing technique is used for the Naive Bayes classifier which process the problem of zero probability in Naïve Bayes.

2.3.3 Random Forest

It is aimed to join the decisions of trees with many variables, each of which is trained on a different tree, instead of producing a single decision tree. The success and accuracy of the result of different data groups with a single classifier increases (Breiman 2001). Random forest is a collective learning method in which more than one decision tree is used. Collective methods are learning algorithms that generate multiple classifiers instead of one, and then classify new data with votes from their predictions. It uses the decision tree structure while performing the classification process. The number of trees determined for the random forest classifier is 400. The importance parameter is set to true in order to use the importance levels of this algorithm in the dataset.

2.3.3 Artificial Neual Networks (ANN)

It is a mathematical model which is revealed by the human brain (NASSER *et al*. 2019). ANN relies on its ability to adapt to changing situations by adjusting its connections (weights) and learning from experience. Artificial neurons are organized into layers, as shown in Figure (2). Each layer consists of a group of neurons that perform similar functions. Input, output and hidden layers are type of the layers. The hidden layer is optional. The input and output layers are mandatory. In this study, multiple layer perceptron (MLP) and deep neural network (DNN) algorithms were used from the ANN family. For DNN, predictions were made using a model with two hidden layers of sizes (30, 3).

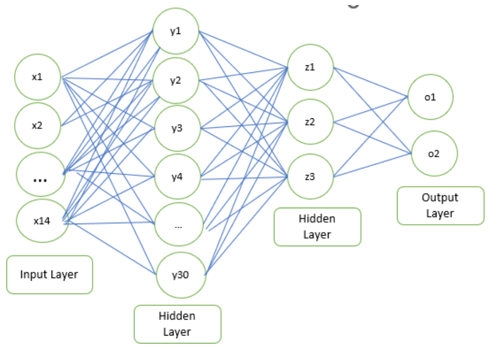

Figure 2. ANN Architecture

The flowchart of the study is depicted in Figure-3.

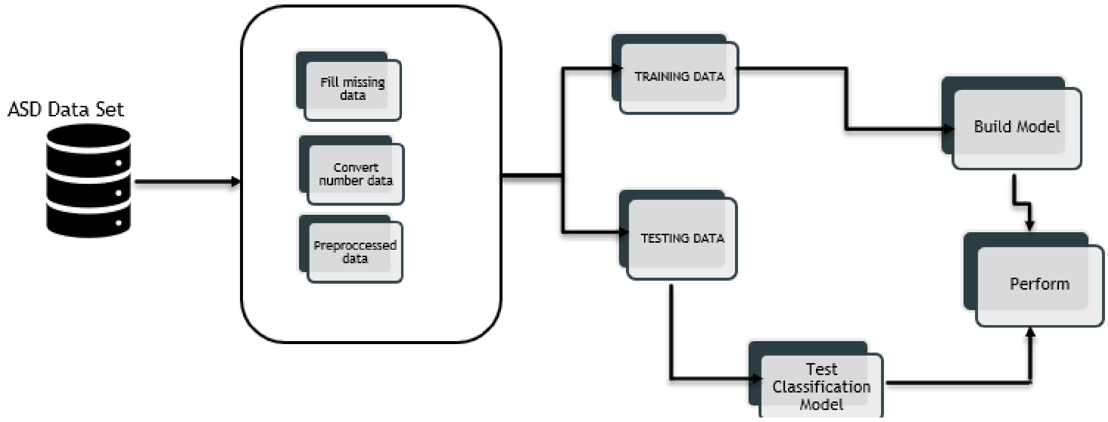

Figure 3. The flowchart of the proposed model

3. Results and Discussion

After the preprocessing, the dataset is divided as training and test. In order to obtain the most successful result of the model, the dataset is divided at different rates. Based on these ratios, the F-measure, accuracy and sensitivity values were calculated. Table 4 shows the accuracy, sensitivity and F-measure values of the classification algorithms.

As can be seen, SVM, random forest,DNN and MLP algorithms are 100% successful in all cases. Although the success rate of the Naive Bayes algorithm is lower, it is also quite effective.
As a result, it was seen that DNN and MLP gave very successful results in detecting ASD. The model that can be used in the diagnosis of ASD has been developed for children between 12-36 months. It has been shown that faster and more successfull detection of ASD can be achieved using DNN and MLP.
When examining the results obtained in this study, it is evident that high success rates were achieved using ANN algorithms. The use of ANN algorithms for ASD detection has been emphasized, as they can demonstrate high performance even with larger datasets

4.Conclusion

Individuals with ASD need early diagnosis. The earlier the disease is diagnosed, the slower the rate of progression. With the development of technology, many diseases can be diagnosed early.
Models created using data sets of individuals with ASD are important for early diagnosis. Based on the autism screening test data provided by the Manukau Institute of Technology, various machine learning algorithms were used to find the appropriate model. An artificial neural network and classical machine learning algorithm models were proposed for autism diagnosis. In the artificial neural network, the input factors were obtained from the dataset of an autism screening application, representing detailed screening results of the users. The ANN model was tested and achieved 100% accuracy in the overall results. This study demonstrated the capability of artificial neural networks in diagnosing ASD.
Obtained results show that ANN, RF and SVM methods give high success rates. Machine learning

methods and artificial neural networks should be applied by replicating the data and the results should be re-evaluated. In future studies, The aim is to utilize deep learning algorithms for autism diagnosis using brain MRI images.

Table 4. Classification result for data set

Tests	Methods	Accuracy	Sensitivity	F-measure
Train(50%) Test(50%)	NaiveBayes	94.47%	91.87%	91.02%
	Random Forest	100%	100%	100%
	SVM	100%	100%	100%
	MLP	100%	99.72%	99.86%
	DNN	100%	100%	100%
Train(60%) Test(40%)	Naive Bayes	97.33%	94.44%	95.54%
	Random Forest	100%	100%	100%
	SVM	100%	100%	100%
	MLP	100%	100%	100%
	DNN	100%	100%	100%
Train(70%) Test(30%)	Naive Bayes	98.6%	95.70%	97.26%
	Random Forest	100%	100%	100%
	SVM	100%	100%	100%
	MLP	100%	100%	100%
	DNN	100%	100%	100%
Train(80%) Test(20%)	Naive Bayes	98.46%	96.72%	97.52%
	Random Forest	100%	100%	100%
	SVM	100%	100%	100%
	MLP	100%	100%	100%
	DNN	100%	100%	100%
Train(90%) Test(10%)	Naive Bayes	100%	100%	100%
	Random Forest	100%	100%	100%
	SVM	100%	100%	100%
	MLP	100%	100%	100%
	DNN	100%	100%	100%

5.References

Al-Diabat, M, 2018. Fuzzy data mining for autism classification of children. *Int. J. Adv. Comput. Sci. Appl.*, **9.**

Baranwal, A. and Vanitha, M., 2020. Autistic spectrum disorder screening: prediction with machine learning models. In *2020 International conference on emerging trends in information technology and engineering*,1-7.

Beykikhoshk, A., Arandjelovi´, O., Phung, D., Venkatesh, S., and Caelli, T. ,2014. Data-mining twitter and the autism spectrum disorder: *2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014)* , 349 – 356

Breiman, L., 2001.Random forests. *Machine learning*, **45**, 5-32.

Catania, L. J., 2021. AI applications in prevalent diseases and disorders. *Foundations of Artificial Intelligence in Healthcare and Bioscience, Elsevier*, 293-444.

Cordova, M., Shada, K., Demeter, D. V., Doyle, O., Miranda-Dominguez, O., Perrone, A., and Feczko, E. ,2020. Heterogeneity of executive function revealed by a functional random forest approach across ADHD and ASD. *NeuroImage: Clinical*, ***26***, 102245.

Cortes, C. and Vapnik, 1995. V. Support-vector networks. *Mach Learn* **20**, 273–297

Duda, M., Ma, R., Haber, N., and Wall, D., 2016. Use of machine learning for behavioural distinction of autism. *Translational Psychiatry*, **221**.

Erkan, U. and Thanh, D. , 2019. Autism spectrum disorder detection with machine learning methods. *Current Psychiatry Research and Reviews Formerly*,**15**, 297-308.

Gultekin, M. , and Kalipsiz, O., 2019. Story Point-Based Effort Estimation Model with Machine Learning Techniques*, Int'l Journal of Software Engineering and Knowledge Engineering*.

Hadi W., Issa G., Ishtaiwi A. ACPRISM, 2017. Associative classification based on PRISM algorithm. *Inf. Sci. (NY)*, **417** , 287-300

Metlek, S. and Kayaalp, K., 2020. Otistik Spektrum Bozukluğunun Makine Öğrenme Algoritmaları ile Tespiti. *Journal of Intelligent Systems: Theory and Applications,* **3.2** , 60-68.

Mohanty, A. S., Parida, P. and Patra, K. C. ,2021. ASD Classification for Children using Deep Neural Network. *Global Transitions Proceedings,* **2,** 461-466

Mohanty, A. S., Parida, P. and Patra, K. C., 2021. Identification of autism spectrum disorder using deep neural network. *In Journal of Physics: Conference Series* , **1,** 012006.

Nayeem, M. J., Rana, S., Alam, F., and Rahman, M. A., 2021. Prediction of Hepatitis Disease Using K-Nearest Neighbors, Naive Bayes, Support Vector Machine, Multi-Layer Perceptron and Random Forest. *2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD)*, 280-284.

Nasser, Ibrahim M.; Al-Shawwa and Mohammed O.; Abu-Naser, Samy S. Artificial neural network for diagnose autism spectrum disorder. 2019.

Negin, F., Ozyer, B., Agahian, S., Kacdioglu, S., Ozyer, G. T. ,2021. Vision-assisted recognition of stereotype behaviors for early diagnosis of Autism Spectrum Disorders. *Neurocomputing*, **446**, 145-155.

Nielsen, J. A., Zielinski, B. A., Fletcher, P. T., Alexander, A. L., Lange, N., Bigler, E. D. and Anderson, J. S. ,2013. Multisite functional connectivity MRI classification of autism: ABIDE results. Frontiers in human neuroscience, **7**, 599.

Nishat, M. M., Hasan, T., Nasrullah, S. M., Faisal, F., Asif, M. A. A. R., and Hoque, M. A, 2021. Detection of Parkinson's disease by employing boosting algorithms. In *2021 Joint 10th International Conference on Informatics, Electronics & Vision (ICIEV) and 2021 5th International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*, 1-7.

Omar, K. S., Mondal, P., Khan, N. S., Rizvi, M. R. K. and Islam, M. N., 2019. A Machine Learning Approach to Predict Autism Spectrum Disorder. *2019 International Conference on Electrical, Computer and Communication Engineering(ECCE)*, 1-6.

Rabbi, M. F., Hasan, S. M., Champa, A. I., and Zaman, M. A. , 2021 A convolutional neural network model for early-stage detection of autism spectrum disorder. *In: 2021 international conference on information and communication technology for sustainable development (icict4sd). IEEE*, 110-114.

Tapak L., Afshar S., Afrasiabi M., Ghasemi M.K. and Alirezaei P., 2021. Applicaiton of Genetic Algorithm-Based Support Vector Machine in Identification of Gene Expression Signatures for Psoriasis Classification. *Biomed Research International*, **10**.

Thabtah, F., 2019. An accessible and efficient autism screening method for behavioural data and predictive analyses, *Health informatics journal* ***25(4)***, 1739-1755.

Virendra S. Dahe, Sai G. Manikandan, Jegadeeshwaran R., Sakthivel G. and Lakshmipathi J, 2021. Tool condition monitoring using Random forest and FURIA through statistical learning Mater. *Today Proc., **46** ,*1161-1166.

Shahamiri, S. R., and Thabtah, F. ,2020. Autism AI: a new autism screening system based on artificial intelligence. *Cognitive Computation*, ***12*(4)**, 766-777.

Vaishali, R., and Sasikala R., 2018. A machine learning based approach to classify autism with optimum behaviour sets, *International Journal of Engineering & Technology* **7(4)** , 18.

References of Internet

1-https://www.cdc.gov/ncbddd/autism/facts.html, (09.12.2022)

2-http://fadifayez.com/publications/#datasets, (07.07.2017)

3-https://www.kaggle.com/datasets/fabdelja/autism-screening-for-toddlers

(06.2018)