Journal of Theoretical and Applied Information Technology
15th November 2022. Vol.100. No 21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

DETECTING AUTISM SPECTRUM DISORDER FOR TODDLERS USING MACHINE LEARNING TECHNIQUES

RIHAM ALBARAZI[1], Dr. BASSEL ALKHATIB[2]

[1]Research Scholar, Department of Web Sciences, Syrian Virtual University, Syria

[2]Professor, Department of Artificial Intelligence, Syrian Virtual University, Syria

E-mail: [1]riham.barazi@gmail.com, [2]b.alkhatib.foit@aspu.edu.sy

ABSTRACT

Autism is a disorder characterized by difficulty in social interactions, communication challenges, and repetitive behaviors. In recent years, treatments for autism have been constantly evolving, therefore, it is essential to diagnose children at an early age to be able to control their symptoms. We have used information about signals that will help us in early detection of autism, to help affected children to integrate into society and live independently.

For these reasons, this article focused on the use of data for only toddlers, and it compared deep learning and traditional classifiers for achieving efficient and accurate classification in the environment of machine learning, in contrast to previous research that focused on using traditional classifiers for different older ages. Among all applied algorithms, Support Vector Machine (SVM), Logistic Regression (LR) and Multilayer Perceptron (MLP) are perhaps worthy of further study on this problem in terms of only scores (Accuracy, Recall, Precision and F1), and only LR in terms of both scores and training runtime.

Keywords: *Autism Spectrum Disorder, Traditional Machine Learning Algorithms, Deep Learning, Classification, Toddlers.*

1. INTRODUCTION

Autism Spectrum Disorder (ASD) represents a broad range of complicated neurodevelopmental disturbances characterized by repeated and specific patterns of behavior, and difficulty with social communication and interaction [1]. This disorder appears at an early stage of life and negatively affects daily functioning [2]. The diagnosis rate is constantly increasing over the last 2 decades, 1 in 54 children are diagnosed with autism in 2020 compared to 1 in 59 in 2018, 1 in 110 in 2006, and 1 in 150 in 2000 [3].

Diagnosing ASD isn't forward because there is no specific test, such as blood or radiological test, instead, the diagnosis depends mainly on clinical approach by looking at the developmental history and behaviors of the child [3] [4]. However, the accredited tests for the diagnosis are, for example: Autism Diagnostic Observation Schedule (ADOS) and Autism Diagnostic Interview Revised (ADI-R) are only used by specialists, and besides that they are time consuming and costly. One of the many difficulties are delays in providing the required interventions and therapies due to not providing autism screening for toddlers at an early stage of their life [5] [6].

The average age of a child's actual diagnosis of autism is currently between 4 and 5 years in USA [7], while 7.2 ± 4.2 years in Japan [8]. Thus, there is a need to minimize the time between the start of symptom appearance of ASD and the actual detection. Indeed, early identification is critical to improve long term results related to cognition, language, adaptive behavior, daily activity and social activity behavior [9].

Recently, Artificial Intelligence (AI) has been used as a diagnosis assistant in many medical fields. It represents a wide spectrum of technologies that can perform cognitive tasks by simulating human intelligence, so we can use it to intervene in a more precise manner in determining the target at the right time.

Machine Learning (ML) is considered one of the most commonly used subfields of AI in research, e.g., detect depression [10]. It can be applied to a variety of tasks like classifications, recommendations, clustering, and prediction. They can be used in health care to produce results that allow early treatment, increase diagnostic accuracy, develop a more accurate treatment plan, and possibly also reduce human errors, thereby allocating more time to caring for patients rather than wasting it on diagnosing them.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ML can take a supervised approach by learning itself with a classified dataset and establishing the best fitting algorithm to predict an outcome of interest, and it can enhance our realization of ASD and may further help construct a strong basis for better screening and diagnosis [11].

Recently, a few scholars in the ASD research field have investigated ML, e.g., Thabtah and Peebles [12], The paper showed hopeful results in detecting ASD cases, but it did not include instances related to toddlers.

Most of the previous work used traditional machine learning approaches and hence are limited in their performance. So, it is evident that there is definitely a need to explore the possibility of applying deep learning-based models for the detection of ASD.

In this article, the performance of several machine learning models has been compared to that of the deep learning model for this purpose.

2. METHODS

The main purpose of this study is to analyze a set of ASD screening data for toddlers by using 8 classification algorithms, thus selecting the best classifier to help in improving the diagnosis process of ASD in healthcare practices.

ML algorithms can learn from the data, where choice of algorithm and features (inputs) are to be fed into algorithm which are made by subject matter experts. This section demonstrates the approaches that we used to achieve this article's goal, including the dataset involved in this study and how to preprocess it, the classification process, and performance criteria.

The stage of the system as in Figure 1.

2.1 Dataset Exploration

Sections The dataset used in this study for toddlers provided by Dr. Fadi Thabtah under the name of "Autistic Spectrum Disorder Screening Data for Toddlers". It is available at Kaggle [12], its task is binary classification and with 18 input values, with their types being categorical, continuous and binary; it has 1054 instances with unbalanced class values, 728 for 'yes', and 326 for 'no'. The features are described in Table 1.

Table 1 List of Features in The Dataset

Attribute Id	Attribute Description
1	Case number
2-11	Q-Chat-10
12	Age
13	Score by Q-Chat-10
14	Sex
15	Ethnicity
16	Born with jaundice
17	Is there a family history of autism?
18	Who is doing the test?
19	Class Variable

Description details of variables mapping to the Q-Chat-10 screening methods which has 2-11 attributes are in Table 2.

Table 2 Q-Chat-10 Of Toddlers

Variable in Dataset	Corresponding Q-chat-10-Toddlers Features
A1	When you call a child by his/her name, does he look at you?
A2	Does the child make eye contact with you?
A3	When a child wants something, does he point to it?
A4	Does your child point at an interesting sight?
A5	Does the child act like he/she talking on the phone or taking care of dolls?
A6	Does the child follow you when you move?
A7	Does the child sympathize with you when you are sad?
A8	Can you understand the child's first words?
A9	Does the child use simple mime?
A10	Does your child look at nothing with no visible purpose?

2.2 Data-preprocessing

Using raw data can lead to the difficulty in the process of training the algorithm, thus, negatively and misleadingly affecting the results. Therefore, pre-processing the data is an important step before starting to apply ML algorithms. As this is a vital part of the process, it can take a great deal of time before starting the training process.

We often begin with pre-processing raw data in several stages according to the algorithm requirements, as a result, we apply the following steps:

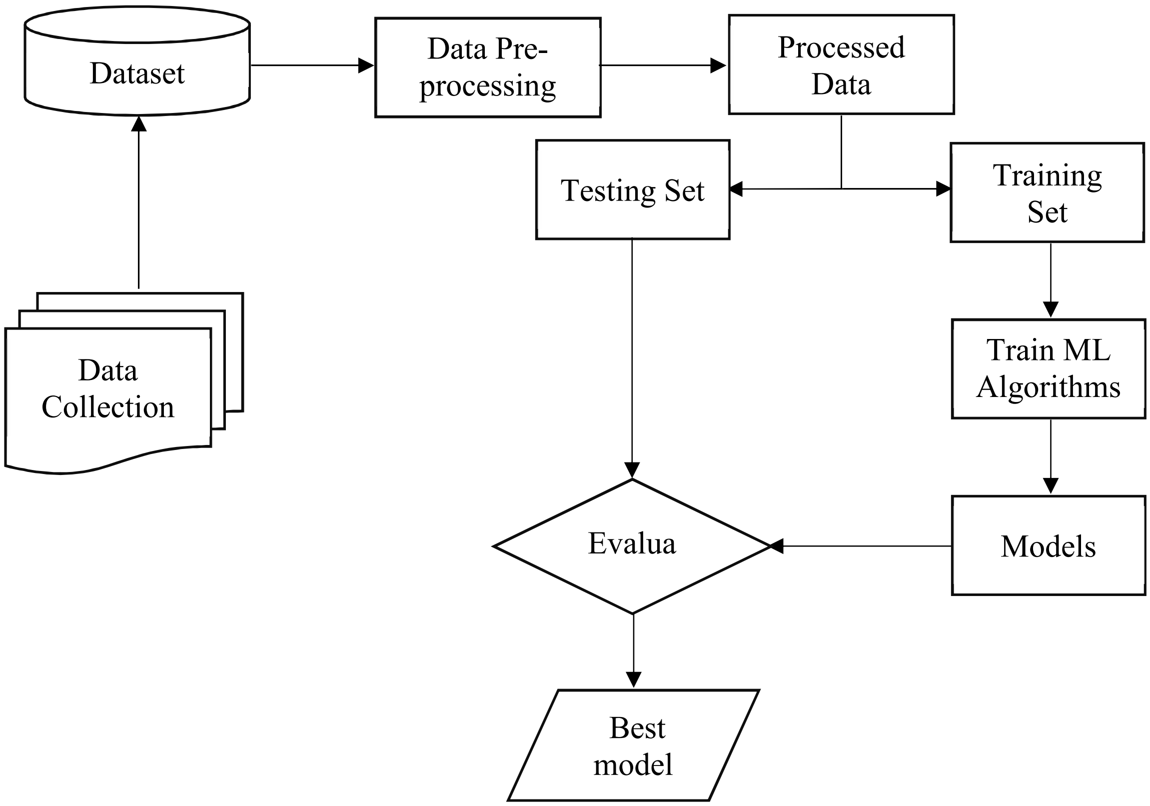

Figure 1. Steps of ASD Detection Solution

- Deleting case number column.
- Handling missing data: typical methods which includes imputation and deletion, according to our raw data, we don't have any missing values.
- Applying one-hot encoding for categorical type of values to transform them to binary values of 0 or 1.
- Normalization: by applying min-max scaler (for age and result of test app (Qchat-10-Score)) that transformed the data to binary values.
- Data splitting: data is a valuable asset and we want to make use of every bit of it; so, we will use cross-validation technique because it is maximizing the availability of training data by dividing it into two parts, one for training and the other for testing for k time. Each k section will be trained with one of the sections for testing at any one time, so it will train the model for the same algorithm for k sections. So, in the end, the result will be the mean of the k results and the model will be more accurate, and it will recognize unseen data and avoid overfitting [13].

2.3 Classification Algorithms

After the pre-processing the algorithms will be able to use the data in the right way to make models, for optimal performance of obtaining better results. Some of the most important learning algorithms are:

Logistic Regression (LR):
It is used in classification even though the word regression is present in its name, and it can also be used for binary or multiple classification. It calculates the probability of the output belonging to a particular class. If the probability is greater than 50%, it is considered that it belongs to that class. So, if we face a binary classification problem, we say that when the probability is above or equal 50 percent, we consider it to belong to class '1', otherwise it belongs to class '0' [14].

Linear Discriminant Analysis (LDA):
It is used for pattern classification and is also used to reduce dimensions [14] [17]. It combines the features in a way that maximize the separation between two or more groups [17].
K-Nearest Neighbors (KNN):
This algorithm is one of the simplest algorithms due to the ease of its principle. The k symbol indicates the number of neighbors surrounding the test point.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

So, building the model consists only of storing the training dataset.

To make a prediction for a new data point, if k=1, the algorithm finds the nearest neighbor to the testing point and considers its class as belonging to the same class of neighbor; but when the number k is greater than 1, the algorithm calculates the number of neighbors that belong to each class, and the class that is more neighbors to, this will be considered as belonging to their testing point class [15].

Classification and Regression Tree (CART):
It can be used not only for classification but also regression. Creating a CART Decision Tree is a process of recursively building a binary decision tree. The algorithm creates a decision tree based on the training data set, dividing it into two parts and dividing the branches into two parts as well, and so on until all restrictions are checked. For example, it reaches the largest depth that was specified before. Or the algorithm can create a decision tree without restrictions and the generated decision tree must be as large as possible, then prunes it with the validation dataset using the minimum loss function and selects the optimal subtree [14].

Naive Bayes (NB):
It is considered one of the fast algorithms and it can be used in instantaneous classification, which is a simple probabilistic classifier with strong independent assumptions between the features [15]. There are three kinds of naive Bayes classifiers:

1. Gaussian NB:
    - applied to continuous data.
    - mostly used on very high-dimensional data.
2. Bernoulli NB:
    - applied to binary data.
    - mostly used in text data classification.
3. Multinomial NB:
    - applied to count data.
    - mostly used in text data classification.

Naive Bayes models are often used on very large datasets [15].

Support Vector Machine (SVM):
The goal is to find the best separator level to classify the data. When we start training the model, we have more than one separator level and the algorithm has to find the best one. The main criterion for determining the level is the so-called margin, which is the distance between the separator level and the data. So, the best separator level is the one that divides the data correctly and has a large margin between it and the data points, and thus it can predict unseen data [14].

Random Forest (RF):
It works by training many decision trees on random subsets of the features, so instead of relying on a single decision tree, it takes the prediction from each tree, then averages out their predictions, so the results are at higher accuracy and avoids overfitting [14].

Multilayer Perceptron (MLP):
The most widely used type of neural networks, and it is generally used for classification. There is one input layer and one output layer for making prediction, and between these two layers, there are several numbers of hidden layers depending on the complexity of the problem. These hidden layers are the main computational power of the multilayer perceptron algorithm [16]. Therefore, every layer is fully connected to the next layer except the output layer, which includes a bias neuron. For a binary classification problem, it has a single output neuron using the logistic activation function, and the output will be a probability of the positive class with a number between 0 and 1, while the probability of the negative class is one minus that number [14].

2.4 Performance Criteria

To identify how well a model has performed can be measured by Accuracy, Precision, Recall (Sensitivity) and F1 Score.

Accuracy:
Determines the number of times the classifier's answer was correct [15].

$$\text{Accuracy} = TP + TN / TP + FP + FN + TN \quad (1)$$

TP is the number of true positives. TN is the number of true negatives. FP the number of false positives. FN the number of false negatives.

Precision:
Indicates the adjacency of two or more measurements to each other. It determines the percentage of how accurate the classifier is in its correct predictions. It is used as a performance metric when the goal is to limit the number of false positives [15]. So, precision does not depend on accuracy.

$$\text{Precision} = TP / TP + FP \quad (2)$$

Recall:

Journal of Theoretical and Applied Information Technology
15th November 2022. Vol.100. No 21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

It calculates the data that we expect to be correct from among all the data that we expected to be correct. This is used as a performance metric when we need to identify all positive samples; when it is important to avoid false negatives [15].

Recall = TP / TP + FN (3)

F1 Score:
It is the harmonic mean of precision and recall [15].

F1 = 2*(Precision*Recall/Precision+Recall) (4)

4. RESULTS

Figure 2 shows the spread of the accuracy scores across each cross-validation for each algorithm using stratified k-fold. Table 3 shows comparison of algorithms by training runtime while Table 4 shows a comparison of algorithms by the mean of scores, and Table 5 shows a comparison of algorithms by the standard deviation of scores.

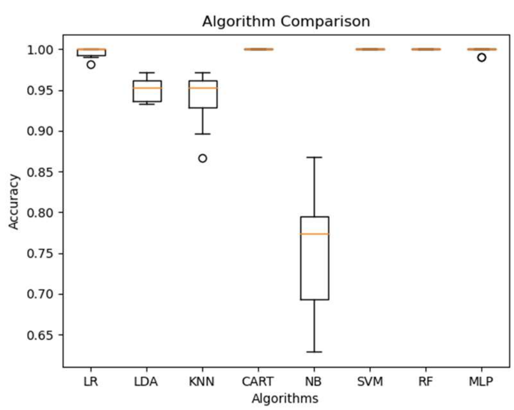

Figure 2 Comparing ML Algorithms

Table 3 Comparing the Algorithms by The Training Runtime (sec)

Algorithm	Training runtime (sec)
LR	0.026978
LDA	0.051857
KNN	0.053761
CART	0.023245
NB	0.025729
SVM	0.208184
MLP	6.229822
RF	1.123399

Table 4 Comparing Algorithms by The Mean Of Accuracy (MA), Recall (MR), Precision (MP) and F1 (MF1)

	MA	MR	MP	MF1
LR	0.996208	0.997260	0.997279	0.997260
LDA	0.951617	0.949763	0.980497	0.964396
KNN	0.940252	0.939593	0.972983	0.955622
CART	1.000000	1.000000	1.000000	1.000000
NB	0.751276	0.646861	0.987369	0.778855
SVM	1.000000	1.000000	1.000000	1.000000
MLP	0.998104	1.000000	0.997279	0.998630
RF	1.000000	1.000000	1.000000	1.000000

Table 5 Comparing Algorithms by The Standard Deviation of Accuracy (STDA), Recall (STDR), Precision (STDP) and F1 (STDF1).

	STDA	STDR	STDP	STDF1
LR	0.006274	0.005479	0.005443	0.004543
LDA	0.014320	0.017420	0.016353	0.010596
KNN	0.033143	0.040229	0.016713	0.025274
CART	0.000000	0.000000	0.000000	0.000000
NB	0.069369	0.090029	0.023540	0.071937
SVM	0.014199	0.009110	0.015227	0.010227
MLP	0.003792	0.000000	0.005443	0.002740
RF	0.000000	0.000000	0.000000	0.000000

We suggest deleting the column under the name 'Qchat-10-Score', and monitoring the algorithms with the new dataset.

The result of accuracy scores for each algorithm in Figure 3 shows the spread of the accuracy scores across each cross-validation for each algorithm after deleting the 'Qchat-10-Score' column using stratified k-fold, while Table 6 shows the new results for comparison of algorithms by execution time. In addition, Table 7 shows the new results for comparison of algorithms by the mean of scores and Table 8 shows a comparison of algorithms with the standard deviation of scores.

Journal of Theoretical and Applied Information Technology
15th November 2022. Vol.100. No 21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

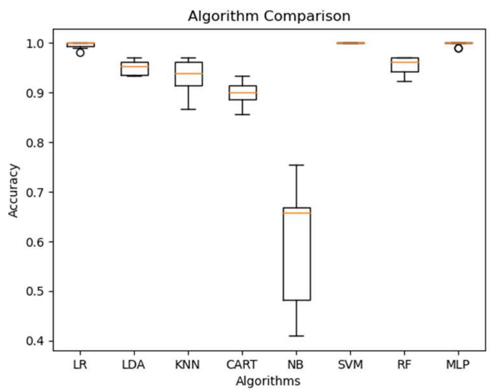

Figure 3 Comparing ML Algorithms After deleting 'Qchat-10-Score' Column

Table 6 Comparison Algorithms by The Execution Time After Deleting 'Qchat-10-Score' Column

Algorithm	Training runtime (sec)
LR	0.027442
LDA	0.050258
KNN	0.059697
CART	0.025166
NB	0.017176
SVM	0.227151
MLP	5.278182
RF	1.222410

Table 7 The New Result for Comparison of Algorithms by The Mean of Accuracy (MA), Recall (MR), Precision (MP) And F1 (MF1) After Deleting 'Qchat-10-Score' Column

	MA	MR	MP	MF1
LR	0.996208	0.997260	0.997279	0.997260
LDA	0.951617	0.949763	0.980497	0.964396
KNN	0.931716	0.929966	0.969983	0.949121
CART	0.895580	0.934018	0.929354	0.929395
NB	0.601249	0.429718	0.968333	0.579377
SVM	1.000000	1.000000	1.000000	1.000000
MLP	0.998104	1.000000	0.997279	0.998630
RF	0.957314	0.980746	0.957036	0.969389

Table 8 A comparison Of Algorithms by The Standard Deviation of Accuracy (STDA), Recall (STDR), Precision (STDP) And F1 (STDF1) After Deleting 'Qchat-10-Score' Column

	STDA	STDR	STDP	STDF1
LR	0.006274	0.005479	0.005443	0.004543
LDA	0.014320	0.017420	0.016353	0.010596
KNN	0.034119	0.041749	0.017998	0.026280
CART	0.031110	0.022950	0.011345	0.014828
NB	0.120581	0.163674	0.057861	0.177473
SVM	0.000000	0.000000	0.000000	0.000000
MLP	0.003792	0.000000	0.005443	0.002740
RF	0.014866	0.012647	0.017049	0.006825

By comparing our results in this paper, we found that the performance of classifiers is affected by the type of dataset, as well as, the number of features involved in the experiment.

From the performance of the models, it would suggest that SVM, LR and MLP are perhaps worthy of further study on this problem in terms of only scores (Accuracy, Recall, Precision and F1), and only LR in terms of both scores and training runtime.

5. CONCLUSION

At the time of using technology to detect most diseases, it has become extremely important for it to be used to diagnose autism at an early age. This study shows that autism can be detected for toddlers using machine learning and deep learning techniques in a large percentage of accuracy. So, we can use one of these classifiers (SVM, LR, or MLP) to build a new mobile app to be used by parents to know if their toddler may have autism or not. This in turn leads to the significance of giving attention to affected children at an early stage, through providing them with speech, behavioral, and occupational therapy. Thus, these programs can help reduce symptoms, and help the child be more independent and integrated in the wider society.

6. FUTURE WORK

We need to collect more real data (via social media, centers of autism, etc.) and compare its result with the results of the data used in this article.

REFERENCES:

[1] Samata R. Sharma, Xenia Gonda and Frank I. Tarazi, "Autism Spectrum Disorder: Classification, diagnosis and therapy",

Pharmacology & Therapeutics, Vol. 190, October 2018, pp. 91-104 https://doi.org/10.1016/j.pharmthera.2018.05.007

[2] Fadi Thabtah, Neda Abdelhamid and David Peebles, "A machine learning autism classification based on logistic regression analysis", National Library of Medicine, Vol. 7, No. 1, 2019, pp. 12. https://doi.org/10.1007/s13755-019-0073-5

[3] Data & Statistics on Autism Spectrum Disorder. Centers for Disease Control and Prevention. https://www.cdc.gov/ncbddd/autism/data.html (02/03/2022).

[4] Sahr Yazdani, Angela Capuano, Mohammad Ghaziuddin and Costanza Colombi, "Exclusion criteria used in early behavioral intervention studies for young children with autism spectrum disorder", Brain Sciences, Vol. 10, No. 2, 2020, pp. 99. https://doi.org/10.3390/brainsci10020099

[5] Kazi Shahrukh Omar, Prodipta Mondal, Nabila Shahnaz Khan, Md. Rezaul Karim Rizvi and Md Nazrul Islam, "A Machine Learning Approach to Predict Autism Spectrum Disorder", International Conference on Electrical, Computer and Communication Engineering (ECCE), 2019, pp. 1-6. https://doi.org/10.1109/ECACE.2019.8679454

[6] C. Lord, M. Rutter and A. Le Couteur, "Autism Diagnostic Interview-Revised: A revised version of a diagnostic inter- view for caregivers of individuals with possible pervasive developmental disorders", Journal of Autism and Developmental Disorders. Vol. 24, No. 5, pp. 659-85. https://doi.org/10.1007/BF02172145

[7] Halim Abbas, Ford Garberson, Stuart Liu-Mayo, Eric Glover and Dennis P. Wall, "Multi-modular AI Approach to Streamline Autism Diagnosis in Young Children", Scientific Reports, Vol. 10, No. 5014, 2020. https://doi.org/10.1038/s41598-020-61213-w

[8] Shigeki Kurasawa, Kiyomi Tateyama, Ryoichiro Iwanaga, Taro Ohtoshi, Ken Nakatani and Katsushi Yokoi, "The Age at Diagnosis of Autism Spectrum Disorder in Children in Japan", International Journal of Pediatrics, vol. 2018, No. 5374725, 2018. https://doi.org/10.1155/2018/5374725

[9] Jennifer Harrison Elder, Consuelo Maun Kreider, Susan N Brasher and Margaret Ansell, "Clinical impact of early diagnosis of autism on the prognosis and parent–child relationships", Psychol Res Behav Manag, Vol. 10, 2017, pp. 283-292. https://doi.org/10.2147/PRBM.S117499

[10] Md. Rafiqul Islam, Muhammad Ashad Kabir, Ashir Ahmed, Abu Raihan M. Kamal, Hua Wang, and Anwaar Ulhaq, "Depression detection from social network data using machine learning techniques", Health information science and systems, Vol. 6, No. 1, 2018, pp. 8. https://doi.org/10.1007/s13755-018-0046-0

[11] Da-Yea Song, So Yoon Kim, Guiyoung Bong, Jong Myeong Kim, and Hee Jeong Yoo, "The Use of Artificial Intelligence in Screening and Diagnosis of Autism Spectrum Disorder", Soa Chongsonyon Chongsin Uihak, Vol. 30, No. 4, 2019, pp. 145-152. https://doi.org/ 10.5765/jkacap.190027

[12] Fadi Thabtah, David Peebles, "A new machine learning model based on induction of rules for autism detection", Health Informatics J, Vol. 26, No. 1, 2020, pp. 264-286. https://doi.org/10.1177/1460458218824711

[13] O. Theobald, Machine Learning for Absolute Beginners, 2nd ed., Scatterplot Press, 2017

[14] A. Géron, Hands-on Machine Learning with Scikit-Learn Keras and TensorFlow, 2nd ed., O'Reilly, Canada, 2019.

[15] A. C. Müller, S. Guido, Introduction to Machine Learning with Python, 1st ed., O'Reilly, United States of America, 2017.

[16] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2017.

[17] A. Tharwat, T. Gaber, A. Ibrahim, A. E. Hassanien, "Linear discriminant analysis: A detailed tutorial", IOS Press, Vol. 30, No. 2, 2017, pp. 169-190. https://doi.org/10.3233/AIC-170729