



Autism Spectrum Diagnosis using Adaptive Learning Algorithm for Multiple MLP Classifier

Qiang Zhang^{1*} and Fatemeh Safara²

¹ School of Transportation, Southeast University, Nanjing, 211189, China

*Corresponding Author's Email: mrzhang036@163.com, w.zhang82@yahoo.com

² Department of Computer Engineering, Islamshahr Branch, Islamic Azad University, Islamshahr, Iran

Email : fsafara@yahoo.com

Abstract

A medical condition that causes disability and early neurological and cognitive condition is autism spectrum disorder (ASD). Gene expression and environment have an impact on this medical condition. Development of diagnostic instruments and skills improved the autism recognition and increased the society awareness about it. To cope with this disorder collaboration between families, service providers, and autistic individuals is a necessity. Early diagnosis of ASD could help in lessening stress, increase adaptation, and support welfare in healthcare systems. Therefore, a large body of research is attempting to provide an intelligent medical diagnostic system to identify and diagnose ASD in early stages using machine learning methods. In this paper, several multilayer perceptron neural network is proposed for ASD detection in healthcare systems. The learning rate is adaptively tuned to achieve the best results. The results show that the approach proposed in this study achieved 99.6% accuracy, which indicates the superiority of the proposed method in identifying and detecting autism disorder in comparison with similar previous methods.

Keywords: Autism spectrum disorder, Intelligent medical diagnostic system, multilayer perceptron neural network, adaptive learning rate, machine learning

1.Introduction

One of the serious mental disorders in the developmental group is autism spectrum disorders (ASD) that appear in early childhood (usually before age 3) [1, 2]. It is referred to as a spectrum disorder because there is an extensive variation in the type and severity of the symptoms [. Although symptoms and severity vary from

person to person, all types of autism affect the ability of communication. A disturbance in the nervous system that causes the brain to fail can contribute to the development of this disease [3, 4]. Moreover, the psychological conditions of parents, in particular mothers, are affected by the emergence of their children's behavioral problems. The prevalence of autism in high-income countries is 1.5% [2], which indicates the production of large amount of data every day. Such a big data needs to be processed with powerful and precise techniques in e-healthcare systems [3, 4]. Therefore, there is an urgent need to develop an intelligent medical diagnostic system to identify and diagnose autism disorder in early stages. Providing a computer aided diagnosis (CAD) system with the capability of automatic autism detection and prediction is an important challenge in big data of medical engineering [5, 6].

To date, several studies have been conducted to provide intelligent methods for the diagnosis and prediction of autism spectrum disorder. In [7], the classification of spectral autism disorders was performed using discriminant analysis and support vector machine (SVM) with polynomial kernel. Fearful, happy and neutral subjects are classified. Data is collected from the electro-encephalogram signal called EEG. The results of 94.7% accuracy is obtained with their proposed method, as well as the sensitivity of 85.7% and specificity of 100%. This research conducted on continuous wavelet transform of EEG signal [8], which are entered as input, and then normalizing and filtering is performed as preprocessing. Then the signal timing is measured and the phase difference of the signal is measured. The average of the frequencies in a selected band of the signal is measured and then the clustering operation is performed using the K-means algorithm. The signal is synchronized, and then the synchronization index is calculated to measure the brain's connections.

In [9], the logical regression approach of the Byzantine cycle that took Laplace cycle to prevent overlapping and the production of the SPARS forecast model for textual data was taken. In [10], the use of neural networks in the classification of autism has been studied. The authors of the article [11] have developed a new method which is based on the neuro-fuzzy network approach for classifying and predicting autism. The reason for the use of the fuzzy approach is also due to the uncertainty in the diagnosis and prediction of autism, which the research presented in this paper also emphasizes this issue.

In [12], an EEG signal is used to predict autism, where the author explores a series of methods for predicting autism. The methods used include K-CM, NaïveBayes, K-nearest Neighbour, FF_Sn, Random forest, Sequential minimal optimization (SMO), and logistic regression. The best results obtained by SMO. In [13], a precise diagnostic method is proposed based on a brain MRI image that operates based on the neural network of the radial base function. The accuracy of the proposed method is up to 95.34% based on prediction based on age and 96.92% based on prediction of gender disorders and 70.37% for prediction of male sexual disorders. In [14], searching for a minimum set of behaviors for the diagnosis of autism has been developed using a feature extraction approach using machine learning methods. The proposed method has a precision of 97.66% in the diagnosis of autism patients. The reduction of computational complexity of most of the previous methods is one of the goals of this research.

Overall, it is necessary to improve the accuracy of ASD diagnosis. The impact of convergence behaviour on neural network performance has been confirmed in previous researches [15-19]. In this study, we employ a number multilayer neural network (MLP) to classify autism dataset into two categories including: normal and autism. In the proposed method, we exploit back propagation error to improve the network generalization behaviour. An adaptive learning rate algorithm is used to improve the convergence rate of the back-propagation as well. The contributions of this paper can be summarized as follows:

- To diagnose two types of autism medical condition using multiple MLP.
- To improve the accuracy of classifying features selected from the row dataset.
- To use adaptive learning to improve the performance of ANNs.
- To achieve better results than previous researches using the proposed method.

In this paper, a multiple MLP is proposed to increase the accuracy of ASD detection. The organization of this research is structured as follow. Section 2 illustrates the proposed classification method and evaluation criteria. Section 3 presents the simulation and experimental results. Section 4 concludes the paper and presents future work.

2. Proposed Method

In this section, first the dataset employed in this paper for the experiment and comparing the results with that of previous studies is introduced. Then, multiple MLP system is explained. At the end of the section, evaluation criteria are presented.

2.1 Dataset

For autism spectrum diagnosis, UCI dataset is used. Table 1 shows existing features of dataset related to autism disorder that contained 14 features to evaluate classification of ASD instances.

Table 1. Features and their descriptions

#	Feature	Domain	Comment
1	Age	Number	years
2	Gender	String	Male or Female
3	Ethnicity	String	Set of common ethnicities
4	Born with jaundice	Boolean (yes or no)	Location of person that born with jaundice
5	Family member with PDD	Boolean (yes or no)	Whether any immediate family member
6	Who is completing the test	String	Parent, self, caregiver, medical staff, clinician ,etc.
7	Country of residence	String	Set of countries
8	Used the screening app before	Boolean (yes or no)	Whether the user has used a screening app
9	Screening Method Type	Number	The type of screening methods chosen based on age category (0=toddler, 1=child, 2= adolescent, 3= adult)
10	genetics	String	A set of genetic factors

11	pesticides	String	A set of pesticides factors
12	drugs	String	A set of drugs factors
13	parental age	Number	Age in years
14	brain development	Binary (0, 1)	Lazy = 0 / Normal =1

2.2 Feature scaling

Feature scaling is an important data preparation step to rescale the value of features from dynamic range into a specific range. Standardization is a well-known method for scaling the value of features with different scales. In artificial neural network, standardizing the input vectors helps to have a fast and accurate convergence. It standardizes the value of features into a small specific range which can improve the accuracy of classification. After the feature selection phase of this study, input variables have been standardized to map the values of each row in the data to have a mean of 0 and a standard deviation of 1.

2.3 Multiple MLP design

MLP is a successful and therefore popular machine learning algorithm. As reviewed in literature review section, several ANN used in previous researches for Autism disease diagnosis.

We examined an MLP with different configuration to improve the accuracy of Autism disease detection which were reported before. However, desired accuracy was not achieved with a single MLP. Therefore, we tried to investigate the power of multiple MLP (MMLP) for the same goal. After training a set of multiple networks individually, each network is tested on the evaluation set with two measures: MSE of each individual MLP and average MSE for different combinations of MLPs. As will be shown later in this section, the accuracy is increased and MSE of validation set is decreased gradually by training more networks. However, the improvements of the networks will not improve after adding a certain number of networks. The best accuracy was obtained by 3 MLPs.

One of the important issues with MLP is its convergence rate. The impact of convergence behaviour on neural network performance has been confirmed in previous researches [2-6]. In this paper, to employ an adaptive learning rate has been proposed to increase the generalization and convergence rate of the proposed multiple MLP in medical environment.

Therefore, an adaptive learning rate is used to overcome this problem; the value of the learning rate is changed in each epoch during the learning process. All the experiment steps are performed on the Autism disease dataset taken from the UCI machine learning repository, which is a widely used benchmark dataset.

2.4 Adaptive learning rate

The convergence rate of the back-propagation is highly dependent on the selection of the learning rate which is a control parameter to controls the step size for each weight update. The value of the learning rate should be a value between 0 and 1. A large value of the learning rate may skip the optimal solution and the convergence may

never be achieved. On the other hand, a small learning rate will increase the total time to converge to the optimal value or trap in local minima [20, 21]. Therefore, selecting the optimal value of the learning rate is a major challenge regarding the convergence rate. To avoid these problems, an adaptive learning algorithm has been used, which adapts the value of the learning rate during the training process.

The most common learning rate include time-based, exponential, and step-based learning rate. Time-based learning rate modifies the learning rate regarding to the learning rate of the earlier time iteration. Exponential learning rate is similar to step-based learning rate, however, as a substitute of steps a decreasing exponential function is included. Step-based learning rate sets a new learning rate based on predetermined steps. In this paper we employed step-based learning rate. The decreasing formula is defined as:

$$\eta_n = \eta_0 d^{\text{floor}(\frac{1+n}{r})} \quad (1)$$

where η_n is the learning rate at iteration n , and η_0 is the initial learning rate, d is how much the learning rate should change at each decrease (for example 0.5 corresponds to a splitting) and r corresponds to the decrease rate, or how often the rate should be decreased (10 corresponds to a decrease every 10 iterations). The *floor* function here decreases the value of its input to 0 for all values smaller than 1.

At each epoch, if the performance of the network decreases, the learning rate will increase by a constant parameter and vice versa. In this procedure, the learning rate is increased whenever a large value could result in stable learning. When the learning rate is too large to guarantee a decrease in error, it is decreased until stable learning resumes.

2.5 Feature Selection

Feature selection is a frequently used techniques in classification applications. The quality of classification result is highly dependent on the selected features. In feature selection process, noisy, redundant, and irrelevant features would be removed while informative features would be preserved. Three major feature selection methods are filter method, wrapper method, and embedded method.

Filter method statistically investigates the inherent characteristics of dataset and calculates the score for each feature independently from any classifier result. Wrapper method scores features based their usefulness to improve classification performance. In embedded method, the search process for the best feature subset is indirectly integrated into classifier construction such as the decision tree classifier. In this paper, we tries several feature selection methods, however the results were not satisfiable. Therefore, we used all the 14 features for classification.

2.6. Evaluation parameters

Accuracy, sensitivity, specificity and Root mean square error (RMSE) were calculated. Accuracy is obtained from the following equation:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (4)$$

In these relationships, TP, FP, TN and FN are true positives rate, false positives rate, true negatives rate and false negatives rate, respectively.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (5)$$

In equation above values are representing as follows: y_i is the type of traffic class number, which could be normal or abnormal for i th traffic, \hat{y}_i is the estimated type of traffic class number for i th traffic, N is the number of testing samples in the proposed method.

3. Simulation and Result

The autism dataset is from UCI with 704 instances. The multiple MLP is built with neural network toolkit in the MATLAB environment. The multiple MLP system has 14 entries presented in Table 1. All MLPs have 2 layers and in each layer there were 8 neurons. There are two outcomes that include has autism, don't have autism. For training 80% of the data was used and the rest 20% was used for testing.

At the beginning of the experiments, we tried just one MLP. The good results were achieved. Back propagation is used as a learning algorithm. Figure 1 illustrates the confusion matrix of standard MLP with the back propagation error algorithm.

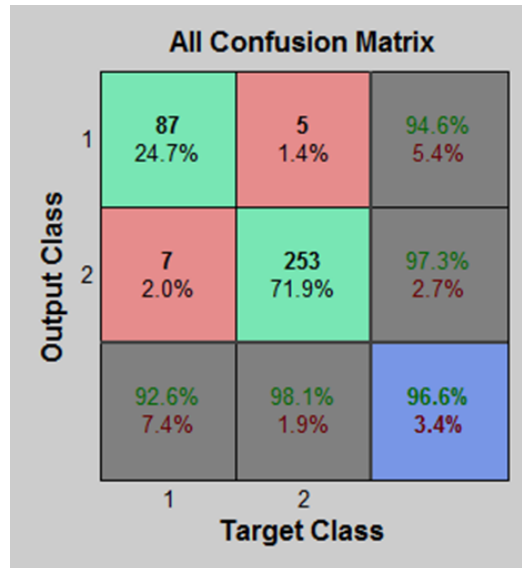


Figure 1. The Confusion Matrix for the Standard BP

Accuracy of 96.6%, specificity of 98.1%, and sensitivity of 92.6% is achieved. MSE of 0.0435 was obtained in 6 seconds. ROC of the results gained with standard BP is shown in Figure 2.

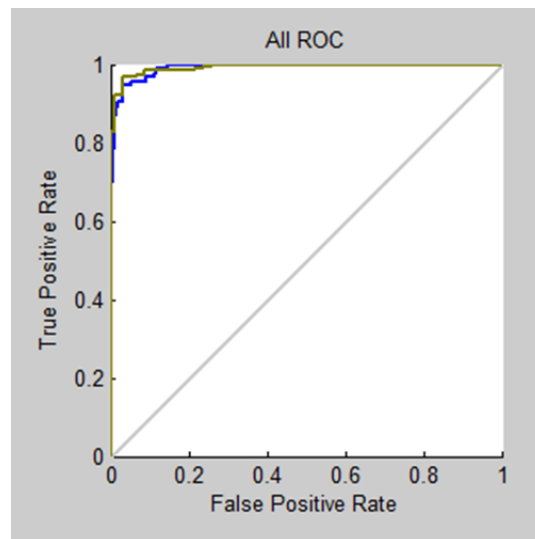


Figure 2. The ROC curve for the Standard BP

Then a number of MLPs is used to gather. We examined a different number of MLPs where the best results achieved with three MLPs. Moreover, learning rate was not a constant. It was adjusted adaptively. Figure 3 illustrates the confusion matrix of standard MLP with the back propagation error algorithm.

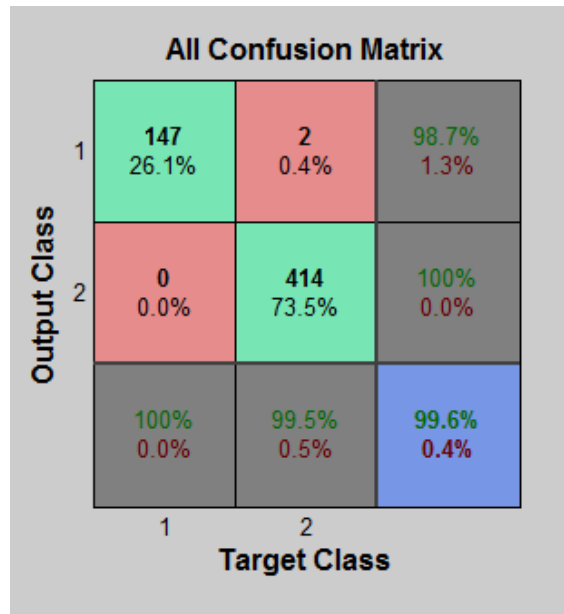


Figure 3. The Confusion Matrix for the Standard BP

With the improved MLP, which is constructed from three MLPs with adaptive learning, accuracy of 99.6%, specificity of 99.5%, and sensitivity of 100% were obtained. MSE of 0.0141 was obtained in 3 seconds Figure 4 illustrates the confusion matrix of standard MLP with the back propagation error algorithm.

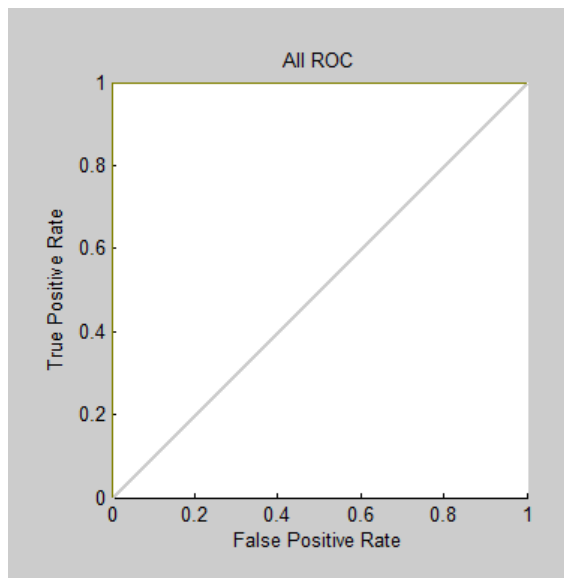


Figure 4. The ROC curve for the Improved BP and adaptive learning

In order to show the impact of the number of MLPs we presented the results of one, two and three MLPs separately, in Figure 5.

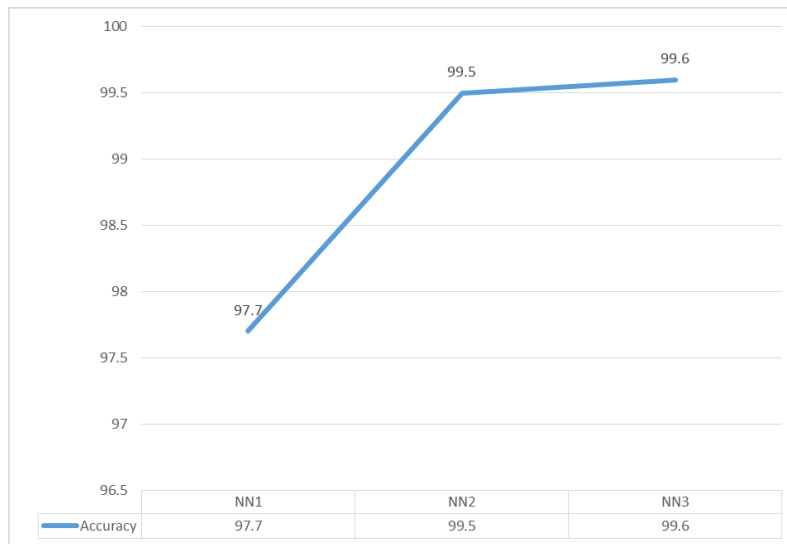


Figure 5. The comparison of the different number of networks regarding the Accuracy

4. DISCUSSION

MLP is shown to be a successful classifier in machine learning application developed for medical systems. It was motivated to use the MLP for diagnosis ASD in this paper. In previous studies single MLP was used and acceptable results were achieved. In order to improve the results reported before, we tried using multiple MLPs. In addition, learning rate was constant on most of the previous studies. In this paper, we used adaptive learning rate with the back propagation learning algorithm. Evaluation of the proposed method in comparison to single MLP with standard back propagation algorithm is provided in Figure 6.

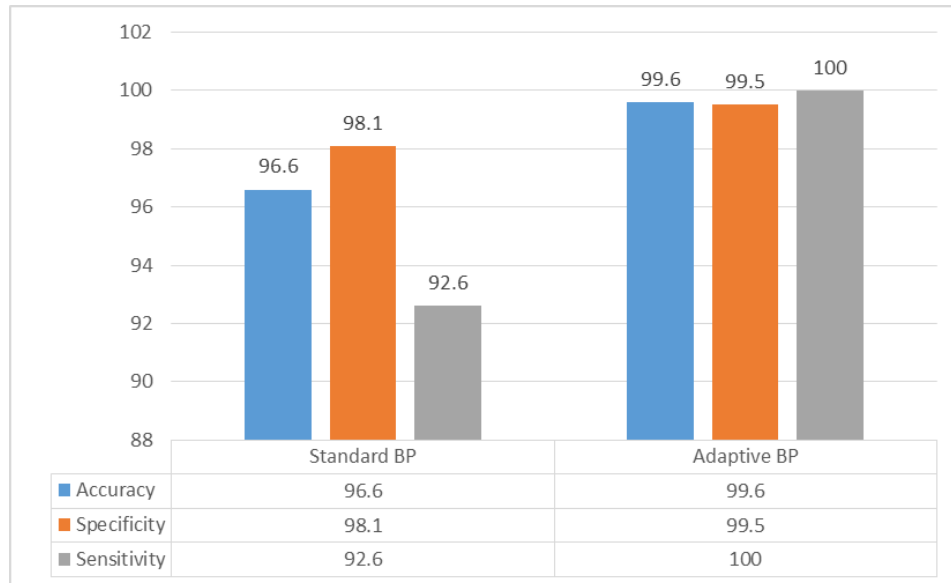


Figure 6. The comparison of the Standard BP and the Improved BP regarding the accuracy, specificity and sensitivity

As the above result illustrates, the proposed multiple MLP with adaptive learning rate outperformed single MLP with standard back propagation.

5. CONCLUSION

In recent years, one of the psychological problems in parents that affects the child has also been studied, which is commonly referred to as autism. This disease, which leads to problems with the behavior of children, needs to be treated in the best possible time before any development and development, so as not to exacerbate the severity of the child's nervous system. One of the principles in medical systems with the help of computers is in the medical field. This field is trying to provide intelligent diagnostic methods in order to diagnose and predict diseases in the shortest possible time based on previous data and current data. Therefore, the present study seeks to provide a new method based on data mining and training-based principles for the detection and prediction of autism. It is essential to use a dataset that has a number of key features of autism disorders. In the world of intelligence, the use of multiple MLPs is proposed in this paper. The result of 99.6% accuracy of the proposed method has a superiority to its previous methods in the same conditions. In the future work, we will apply deep learning and some hybrid feature selection methodologies with meta-heuristic algorithms to improve the accuracy of the ASD diagnosis.

References

- [1] S. Oueida, Y. Kotb, M. Aloqaily, Y. Jararweh, and T. Baker, "An edge computing based smart healthcare framework for resource management," *Sensors*, vol. 18, no. 12, p. 4307, 2018.

- [2] M.-C. Lai, E. Anagnostou, M. Wiznitzer, C. Allison, and S. Baron-Cohen, "Evidence-based support for autistic people across the lifespan: maximising potential, minimising barriers, and optimising the person–environment fit," *The Lancet Neurology*, 2020.
- [3] B. Deebak, F. Al-Turjman, M. Aloqaily, and O. Alfandi, "An Authentic-Based Privacy Preservation Protocol for Smart e-Healthcare Systems in IoT," *IEEE Access*, vol. 7, pp. 135632-135649, 2019.
- [4] S. Oueida, M. Aloqaily, and S. Ionescu, "A smart healthcare reward model for resource allocation in smart city," *Multimedia Tools and Applications*, vol. 78, no. 17, pp. 24573-24594, 2019.
- [5] G. Todorov, N. Nikolov, Y. Sofronov, N. Gabrovski, M. Laleva, and T. Gavrilov, "Computer Aided Design of Customized Implants Based on CT-Scan Data and Virtual Prototypes," in *International Conference on Future Access Enablers of Ubiquitous and Intelligent Infrastructures*, 2019: Springer, pp. 339-346.
- [6] M. Al-Khafajiy *et al.*, "Intelligent Control and Security of Fog Resources in Healthcare Systems via a Cognitive Fog Model," *ACM Transactions on Internet Technology*.
- [7] K. Sarmukadam, V. Bitsika, C. F. Sharpley, M. M. McMillan, and L. L. Agnew, "Comparing different EEG connectivity methods in young males with ASD," *Behavioural Brain Research*, vol. 383, p. 112482, 2020.
- [8] M. H. Bhatti *et al.*, "Soft computing-based EEG classification by optimal feature selection and neural networks," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 10, pp. 5747-5754, 2019.
- [9] A. Genkin, D. D. Lewis, and D. Madigan, "Large-scale Bayesian logistic regression for text categorization," *technometrics*, vol. 49, no. 3, pp. 291-304, 2007.
- [10] R. Anirudh and J. J. Thiagarajan, "Bootstrapping graph convolutional neural networks for autism spectrum disorder classification," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019: IEEE, pp. 3197-3201.
- [11] R. Ahsan *et al.*, "Prediction of Autism Severity Level in Bangladesh Using Fuzzy Logic: FIS and ANFIS," in *International Conference on Multimedia and Network Information System*, 2018: Springer, pp. 201-210.
- [12] E. Grossi, C. Olivieri, and M. Buscema, "Diagnosis of autism through EEG processed by advanced computational algorithms: a pilot study," *Computer methods and programs in biomedicine*, vol. 142, pp. 73-79, 2017.
- [13] V. Subbaraju, S. Sundaram, S. Narasimhan, and M. B. Suresh, "Accurate detection of autism spectrum disorder from structural MRI using extended metacognitive radial basis function network," *Expert Systems with Applications*, vol. 42, no. 22, pp. 8775-8790, 2015.
- [14] J. Kosmicki, V. Sochat, M. Duda, and D. Wall, "Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning," *Translational psychiatry*, vol. 5, no. 2, pp. e514-e514, 2015.
- [15] T. Takase, S. Oyama, and M. Kurihara, "Effective neural network training with adaptive learning rate based on training loss," *Neural networks : the official journal of the International Neural Network Society*, vol. 101, pp. 68-78, May 2018, doi: 10.1016/j.neunet.2018.01.016.
- [16] J. Liang, Y. Xu, C. Bao, Y. Quan, and H. Ji, "Barzilai–Borwein-based adaptive learning rate for deep learning," *Pattern Recognition Letters*, vol. 128, pp. 197-203, 2019, doi: 10.1016/j.patrec.2019.08.029.

- [17] J. Kolbusz, P. Rozycki, O. Lysenko, and B. M. Wilamowski, "Error Back Propagation Algorithm with Adaptive Learning Rate," in *2019 International Conference on Information and Digital Technologies (IDT)*, 2019: IEEE, pp. 216-222.
- [18] W. Ahmad, A. Ahmad, C. Lu, B. A. Khoso, and L. Huang, "A novel hybrid decision support system for thyroid disease forecasting," *Soft Computing*, vol. 22, no. 16, pp. 5377-5383, 2018, doi: 10.1007/s00500-018-3045-9.
- [19] N. M. Nawi *et al.*, "The effect of pre-processing techniques and optimal parameters selection on back propagation neural networks," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 7, no. 3, pp. 770-777, 2017.
- [20] E. R. Rene, M. E. López, H. S. Park, D. Murthy, and T. Swaminathan, "ANNs for Identifying Shock Loads in Continuously Operated Biofilters: Application to Biological Waste Gas Treatment," in *Handbook of Research on Industrial Informatics and Manufacturing Intelligence: Innovations and Solutions*: IGI Global, 2012, pp. 72-103.
- [21] S. Sheel, T. Varshney, and R. Varshney, "Accelerated learning in MLP using adaptive learning rate with momentum coefficient," in *2007 International Conference on Industrial and Information Systems*, 2007: IEEE, pp. 307-3