

Classification of Autism Spectrum Disorder Data using Machine Learning Techniques

V.Jalaja Jayalakshmi, V.Geetha, R.Vivek

ABSTRACT--- *Autism is a neuro-developmental disability that affects human communication and behaviour. It is a condition that is associated with the complex disorder of the brain which can lead to significant changes in social interaction and behaviour of a human being. Machine learning techniques are being applied to autism data sets to discover useful hidden patterns and to construct predictive models for detecting its risk. This paper focuses on finding the best machine learning classifier on the UCI autism disorder data set for identifying the main factors associated with autism. The results obtained using Multilayer Perceptron, Naive Bayes Classifier and Bayesian Network were compared with J48 Decision tree algorithm. The superiority of Multilayer Perceptron over the well known classification algorithms in predicting the autism risk is established in this paper.*

Keywords - Machine Learning, Classification, Accuracy, Autism

I. INTRODUCTION

Autism Spectrum Disorder (ASD) is a lifelong neuro-developmental disorder that has issues related to social interaction, speech and communication. The symptoms though not hard to detect, the diagnosis requires trained physicians to oversee autism behavioural evaluations which are measured depending on the occurrence of various symptoms that interrupt a person's ability to talk, play and formulate communication relationships. ASD exists in many forms ranging from very mild condition to very severe disorder, depending on the rigorousness of the symptoms. It has been found that 1 in 68 children are found with ASD which has been increasing in the recent years. ASD is a significant burden for both the patients and their families.

Early discovery of Autism Spectrum Disorder will enhance the overall mental health of the individual. The use of machine learning techniques on autism data sets to discover useful hidden patterns and to construct predictive models for detecting its risk has been studied in this work. The paper focuses on finding the best machine learning classifier on the UCI autism disorder data set [11] to provide an insight into the causes of autism. Machine learning algorithms are useful for detecting complex patterns in the data. Data classification is one of the most important machine learning techniques. Classification plays an important role in organizing the data. Classification assigns items in a collection to target categories or classes and the

goal of classification is the accurate prediction of target class for each class of data. A number of classification algorithms such as decision tree (Quinlan 1993) [13], rule-based classifiers (Cohen 1995) [15], Bayesian classifiers (Langley et al 1992) [16], Support Vector Machines (SVM) (Vapnik 1995) and Artificial Neural Networks have been proposed in literature.

Machine learning methods have been widely used in a number of research studies on autism data sets [10]. The behavioural disorders in ASD in adult data have been classified using Support Vector Machines, k-nearest neighbour & Random Forest (RF) and compared using accuracy measure. The results indicate that RF has the highest classification accuracy. The authors have used weighted decision tree prediction model is proposed for Autism Risk Analysis [1]. The dataset required for this research work was gathered from multiple questionnaires, interviews carried out by experts from observations on ASD affected children. Multiple SVMs were used to classify ASD patients and this method gave an excellent classification performance of 96.15% based on all the features [2]. Naive Bayes, Support Vector Machines (SVM), J48 Decision Tree and Random Forest algorithms were applied on NSCH autism data with 95577 records of children to determine a set of conditions that prove to be predictive of the severity of ASD [3]. Kosmicki et al. experimented if it is possible to differentiate between children with and without ASD by using only a subset of features from autism diagnostic observation schedule (ADOS) [4]. They were able to achieve high values for sensitivity and specificity with using only a small feature subset by combining of ADOS and autism diagnostic interview revised (ADI-R) items [5].

This study has attempted to provide an insight into the process of applying machine learning techniques on UCI ASD adult data set using WEKA tool to identify the autism spectrum disorder. The organization of the paper is as follows: Section 2 discusses about the theoretical aspects of the machine learning algorithms are relevant to the proposed work. Experimental setup, results, analysis and comparison with J48 algorithm are discussed in Section 3. Section 4 concludes the work and discusses the scope for future work.

Revised Manuscript Received on August 14, 2019.

V.Jalaja Jayalakshmi, Assistant Professor, Dept. of Computer Applications, Kumaraguru College of Technology, Coimbatore, T.N, India. (E-mail: jalajajayalakshmi.v.mca@kct.ac.in)

V.Geetha, Professor, Dept. of Computer Applications, Kumaraguru College of Technology, Coimbatore, T.N, India. (E-mail: geetha.v.mca@kct.ac.in)

R.Vivek, Student, Dept. of Computer Applications, Kumaraguru College of Technology, Coimbatore, T.N, India.

II. MACHINE LEARNING ALGORITHMS

A. Multilayer Perceptron (MLP)

MLP is a classifier that uses back propagation for classification. It has any number of Input layers, one or more hidden layers with any number of inputs and output layers. There are connections between the input layers the hidden layers, and between the hidden layers and the output layers, with each connection having a weight associated with it. A MLP (or Artificial Neural Network - ANN) with a single hidden layer is represented graphically in Fig 1.

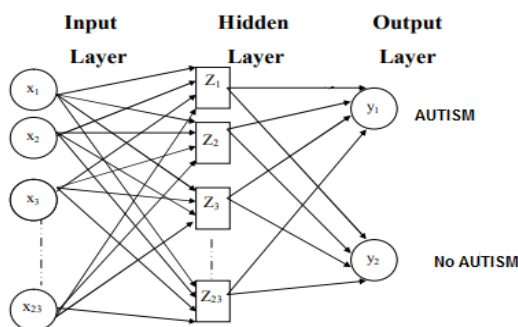


Fig 1. Graphical representation of MLP

Each training record is fed forward to the network and classified by adjusting the weights continuously. The correct class of each record is known in the training phase. The network's calculated values for the output nodes are compared to the correct values. The error term is calculated to adjust the weights of the hidden layers for each node. During the learning phase, the network learns by adjusting the weights, to predict the correct class label of input samples.

B. Bayesian Networks

Bayesian classifiers are statistical classifiers predicting class membership probabilities. They describe the probability distribution governing a set of variables. They specify a set of conditional independence assumptions along with a set of conditional probabilities. It refers to a type of model that uses Directed Acyclic Graphs (DAGs) where nodes represent features and arcs represent the causal relationships between them. There is one conditional probability table (CPT) for each node given its parents. If the node has no parents, then the probability distribution is unconditional; otherwise it is conditional, where the probability of each feature value depends on the values of its parents.

C. Naïve Bayes Classifier

Naïve Bayes classifier is a variant of Bayesian network with the simple structure that has the class node as the parent node of all other attributes. Naive Bayes classifier estimates the class-conditional probability by assuming that

the attributes are conditionally independent, given the class label C_k . Suppose that there are n classes, C_1, C_2, \dots, C_n , the conditional independence assumption can be formally stated in Equation (Eq. 1).

$$P(X | C_k) = \prod_{i=1}^m P(X_i | C_k) \quad (\text{Eq. 1})$$

where every attribute set $X = \{x_1, x_2, \dots, x_m\}$ consists of m attributes.

D. J48

Decision trees are powerful tools for classification and prediction. J48 is an inductive learning algorithm designed by Quinlan [13]. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the attributes. The algorithm constructs the tree starting from a training set. It uses divide and conquer strategy to create a tree structure which represents the rules. It calculates the information gain for each attribute and then selects the attribute with the highest information gain. It is a hierarchical structure where, at each level, a test is applied to one or more attribute values that may have one of two outcomes (Mitchell 1997)[14]. The outcome may be a leaf which allocates a class, or a decision node which specifies an additional test on the attribute values and forms a sub-tree for each possible outcome of the test.

III. RESULTS AND DISCUSSION

A. Experimental Setup

The experiment was carried out by using UCI Autism Adult Data set was analyzed using J48 Decision Tree classifier, Multilayer Perceptron classifier, Naïve Bayes Classifier and Bayesian Network classifier using WEKA open source data mining tool. The results obtained by these algorithms were compared based on various measures. The steps followed in analysing the algorithms performance on Autism data set is specified in the Fig.2.

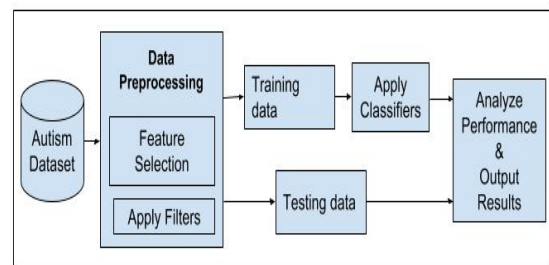


Fig. 2. Experimental Setup

The Autism Adult data set consists of 21 attributes with 704 instances. Out of these 21 attributes, the attributes that does not contribute for early prediction autism spectrum disorder are removed using feature selection algorithms. The missing values are replaced and the continuous data attributes were discretized. Four different classifiers are applied on the autism data set with different runs of 10% to 90% trained data. The J48 classifier generated results in the form J48 pruned tree as in the Fig. 3.



```

==== Classifier model for training split (141 instances) ====
J48 pruned tree
-----
A9_Score <= 0
| A5_Score <= 0: NO (55.0)
| A5_Score > 0
| | A1_Score <= 0: NO (10.0)
| | A1_Score > 0
| | | A10_Score <= 0: NO (11.0/1.0)
| | | A10_Score > 0
| | | | A4_Score <= 0: NO (6.0/1.0)
| | | | A4_Score > 0: YES (7.0/1.0)
A9_Score > 0
| A4_Score <= 0
| | A3_Score <= 0: NO (8.0)
| | A3_Score > 0: YES (3.0/1.0)
| A4_Score > 0
| | A10_Score <= 0
| | | austin = no: NO (6.0/1.0)
| | | austin = yes: YES (2.0)
| | A10_Score > 0
| | | A1_Score <= 0
| | | | A2_Score <= 0: NO (2.0)
| | | | A2_Score > 0: YES (2.0)
| | | A1_Score > 0: YES (29.0/1.0)

Number of Leaves :      12

Size of the tree : 23
    
```

Fig. 3. J48 Pruned tree for the 20% test split

Assessment of prediction accuracy is an important aspect of classification. The classifier’s performance can be analyzed and compared by the measures generated by the confusion matrix. The performance of the proposed algorithms in terms of accuracy, precision, recall rates and F-measure are analyzed.

B. Classification accuracy

Accuracy is the proportion of instances whose class the classifier can correctly predict and is calculated by the following Equation (Eq.2).

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (Eq. 2)$$

The accuracy percentage of the four algorithms is shown as a line chart in Fig.4. It illustrates the performance comparison of Multilayer Perceptron(MLP) classifier with J48decision tree (J48) classifiers, Naïve Bayes (NB) classifier and Bayesian network (BN) classifier. It can be seen that Multilayer Perceptron classifier achieved the highest accuracy among the four algorithms for different training sample sizes.

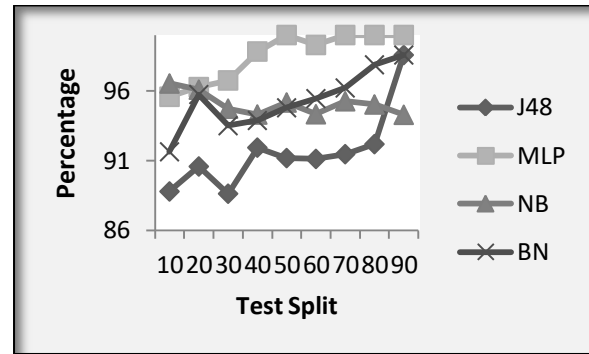


Fig. 4. Classification accuracy for MLP, J48, NB and BN classifiers

It can be clearly seen that the highest accuracy is 100% by Multilayer Perceptron classifier and the lowest is 88.64% by J48 decision tree. The results show that the Multilayer Perceptron classifier are able to obtain classification rate of 100% for 50%, 70%, 80% and 90% of training samples respectively, which is much higher than J48 decision tree, Naïve Bayes and Bayesian network algorithms. The results indicate that Multilayer Perceptron classifier should be favoured over J48 decision tree, Naïve Bayes and Bayesian network classifiers in handling early prediction of autism in adult especially when the accuracy of classification is important.

C. Precision, Recall and F-measure

Measuring accuracy alone is not sufficient to assess the real performance of a classifier, because it will yield misleading results if the data set is unbalanced. Other commonly used measures for evaluation of classifiers are precision, recall and F-measure. They are calculated using the following equations for various training and testing sample sizes.

Precision is the proportion of predicted positives which are actually positive and is given by the following Equation (Eq.3):

$$Precision = TP / (TP + FP) \quad (Eq. 3)$$

Recall (Sensitivity) is the proportion of real positive cases that are correctly predicted as positive. Sensitivity predicts how good a test is at detecting positives and is given by the following Equation (Eq. 4).

$$Recall = TP / (TP + FN) \quad (Eq. 4)$$

The range of Precision and recall values is from 0 to 100%. The precision and recall values for different sample sizes are shown in TABLE.I.

The result shows that Multilayer Perceptron yields an increase of 96.8% in precision and in recall rate when compared to the J48 decision tree classifier for 30% of training samples. It can also be seen that there is an increase of 100% in precision and in recall rate and 100% in precision and in recall rate for 50% and 70% of training samples respectively. Also Naïve Bayes classifier and Bayesian network classifier yields an increase of 94.7% in precision rate and in recall rate when compared to J48

decision tree classifier for 30% of the training samples. It can be clearly seen that Multilayer Perceptron has the highest precision and recall rates for various training sample sizes.

TABLE.I. Comparison of precision and recall rate of the classifiers

Test Split %	J48		Multilayer Perceptron		Naïve Bayes		Bayesian Network	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
10	89.1	88.8	95.8	95.6	96.5	96.5	91.8	91.6
20	90.7	90.6	96.4	96.3	96.1	96.1	95.8	95.7
30	89.5	88.6	96.8	96.8	94.7	94.7	93.5	93.5
40	91.8	91.9	98.8	98.8	94.3	94.3	94.3	94.2
50	91.3	91.2	100.0	100.0	95.1	95.2	94.8	94.8
60	91.0	91.1	99.3	99.3	94.3	94.3	95.4	95.5
70	91.3	91.5	100.0	100.0	95.3	95.3	96.2	96.2
80	92.1	92.2	100.0	100.0	95.0	95.0	97.9	97.9
90	98.6	98.6	100.0	100.0	94.8	94.3	98.6	98.6

Fig.5.illustrates the precision and recall values of Multilayer Perceptron, Bayesian network, Naïve Bayes and J48 decision tree classifier for 30% of the training samples.

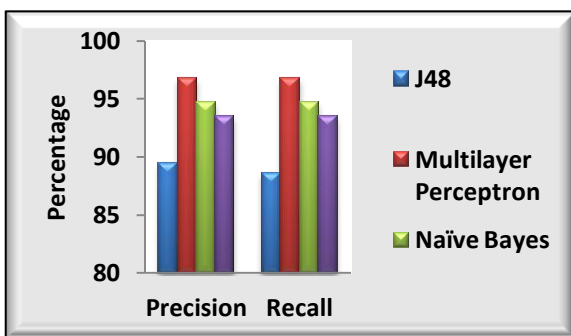


Fig. 5. Comparison of precision and recall rate for 30% test split samples

Both precision and recall focus only on the positive examples and predictions. F-measure combines precision and recall into a single measure of performance and is given by the following Equation (Eq. 5).

$$F\text{-measure} = 2 \times ((\text{precision} \times \text{recall}) / (\text{precision} + \text{recall})) \quad (\text{Eq.5})$$

F-measure value ranges between 0 to 100% and a larger F-measure value indicates a higher classification quality. A comparison of the results based on F-measure for four classifiers are shown in Fig. 6.

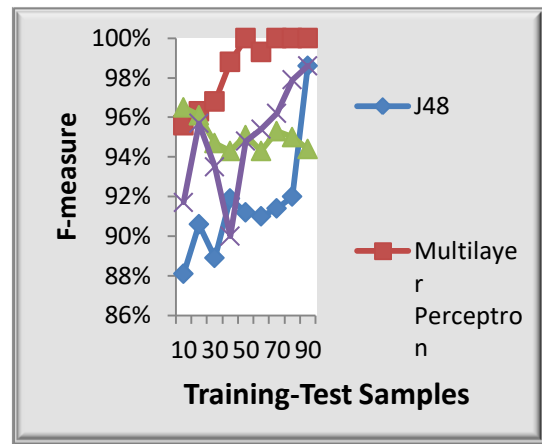


Fig.6. Comparison of F-measure values for the four classifiers

The quality of the best F-measure result of Multilayer Perceptron (100%) is achieved for 50% training samples. This is higher than Bayesian Network and Naïve Bayes by 5% and J48 by 9%. The F-measure values for Multilayer Perceptron are 100 % for 70%, 80% and 90% of the training samples respectively. It can be clearly seen that Multilayer Perceptron has highest value for F-measure which ensures that both precision and recall are reasonably high.

D. Paired t- Test

A best machine learning algorithm for a problem can be found empirically by trial and error method. Paired – t test, a statistical procedure is used to compare the performance of one algorithm with the other as base algorithm. The PairedCorrectedTTester was applied for analysing the difference in accuracy of the four algorithms J48, Multilayer Perceptron, Naïve Bayes and Bayesian Network. With a significance level set to 0.5, the accuracy of correct predictions was tested with runs from 10% to 90% training data. It was found that the Multilayer Perceptron is statistically significant in most of the cases when compared with the other three algorithms. A comparison of the results based on the performance accuracy at 10% training data is shown in Fig.7.

Dataset	(1) trees.J4	(2) funct	(3) bayes	(4) bayes
csv_result-Autism-Adult-D (10)	87.23	95.11	93.37	93.97
	(v/*)	(1/0/0)	(0/1/0)	(0/1/0)

Fig.7. Comparison of performance accuracy of the four classifiers for 10% training data

The symbol cited alongsidethe results indicate that it is statistically better (v) or worse (*) than the baseline classifier. It was found from various test runs, that in most of the cases, the performance accuracy Multilayer Perceptron classifier is denoted as ‘v’(better) and is found to be significantly better than J48 Decision Tree, Naïve Bayes and Bayesian Network classifier algorithms.

IV. CONCLUSION

This study evaluates the performance of machine learningclassification techniques in early prediction of Autism Spectrum Disorder in Adults. UCI autism disorder dataset was used for analysis. The empirical results were analysed using Multilayer Perceptron, J48 Classifier, Naïve Bayes and Bayesian Networks. Comparing the performance, the analysis shows that Multilayer Perceptronalgorithm has the highest classifier accuracy than J48, Naive Bayes Classifier and Bayesian Classifier. The experimental results also reveal that even in terms of other classifier measures, Multilayer Perceptron performs exceptionally better than the other algorithms. The superiority of Multilayer Perceptron in predicting early detection is established through this empirical study.

REFERENCES

1. M. S. Mythili And A. R. Mohamed Shanavas, “An Improved Autism Predictive Mechanism Among Children Using Fuzzy Cognitive Map And Feature Extraction Methods (Feast)”, ARPN Journal Of Engineering And Applied Sciences, Vol. 11, No. 3, February 2016.
2. Xia-an Bi*, Yang Wang, Qing Shu, Qi Sun and Qian Xu, “Classification of Autism Spectrum Disorder Using Random Support Vector Machine Cluster”,Journal of Front Genet. 2018, Vol. 9,Published online Feb 2018.
3. Bram van den, “Using Machine Learning for Detection of Autism Spectrum Disorder”, *Journal of Autism and Developmental Disorders*, 46(7):2317–2326, 2016.
4. Kosmicki, J. A., Sochat, V., Duda, M., & Wall, D. P., “Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning”, *Translational Psychiatry*, 5, e514, (2015).
5. F. Hauck and N. Kliewer, “Machine Learning for Autism Diagnostics: Applying Support Vector Classification”,Int'l Conf. Health Informatics and Medical Systems, HIMS'17, FreieUniversität Berlin, Berlin, Germany.
6. Duda, M., Ma, R., Haber, N., & Wall, D. P. , “Use of machine learning for behavioral distinction of autism and adhd.*Translational Psychiatry*”, 6, e732.
7. Bone, D., Bishop, S. L., Black, M. P., Goodwin, M. S., Lord, C., & Narayanan, S. S, “Use of machine learning to improve autism screening and diagnostic instruments: effectiveness, efficiency, and multi-instrument fusion”, *Journal of Child Psychology and Psychiatry*, 57, 927-937, 2016.
8. Venkatesan.M, ArunkumarThangavelu. “Event Centric Modeling Approach in Co-location Pattern Analysis from Spatial Data”, *International Journal of Database Management Systems (IJDMS)*. 3(3): 125- 133, 2011.
9. Werner E, Dawson G, Munson J, Osterling J, “Variation in early developmental course in Autism and its relation with behavioral outcome at 3-4 years of age”, *J Autism Dev Disorder*. 35(3):337-50, 2005.
10. Tabtah, F. , “Autism Spectrum Disorder Screening: Machine Learning Adaptation and DSM-5 Fulfillment”, *Proceedings of the 1st International Conference on Medical and Health*

- Informatics 2017, pp.1-6, 2017, Taichung City, Taiwan, ACM.
11. Thabtah, F, “Machine Learning in Autistic Spectrum Disorder Behavioural Research: A Review”, *Informatics for Health and Social Care Journal*, December, 2017.
12. Demirhan, A., “Performance of machine learning methods in determining the autism spectrum disorder cases”, *Mugla Journal of Science and Technology*, 4(1), 79-84, 2018.
13. Quinlan, J.R. “C4.5: Programs for Machine Learning”, Morgan Kaufman Publishers, 1993.
14. Mitchell, M. *Machine Learning*, McGrawHill, 1997.
15. Cohen, W.W. “Fast Effective Rule Induction”, in *Proc. of the 12th International Conference on Machine Learning*, pp.115-123,1995.
16. Langley, P., Iba, W. and Thompson, K. “An Analysis of Bayesian Classifiers”, *National Conference on Artificial Intelligence*, pp.223- 228, 1992.
17. Vapnik, V. *The Nature of Statistical Learning Theory*. Springer, Berlin Heidelberg, New York, 1995.