International Journal of Engineering & Technology, 7 (4) (2018) 4216-4219

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET
doi: 10.14419/ijet. v7i4.14907
Research paper

A machine learning based approach to classify autism with optimum behavior sets

Vaishali R [1] *, Sasikala R [1]

[1] *School of Computer Science and Engineering, VIT University, Vellore, India*
**Corresponding author E-mail: sasikala.ra@vit.ac.in*

Abstract

Machine Learning based behavioural analytics emphasis the need to develop accurate prediction models for detecting the risk of autism faster than the traditional diagnostic methods. Quality of prediction rely on the accuracy of the supplied dataset and the machine learning model.To improve accuracy of prediction, dimensionality reduction with feature selection is applied to eliminate noisy features from a dataset. In this work an ASD diagnosis dataset with 21 features obtained from UCI machine learning repository is experimented with swarm intelligence based binay firefly feature selection wrapper. The alternative hypothesis of the experiment claims that it is possible for a machine learning model to achieve a better classification accuracy with minimum feature subsets.Using Swarm intelligence based single-objective binary firefly feature selection wrapper it is found that 10 features among 21 features of ASD dataset are sufficient to distinguish between ASD and non-ASD patients.The results obtained with our approach justifies the hypothesis by producing an average accuracy in the range of 92.12%-97.95% with optimum feature subsets which is approximately equal to the average accuracy produced by entire ASD diagnosis dataset.

Keywords: *Autism Spectrum Disorder; Behavioral Analytics; Machine Learning, Feature.*

1. Introduction

Autism is a childhood disorder which has become more prevalent among younger generations in the recent decade. According to the centre for disease control and prevention, there is a sustainable growth in the number of children diagnosed with Autism disorder. According to them, 1 among 68 Children under the age of 8 in the United States of America is diagnosed with autism [1] Autism diagnosis is a clinical examination procedure conducted according to the DSM-V standards for disorder classification [2] These standards are coined by the US Mental health professionals based on their successful diagnostic experiences and contributions. These procedures are widely incorporated in behavioral analytics for classification of ASD from non-ASD. In addition to DSM-V standards, interview and questionnaire based clinical examinations are also followed for behaviour classification. ADI-R and ADOS are some common behaviour tests carried out by pediatricians for detection of childhood autism symptoms. Certified professionals in laboratory conditions practice these clinical experiments. The assessments can last for 60 minutes of duration based on the patient's responsiveness. The certified professional awards a binary score based on the quality of response. Consolidated scores decide the severity of autism in the patients.

In [3], an ASD diagnostic dataset comprised of 21 behavioural attributes is taken for classification task of ASD patients from non-ASD. This work has adapted a mobile application based ASD screening approach obeying the DSM-V fulfillment for Autism detection. The behaviour dataset has collected 292 samples of children Autism screening episodes. In, the researcher suggests feature selection as a measure for improving prediction accuracy of machine learning models. This process of obtaining better accuracy with an optimum feature subset that represent the structure of entire dataset is called Dimensionality reduction. Among two approaches of dimensionality reduction, feature selection is recommended for real world datasets. Any dataset that exceeds 10 features fall under the problem of high dimensionality. In ASD dataset there are 21 features, which makes it a high dimensional dataset according to the previous claim. In the literature a work to classify ASD from ADHD has applied a filter based forward feature elimination approach on a different ASD behaviour dataset consisting of 64 features [4]. The work claimed [5] attributes are sufficient for efficient classification. In [5], a backward feature elimination approach was to select features for classification of ASD patients from ADHD based on differences in the behaviour patterns. There exist very few research works on machine learning based diagnosis of ASD due to unavailability of datasets for public access. In 2017, Fadi Thabtah has published the ASD screening dataset of children, adult and adolescent in UCI machine learning repository for public access. These datasets were analyzed in their work on ASD prediction with machine learning [3], [6]. R [7] and Weka tool [8] were used for building machine learning models for classification of ASD from non-ASD patients.In this paper we analyze ASD diagnosis dataset of children with 292 instances for behavioural analytics and prediction tasks. As the repository mentions the presence of missing values in the dataset, missing data imputation approach for noise reduction is applied to check completeness of the dataset. The organization of this paper include

i) Discussion on Problem statement and solution
ii) Materials and Methods of study
iii) Dataset and Pre-Processing tasks
iv) Analysis
v) Interpretations from the results
vi) Future Scope of the research

2. Problem statement and proposed solution

ASD dataset with 21 attributes contains 20 features and 1 binary class attribute. This dataset can produce 2^{20} feature subsets for evaluation. Exhaustive search based feature subset selection algorithms will face exponential increase in time complexity as feature subselection is classified as an np-hard problem [9]. Stochastic search algorithms with objective evaluation function and feature elimination algorithms with candidate evaluation are best solutions to overcome the np-hard search problems. In [5] and [4], feature elimination approach by individual candidate evaluation and selection based on ranking is opted as a feature selection strategy. According to [10]features selected by ranking approach are highly prone to inter-feature correlation bias and thus leads to redundancy among features which inversely affect the performance of machine learning model. [11], [12] have proposed swarm intelligence based feature selection wrappers as better alternatives to avoid inter-feature correlation with correlation bias as an objective function in feature subset evaluation. Among stochastic algorithms, bio-inspired swarm intelligence wrappers are better explorers in feature selection. This makes swarm intelligence wrappers as a better choice to explore more possibilities in minimum iterations and produce results that meets the objective of selection. In this paper we propose a swarm intelligence based feature selection wrapper combining Binary firefly algorithm for feature selection with a single objective function considering maximum accuracy and minimum features to decide fitness of subsets [13].

3. Materials and methods of study

3.1. Feature selection for machine learning

In module 1, the ASD children dataset is trained with 8 different machine learning algorithms using 10 fold cross validation.The results of module 1 are compared with results of machine learning models obtained after feature selection with Binary firefly feature selection algorithm. The configuration and pseudocode of feature selection algorithm is listed below. The Binary firefly algorithm for feature selection implemented in this work is introduced as an optimizer in and developed as feature selection algorithm in [21].

3.2. Binary firefly feature selection wrapper

The binary firefly feature selection algorithm is proposed in [22], [23]for optimization of classification and regression algorithms. This feature selection algorithm is a recent and fast performer that has outperformed benchmark algorithms such as on 40 datasets. Binary firefly algorithm is accelerated with a logistic chaotic map to boost attractiveness. The local and global search strategy of feature selection is enhanced by simulated annealing. Thus the algorithm converges towards global best solution within minimum iterations. The binary firefly feature selection algorithm is classified as a swarm intelligence optimizer based wrapper feature selection algorithm with a single objective function. General architectural working model of wrappers is shown in Fig. 1.

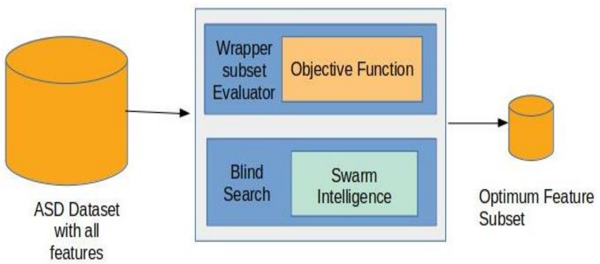

Fig. 1: Architectural Working Model of Swarm Intelligence Based Feature Selection for Feature Subset Evaluation with a Single Objective Evaluation Function on ASD Datasets.

In addition to the architectural working model, the flow chart of the binary firefly feature selection algorithm can provide deep insights into the working of the algorithm.

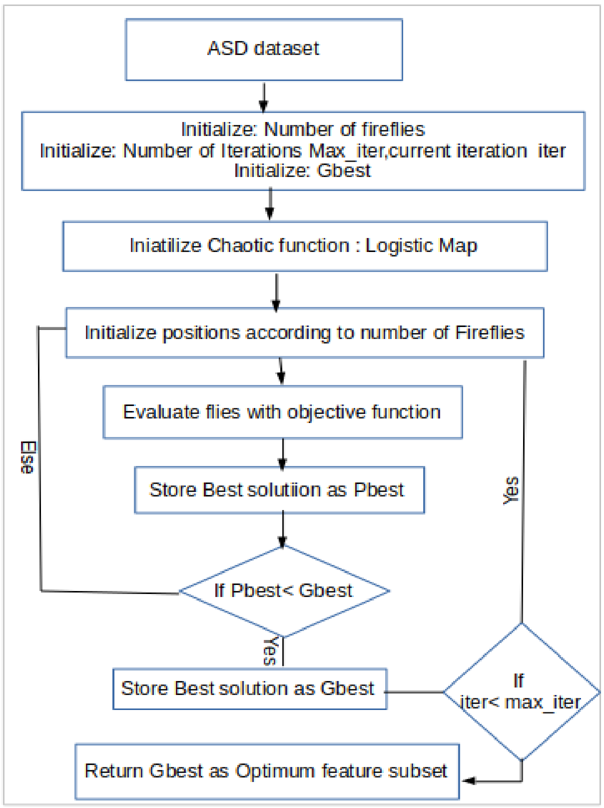

Fig. 2: Flowchart on the Working of Binary Firefly Algorithm for Feature Selection on ASD Diagnosis Dataset.

4. Dataset and pre-processing

Experimental results of machine learning algorithms before and after feature selection of ASD children diagnosis dataset is tabulated for analysis in table 2. The experiments are executed according to the setup and configuration discussed above.

The Binary firefly feature selection algorithm has selected 10 featured subset among 21 features in the dataset as optimum. The feature reduction ratio of this algorithm is 0.48.

Table 1: Parameter Setup of the Firefly Feature Selection Wrapper Algorithm and Machine Learning Algorithms

System Configuration	Intel i5 5th Gen, 12GB RAM
Tools	R and Weka
Feature Selection Algorithm	Firefly wrapper
Category	Swarm Intelligence search
Evaluation algorithm	k-NN wrapper (k=5)
Distance	Euclidean
Number of Particles	30
No of Iterations	100
Objective	Maximum accuracy and Minimum Features
Objective Type	Single Weighted fuzzy fitness function
Chaotic function	Logistic map
Machine Learning model Configuration	
Naive Bayes	A simple probability based Bayesian classification algorithm for prediction [14]
J48 Decision Tree	A tree based decision tree classifier based on C4.5 by R.Quinlan [15]
SVM	A discriminant function based classifier classifies data with hyperplanes and Kernels [16]–[18]
K-NN	A distance based classification algorithm based on nearest values. [19]

MLP	Back propagation neural network based classification algorithm [20]
CV Partition	10 Fold Cross validation
Evaluation Metrics	
Feature Reduction Ratio	Ratio of number of features selected from total feature set
Accuracy	Percentage of Instances correctly classified
TP Rate	Ratio of Correctly Predicted positive instances
RMSE	Bias rate in prediction
ROC Area	Area under the curve calculated by integrated start and end points of a graph

Table 2: Analysis of Performance of Various Machine Learning Algorithms on the ASD Children Dataset before and after Feature Selection with Binary Firefly Algorithm

	Accuracy		TP Rate		ROC area		RMSE	
	B	A	B	A	B	A	B	A
NB	93.15	95.55	0..93	0.96	0.99	1.00	0.22	0.20
J48	91.10	92.12	0.91	0.92	0.89	0.90	0.30	0.28
SVM	99.66	97.95	1.00	0.98	1.00	0.98	0.06	0.14
KNN	87.67	93.84	0.88	0.93	0.97	0.97	0.30	0.23
MLP	99.66	97.60	1.00	0.98	1.00	1.00	0.052	0.14

Evaluation of various machine-learning models on ASD children diagnosis dataset observed an accuracy in the range of (87.67% to 99.66%) on original dataset. K-NN classifier with K=5 has produced the least accuracy of 87.67% with RMSE score of 0.30. Multilayered Perceptron and Support vector machine classifiers produced 99.66% prediction accuracy on original dataset. J48 decision tree and Naive bayes classifier had shown medium performance. MLP and SVM classifiers with considerably minimum RMSE scores of 0.05 and 0.06 respectively achieve maximum ROC of [1]. These algorithms have achieved maximum true positive rate of [1] whereas other algorithms have undergone misclassification errors affecting True positive rate.
After feature selection the number, features are reduced to 10. The features selected are A1_Score A2_Score, A3_Score, A4_Score, A5_Score, A7_Score, A8_Score, A9_Score, A10_Score and relation. On training machine learning models with these selected features, the accuracy obtained are in the range of (92.12%-97.95%). K-NN model produced 93.84% of accuracy which shows 6.17% improvement than K-NN model trained with original dataset. Except SVM and MLP models other three models trained with optimum behaviour set have shown a considerable improvement in the accuracy, TP Rate, ROC and RMSE.

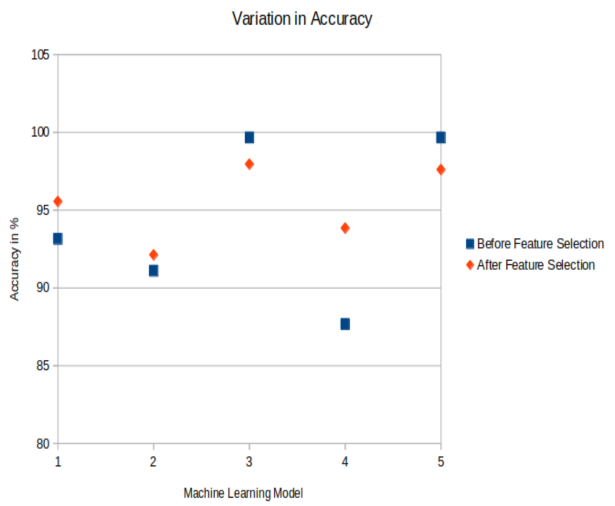

Fig. 3: A Graphical Representation of Variation in Accuracy of the Machine Learning Models before and after Feature Selection with Binary Firefly. from the Figure, It Is Clear That the after Feature Selection That Models Are Able to Perform Better or Perform Closer to the Machine Learning Models Built on Entire Datasets.

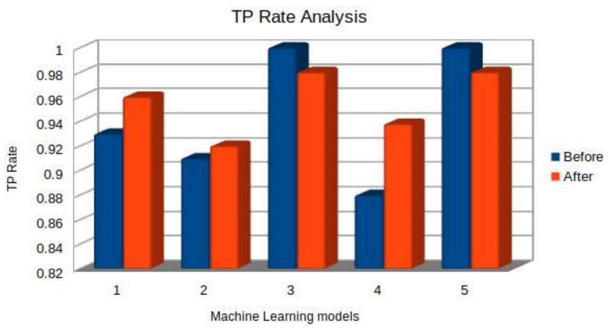

Fig. 4: TP Rate Obtained by Machine Learning Models before and after Feature Selection is visualized in the Figure.

True positive refers to the number of positive instances correctly classified as True. This measure is given importance because the impact of positive instances falsely classified as false may result lethal effect on the prediction model. If a kid suffering from Autism is wrongly classified as Autism free, then it may affect the treatment of the Kid as well as delays diagnostic process resulting in complications. Hence TP rate is considered as an important evaluation factor in terms of medical datasets. False Positive rates are negotiable, as the therapist's intervention may prove it wrong anytime and it does not interrupt or affect the diagnosis of the child.
After feature selection, there is an improvement in the TP rate and in cases of SVM and MLP, the TP rate is considerably better and closer to the actual model.

5. Interpretations

Among 292 instances in ASD children dataset,there are 151 instances with class 'yes' and 141 instances with class 'No'. This shows that the chosen dataset is void of class imbalance problem.
Due to the presence of 21 attributes, the dataset becomes high dimensional and faces NP-hard problem in feature selection. Stochastic Swarm intelligence algorithms with fixed number of iterations and exploration capacity are better choices for optimum feature subset selection.
Binary Firefly algorithm for feature selection opted is a fast explorer than existing swarm intelligence search algorithms.
Comparison of results of machine learning models before and after feature selection showed that 3/5 machine learning models have considerable performance improvement with the optimum behaviour sets.
Presence of 15% missing values in the selected Relation attribute might have caused deterioration of quality of models in the functional classifiers such as SVM and MLP. However the performance of functional models built with optimum behaviour set is better than the other classification models.
Due to lesser amount of instances in the dataset, there exist of chance of model overfitting on the dataset.
From the above interpretations it is clear that the optimum behaviour set has improved the prediction performance of machine learning models in 3/5 cases and in 2/5 cases the behaviour set has exhibited a decent performance with minimum features. These observations validate the alternative hypothesis: Minimum behaviour sets can retain the structure of the entire dataset in machine learning.

6. Conclusion

This paper aimed to design an automated ASD prediction model with minimum behaviour sets selected from ASD diagnosis dataset with Binary Firefly algorithm for feature selection. The hypothesis of this paper is to find whether machine learning models trained with minimum behaviour sets are capable of better performance or not. In order to select features a swarm intelligence based wrapper is considered as a better alternative to Ranking based feature elimination algorithms. From the above results and discussions the hypothesis is validated.

7. Future work

UCI repository indicates the presence of missing instances in the ASD child dataset which is not handled in the present work. Rather it is assumed that the dataset is complete and evaluation is done. This assumption could have impacted on the performance of feature selection and machine learning. In future, a suitable missing data imputation framework should be designed to check the presence of missing data in the dataset. Even Though swarm intelligence wrappers are better explorers than traditional feature selection , there exist their own disadvantages in terms of risk of overfitting, time complexity and search complexity. These factors should be addressed in the future work.

Acknowledgement

The authors thank VIT University for providing 'VIT SEED GRANT' for carrying out this research.

References

[1] J. Baio, "Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years—Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014," *MMWR. Surveill. Summ.* vol. 67, 2018.

[2] A. P. Association and others, *Diagnostic and statistical manual of mental disorders (DSM-5®).* American Psychiatric Pub, 2013. https://doi.org/10.1176/appi.books.9780890425596.

[3] F. Thabtah, "Autism Spectrum Disorder Screening: Machine Learning Adaptation and DSM-5 Fulfillment," in *Proceedings of the first International Conference on Medical and Health Informatics 2017*, 2017, pp. 1–6. https://doi.org/10.1145/3107514.3107515.

[4] M. Duda, R. Ma, N. Haber, and D. P. Wall, "Use of machine learning for behavioral distinction of autism and ADHD," *Transl. Psychiatry*, vol. 6, no. 2, p. e732, 2017. https://doi.org/10.1038/tp.2015.221.

[5] J. A. Kosmicki, V. Sochat, M. Duda, and D. P. Wall, "Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning," *Transl. Psychiatry*, vol. 5, no. 2, p. e514, 2015. https://doi.org/10.1038/tp.2015.7.

[6] F. Thabtah, "Machine learning in autistic spectrum disorder behavioral research: A review and ways forward," *Informatics Heal. Soc. Care*, vol. 0, no. 0, pp. 1–20, 2018. https://doi.org/10.1080/17538157.2017.1399132.

[7] R. C. Team and others, "R: A language and environment for statistical computing," 2013.

[8] G. Holmes, A. Donkin, and I. H. Witten, "Weka: A machine learning workbench," in *Intelligent Information Systems, 1994. Proceedings of the 1994 Second Australian and New Zealand Conference on*, 1994, pp. 357–361. https://doi.org/10.1109/ANZIIS.1994.396988.

[9] W. Siedlecki and J. Sklansky, "On automatic feature selection," *Int. J. Pattern Recognit. Artif. Intel.* vol. 2, no. 02, pp. 197–220, 1988. https://doi.org/10.1142/S0218001488000145.

[10] L. Tolosi and T. Lengauer, "Classification with correlated features: unreliability of feature ranking and solutions," *Bioinformatics*, vol. 27, no. 14, pp. 1986–1994, 2011. https://doi.org/10.1093/bioinformatics/btr300.

[11] X. Wang, J. Yang, X. Teng, W. Xia, and R. Jensen, "Feature selection based on rough sets and particle swarm optimization," *Pattern Recognit. Lett.* vol. 28, no. 4, pp. 459–471, 2007. https://doi.org/10.1016/j.patrec.2006.09.003.

[12] A. Unler, A. Murat, and R. B. Chinnam, "mr2PSO: A maximum relevance minimum redundancy feature selection method based on swarm intelligence for support vector machine classification," *Inf. Sci. (Ny).* vol. 181, no. 20, pp. 4625–4641, 2011. https://doi.org/10.1016/j.ins.2010.05.037.

[13] H. Banati and M. Bajaj, "Fire fly based feature selection approach," *IJCSI Int. J. Comput. Sci. Issues*, vol. 8, no. 4, 2011.

[14] G. H. John and P. Langley, "Estimating continuous distributions in Bayesian classifiers," in *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, 1995, pp. 338–345.

[15] J. R. Quinlan, "C4. 5: Programming for machine learning," *Morgan Kauffmann*, vol. 38, p. 48, 1993.

[16] S. S. Keerthi, S. K. Shevade, C. Bhattacharyya, and K. R. K. Murthy, "Improvements to Platt's SMO algorithm for SVM classifier design," *Neural Comput.*, vol. 13, no. 3, pp. 637–649, 2001. https://doi.org/10.1162/089976601300014493.

[17] J. C. Platt, "12 fast training of support vector machines using sequential minimal optimization," *Adv. kernel methods*, pp. 185–208, 1999.

[18] T. Hastie and R. Tibshirani, "Classification by pairwise coupling," in *Advances in neural information processing systems*, 1998, pp. 507–513.

[19] D. W. Aha, D. Kibler, and M. K. Albert, "Instance-based learning algorithms," *Mach. Learn.*, vol. 6, no. 1, pp. 37–66, 1991. https://doi.org/10.1007/BF00153759.

[20] S. K. Pal and S. Mitra, "Multilayer perceptron, fuzzy sets, and classification," *IEEE Trans. neural networks*, vol. 3, no. 5, pp. 683–697, 1992. https://doi.org/10.1109/72.159058.

[21] X.-S. Yang, "Firefly algorithm, Levy flights and global optimization," in *Research and development in intelligent systems XXVI*, Springer, 2010, pp. 209–218. https://doi.org/10.1007/978-1-84882-983-1_15.

[22] H. Banati and M. Bajaj, "Fire fly based feature selection approach," *IJCSI Int. J. Comput. Sci. Issues*, vol. 8, no. 4, 2011.

[23] L. Zhang, K. Mistry, C. P. Lim, and S. C. Neoh, "Feature selection using firefly optimization for classification and regression models," *Decis. Support Syst.*, 2017.