



DRN HYBRID MODEL FOR PREDICTING AUTISM USING RAPID MINER TOOL

R. Ramya

Research Scholar

Department of Computer Science
Bharathidasan University
Trichy, Tamilnadu, India

B.S.E. Zoraida

Assistant Professor

Department of Computer science
Bharathidasan University
Trichy, Tamilnadu, India

Abstract: Autism is wrong connection between cells in the human brain which cause abnormalities in the brain structure or function. Every human being with Autism Spectrum Disorder (ASD) has unique symptoms and abilities. Symptoms of ASD typically appear during the first three years of human life. Autism had been classified as three different types such as severe autism, moderate autism, and mild autism. Diagnosing ASD is based on ASD historical dataset because there is no blood or other medical test. With this in mind this paper focuses on developing new hybrid DRN model is created by combining three different models like Deep Learning, Random Forest, and Naïve Bayes (DRN) models. DRN hybrid model is implemented in Rapid Miner tool to find the Accuracy, Precision, recall, Classification error and Executed time. The result obtained shows DRN model is better when compared to the existing models like, Ada Boost, Bagging, Vote, Stacking and Bayesian Boosting models. Hence DRN hybrid can be used to predicting autism using the historical dataset.

Keywords: Autism Spectrum Disorder (ASD), New DRN (Deep Learning, Random Forest, Naïve Bayes) Model, Stacking, Vote, Ada Boost, Bagging, Bayesian Boosting, Rapid Miner.

I. INTRODUCTION

Autism is a lifelong complex developmental disorder in the human brain. Autism occurs in the early stage of human brain development especially, during the first three years of life. Everyday human brains interpret the things happening in the environment to smell, hear, taste and touch but a child with ASD has problem of interpreting. Every child with autism has unique abilities, wide range of symptoms and challenges. Some symptoms of ASD are lack of communication, social-behavioral challenges and Repetitive behaviors. In this symptoms based Autism has been divided into three different level as severe autism (Low level), moderate autism (medium level), and mild autism (high level). ASD diagnosing is very complex and time consuming because early stage symptoms are found common in adult. For this reason of diagnosing is based on the historical dataset in ASD. In this work historical dataset of ASD is collected from VAERS website. In VAERS website has collected information during the adverse events that occur after administration of vaccine. In this website approximately 30,000 reports are filled and each fields of the report has patient details.

In this paper a create new Hybrid DRN model is developed by combining three different models like Deep Learning, Random Forest and Naïve Bayes models. A new DRN hybrid model is used to predict autism using historical dataset.

II. LITERATURE REVIEW

Michael Siller and Marian Sigman proposed developmental relationships connecting the child's subsequent development of communication skills and parental understanding in adult with ASD [1]. Tony Charman *et al.* described an consideration should be directed at assessing these skills in 2

and 3 year old children referred for a diagnosis of autism spectrum disorder [2]. G. Leroy *et al.* proposed a data mining techniques for decision tree and association rule using to predict autism affect child behavior provided more detailed insights for high and low levels of appropriate and inappropriate behavior [3]. S.Wheelwright *et al.* investigating by the relationship between scores on the Empathy Quotient, Systemizing Quotient-Revised and Autism Spectrum Quotient in both a large sample of typical participants and sample of adults with autism spectrum conditions (ASC) [4]. Sheena angra and Sachin ahuja described an predict grades of students of one class by analyzing the grades of previous classes using K-Means, Linear Regression and Decision Tree in Rapid Miner environment [5]. Brian P. Keane *et al.* suggest that the ability to integrate between the auditory and visual sense modalities is unimpaired among high-functioning adults with autism [6]. A.Martin *et al.* used to the hybrid model of genetic algorithm, Fuzzy c-means algorithm, MARS for prediction of bankruptcy [7]. Tanaya Guha *et al.* describe reduced complexity in facial expression dynamics of subjects with High Functioning Autism relative to their typically developing peers. Significant difference is observed for expressions related to disgust, joy and sadness [8]. Ahmed Hassan *et al.* used Naïve Bayes (NB), Random Forest (RF), Decision Trees (C4.5), Support Vector Machines (SVM), K-Nearest Neighbor (KNN) and Logistic Regression (LR) to evaluate the classification performance by using three microarray gene expression datasets namely the Leukemia Cancer, Lung cancer and breast Cancer datasets [9]. Niyati Gupta *et al.* used to the various classification methods and compared the results of various algorithms on Weka on the basis of accuracy, sensitivity and specificity with the results of various other algorithms

implemented on Matlab and Rapid Miner [10]. E. M. Albornoz *et al*. presented an evolutionary method for automatic selection of features of speech in a classification task for autism and It is based on a genetic algorithm that selects the best combination of acoustic features using support vector machines as classifier [11]. M.S. Mythili and A.R.Mohamed Shanavas comparisons on data mining classification methods used to predict learning skills of autistic children. The classification algorithms like, J48 and Support Vector Machine (SVM) algorithms are compared check the efficiency and outperform the SVM [12]. Vitthal Manekar and Kalyani Waghmare comparisons on various data mining algorithms like, decision tree, Naïve Bayes, KNN and SVM and Hybrid Approach of SVM algorithm and the SVM algorithm with evolutionary approach is more Accurate [13]. Parvathi I and Siddharth Rautaray describes about the proposal of hybrid data mining model to extract classification knowledge for various diseases in clinical decision system and presents a framework of the tool various tools used for analysis [14]. Priyanka Juneja and Anshul Anand has been analysis on the ASD patient symptom based a pre-level decision is taken about to recognize the probability of ASD [15]. Mohana E and Poonkuzhali.S describes the result for best suited classifier using to predicting the risk level of ASD [16]. Motaz M. H. Khorshid *et al*. proposed an hybrid machine learning models using to predict terrorist groups responsible of attacks in Middle East and North Africa [18]. Ionu Taranu describes data mining techniques is most beneficial in health care organization [19]. Priyanka Sanjay Podutwar and Prof. Ms. R. R. Tuteja proposed a soft computing technique using to find the accuracy of ASD [20]. Ashmeet Singh and R Sathyaraj describe various Methodologies used to predicting small and large dataset in rapid miner environment [21].

III. PROPOSED SYSTEM

The implementation work is to predict different level of autism (low, middle, high). Diagnosing initial stage of autism is difficult because early stages of symptoms are very common. Autism diagnosed based on multiple symptoms in the historical dataset. Autism symptoms dataset is collected from VAERS and it is used hybrid model to predict ASD [22]. In Figure 1 represent the process flow architecture of this proposed system.

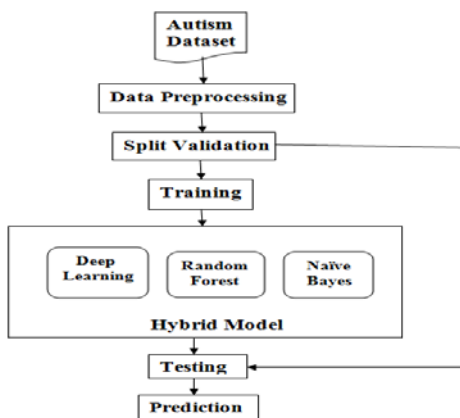


Figure 1: Process Flow Architecture

A. Implementation using Rapid Miner:

Rapid Miner combines tools and applicability to provide user-friendly integration environment of the newest data mining techniques [17]. Rapid Miner is open-source data mining with java software and platform for data science software. It is provides an integrated environment for data preparation, deep learning, machine learning, predictive analysis and text mining. Rapid Miner is developed on open core model. An application of Rapid Miner covers a wide range of real-world data mining task. Rapid Miner tool supports the DRN hybrid algorithm [23].

B. Dataset:

The raw autism symptoms dataset had been downloaded from the government website of “https://vaers.hhs.gov/data/data” [22]. The ASD dataset is a CSV file format and it is contained eight attribute. Each attribute define autism patient details show in table I.

Table I: ASD Dataset Explanation

Field No.	Attributes	Explanation
1	VAERS_ID	It is each data assigned a unique identification number in VAERS website.
2	Gender	It is defined autism patient of the gender.
3	Symptom level 1	It is defined symptom stage level 1
4	Symptom level 2	It is defined symptom stage level 2
5	Symptom level 3	It is defined symptom stage level 3
6	Symptom level 4	It is defined symptom stage level 4
7	Symptom level 5	It is defined symptom stage level 5
8	Autism symptom level	It is label attribute and it is predicting three different levels as low, medium, high.

C. Data Preprocessing:

Data preprocessing is an important steps of this work. The raw dataset cannot be used directly because raw dataset associated with huge amount of unwanted data. Unwanted data is includes Null Values, Redundant Values, and Missing Values. Autism Dataset is removed the unwanted values by the data preprocessing. Preprocessed data will be clear, clean and noiseless.

D. Split Validation:

Split Validation is a nested operation. It has split into two subprocess as training subprocess and testing subprocess. The training subprocess is used for learning or building a hybrid model and trained model is then applied in the testing subprocess. The performance of the hybrid model is also measured during the testing subprocess. Split validation is mainly used to estimate how accurate a hybrid model will perform [23].

E. DRN Hybrid Model:

In this work the proposed hybrid model combined of three models like Deep Learning, Random Forest and

Naïve Bayes. The result obtains from the combining DRN Hybrid model outperform in classification parameters than Accuracy, Recall, Precision, Classification-Error, Kappa Statistic and Executed time when implement individual.

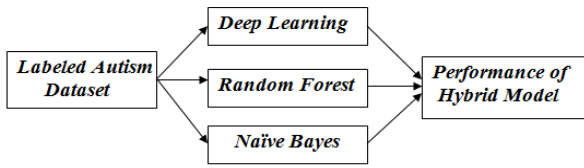


Figure 2: DRN Hybrid Model

In figure 2 is defined the Hybrid Model for DRN. To get the input is labeled autism dataset to apply the DRN hybrid model and the produced output for performance of DRN Model

a. Deep Learning:

Deep Learning is a set of Machine Learning algorithms which have one or more hidden layers of multiple nonlinear processing units between input and output layers. This is used in the learning phase during training process. The deep learning algorithm makes multiple hidden layers in the autism symptoms dataset [23].

b. Random Forest:

Random Forest in Rapid Miner environment generates a set of random trees. In Random Forest to create a classification model that predicts the value of a label attribute. The resulting of Random Forest model contains a specified number of trees in the ASD dataset [23].

c. Naïve Bayes:

A Naïve Bayes is a probabilistic classifier and it is a prediction models based on Baye’s theorem, with focus on independent attributes. Naïve Bayes classifier combines this model with a decision rule. In these model describe the conditional probability of each set of possible causes for a given observed outcome can be computed [23].

The pseudo code for proposed DRN Hybrid model:

Method: new DRN Hybrid model (Deep Learning, Random Forest, Naïve Bayes)

Input: D = Autism Dataset

```

Start
D= Z- Norm(D) /* Z-Normalization of dataset*/
Do
    Split Validation (D) /* Split into training and testing phases*/
    If (Training phase(D))
    
```

```

D= Training (DRN(D)) /* Construct DRN
hybrid model to apply the dataset*/
If (Testing Phase(D))
    Training (DRN(D)) = Predict(D)
Return
    Predict (D) = Label {Low, Middle, High} /*Predict
Labels in the dataset*/
End
Output: Predicting Dataset Label (Low, Middle, High)
    
```

IV. RESULT AND ANALYSIS

The New DRN hybrid model used to predict three different level of autism like low, middle and high. In DRN hybrid model is implemented in the Rapid Miner environment to find the Accuracy, Precision, Recall, Classification Error, Kappa Statistic and Executed Time

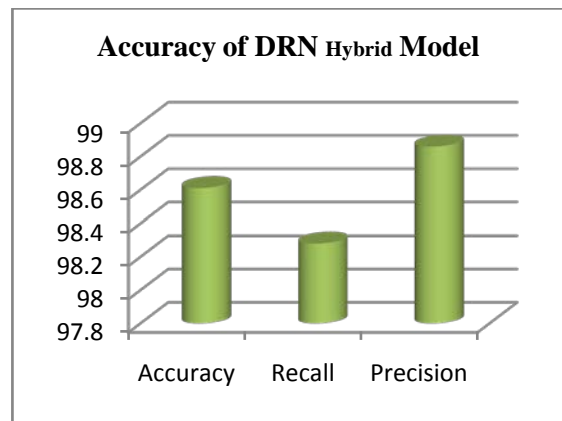


Figure 3: Accuracy of DRN Hybrid Model

Figure 3 gives the accuracy parameters of the proposed DRN hybrid model. The graphical representation for Accuracy, Recall and Precision is shown in Figure 3.

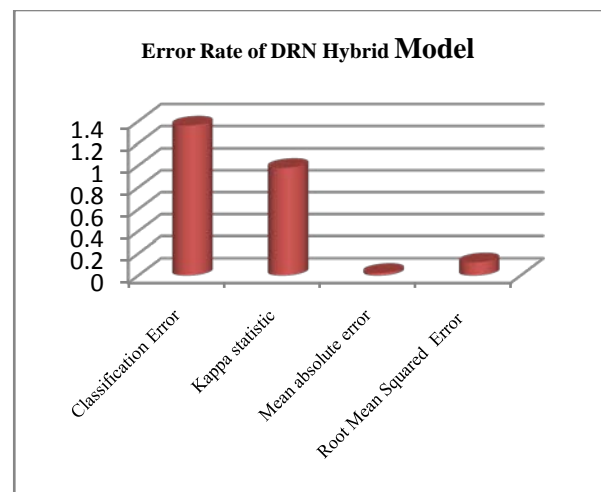


Figure 4: Error-Rate of DRN Hybrid Model

Figure 3 gives the Error-Rate parameters of the proposed DRN hybrid model. The graphical representation for Classification error, Kappa statistic, Mean Absolute and Root Mean Squared Error is show in Figure 4.

DRN- Deep Learning, Random Forest and Naïve Bayes, S- Stacking, V- Vote, AB- Ada Boost, B- Bagging, BB- Bayesian Boosting.

Table II: Performance Result of Hybrid Models

Evaluation Criteria	DRN %	S %	V %	A B %	B %	B B %
Timing to build the model(sec)	0.05	0.10	0.18	0.10	0.10	0.08
Accuracy	98.6	89.5	74.9	47.7	47.7	47.4
Recall	98.2	80.2	65.4	33.3	33.3	33.3
Precision	98.8	78.2	54.8	15.9	15.9	15.9
Classification Error	1.39	10.4	25.1	53.2	53.2	53.5

In Table II the percentage of Accuracy, Recall and Precision is listed for various hybrid Models. DRN Hybrid Model outperform than other models like Stacking, Vote, Ada Boost, Bagging, and Bayesian Boosting.

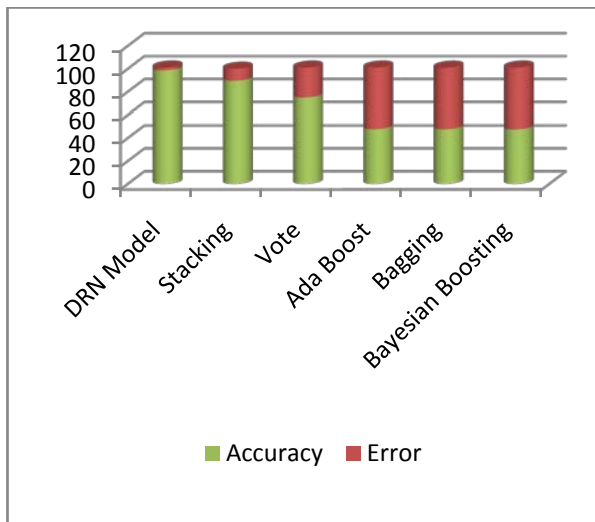


Figure 5: Efficiency of Different Hybrid Models

The Figure 5 gives the efficiency of different Hybrid Models in the graphical representation. The highest percentage of efficiency is seen in DRN hybrid model.

DRN- Deep Learning, Random Forest and Naïve Bayes, S- Stacking, V- Vote, AB- Ada Boost, B- Bagging, BB- Bayesian Boosting.

Table III: Error Measurement for Hybrid Models

Evaluation Criteria	DRN %	S %	V %	A B %	B %	B B %
Classification Error	1.3	10.4	25.0	53.2	53.2	53.5
Kappa statistic	0.97	0.83	0.56	0.0	0.0	0.0
Mean absolute error	0.02	0.53	0.26	0.65	0.61	0.61
Root Mean Squared Error	0.11	0.54	0.37	0.65	0.64	0.64

In table III displays the degree of incorrect classification error like Classification error, kappa Statistic, Mean absolute error and Root mean squared error in the different predictive models. DRN model is gives minimum error when compared to other classifiers.

V. CONCLUSION

In this work, predicting different level of autism using new DRN hybrid model is implemented Rapid Miner environment. The DRN hybrid model is compare to check the accuracy, classification error and executed time with other hybrid models like Ada Boost, Bagging, Vote, Stacking and Bayesian Boosting. The proposed DRN hybrid model is outperforms the other models and provides Higher Accuracy and Lower classification error. Hence the DRN hybrid model is efficient predictive model using the historical dataset for predicting the different level of autism.

VI. REFERENCES

- [1] Michael Siller and Marian Sigman, "The Behaviors of Parents of Children with Autism Predict the Subsequent Development of Their Children's Communication", Journal of Autism and Developmental Disorders, Vol. 32, No. 2, April 2002.
- [2] Tony Charman, Simon Baron-Cohen, John Swettenham, Gillian Baird, Auriol Drew and Antony Cox, "Predicting language outcome in infants with autism and pervasive developmental disorder", International Journal of Language & Communication Disorders, Vol.38, No.3, 2003.
- [3] G. Leroy, A. Irmscher, and M.H. Charlop-Christy, "Data Mining Techniques to Study Therapy Success with Autistic Children", 2006 International Conference on Data Mining, 26 - 29 June 2006, Monte Carlo Resort, Las Vegas, USA.
- [4] S. Wheelwright, S. Baron-Cohen, N. Goldenfeld, J. Delaney, D. Fine, R. Smith, L. Weil and A. Wakabayashi, "Predicting Autism Spectrum Quotient (AQ) from the Systemizing Quotient-Revised (SQ-R) and Empathy Quotient (EQ)", 2006.
- [5] Sheena angra and Sachin ahuja, "Analysis of student's data using rapid miner", Journal of Today's Ideas – Tomorrow's Technologies, Vol. 4, No. 1, June 2016 pp. 49–58.
- [6] Brian P. Keane, Orna Rosenthal, Nicole H. Chun and Ladan Shams, "Audiovisual integration in high functioning adults with autism", Research in Autism Spectrum Disorder for Elsevier journal, Volume. 4, Issue 2, April-June 2010.

- [7] A.Martin, V.Gayathri, G.Saranya, P.Gayathri and Dr.Prasanna Venkatesan, "A Hybrid Model For Bankruptcy Prediction Using Genetic Algorithm, Fuzzy C-Means and Mars", International Journal on Soft Computing (IJSC), Vol.2, No.1, February 2011.
- [8] Tanaya Guha, Member, IEEE, Zhaojun Yang, Student Member, IEEE, Ruth B. Grossman, and Shrikanth S. Narayanan, Fellow, IEEE, "A Computational Study of Expressive Facial Dynamics in Children with Autism", IEEE TRANSACTIONS ON AFFECTIVE COMPUTING, VOL. XX, NO. X, MARCH 2016.
- [9] Prof. Dr. Ahmed Hassan, Assistant Prof. Osama Abdo Mohamed, Prof. Dr. Ahmed Soufi Abou-Taleb, and Mr. Amr Hassan, "A Hybrid Feature Selection Approach Of Ensemble Multiple Filter Methods And Wrapper Method For Improving The Classification Accuracy Of Microarray Data Set", IRACST - International Journal of Computer Science and Information Technology & Security (IJCSITS), ISSN: 2249-9555 Vol. 3, No.2, April 2013.
- [10] Niyati Gupta, Arushi Rawal, Dr. V.L. Narasimhan, and Savita Shiwani, "Accuracy, Sensitivity and Specificity Measurement of Various Classification Techniques on Healthcare Data", IOSR-Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p- ISSN: 2278-8727 Volume 11, Issue 5 (May. - Jun. 2013), PP 70-73.
- [11] E. M. Albornoz, L. D. Vignolo, C. E. Martínez & D. H. Milone, "Genetic Wrapper Approach for Automatic Diagnosis of Speech Disorders related to Autism" 14th IEEE International Symposium on Computational Intelligence and Informatics (CINTI), nov, 2013.
- [12] M.S. Mythili and A.R.Mohamed Shanavas, "A Novel Approach to Predict the Learning Skills of Autistic Children using SVM and Decision Tree", International Journal of Computer Science and Information Technologies, Vol. 5 (6) , 2014, 7288-7291.
- [13] Vitthal Manekar and Kalyani Waghmare, " Improving Accuracy of SVM Using Hybrid Cultural Algorithm", Int.J.Computer Technology & Applications, Vol 5 (3),1194-1197.
- [14] Parvathi I and Siddharth Rautaray, "Survey on Data Mining Techniques for the Diagnosis of Diseases in Medical Domain", International Journal of Computer Science and Information Technologies, Vol. 5 (1), 2014, 838-846.
- [15] Priyanka Juneja and Anshul Anand, "Analyses of Autistic Patients By using Interpretation Value Analysis", International Journal of Computer Science and Mobile Computing, Vol.3 Issue.7, July- 2014, pg. 585-593.
- [16] Mohana E and Poonkuzhali.S, "Categorizing The Risk Level Of Autistic Children Using Data Mining Techniques", International Journal of Advance Research In Science And Engineering IJARSE, Vol. No.4, Special Issue (01), April 2015.
- [17] Priti S. Patel and Dr. S.G. Desai, "A Comparative Study on Data Mining Tools", International Journal of Advanced Trends in Computer Science and Engineering, Vol.4(2), March - April 2.
- [18] Motaz M. H. Khorshid , Tarek H. M. Abou-El-Enien and Ghada M. A. Soliman, "Hybrid Classification Algorithms For Terrorism Prediction In Middle East And North Africa", International Journal of Emerging Trends & Technology in Computer Science, Volume 4, Issue 3, May-June 2015.
- [19] Ionuț Taranu, "Data mining in healthcare: decision making and precision", Database Systems Journal vol. VI, no. 4/2015.
- [20] Priyanka Sanjay Podutwar and Prof. Ms. R. R. Tuteja, "Enhancing technique for Predictive Grading of Childhood Autism using Soft Computing", International Journal of Research In Science & Engineering, Special Issue: Techno-Xtreme 16.
- [21] Ashmeet Singh and R Sathyaraj, "A Comparison Between Classification Algorithms on Different Datasets Methodologies using Rapidminer", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 5, Issue 5, May 2016.
- [22] "Vaccine Adverse Event Reporting System (VAERS), <https://vaers.hhs.gov/data/data>.
- [23] "Rapid Miner tool", <http://rapidminer.com/>