

Prediction of ASD among Children using Machine Learning Techniques

Ishwaria A, Ajaypradeep N, Sasikala Ra

Abstract- Autism spectrum disorder (ASD) is a neurological issue that begins from early in childhood and proceeds all through such an individual's reality. It will influence that individual's conduct, speech with others, interference, and learning. Right now anyway, Autism Spectrum Disorder is to be distinguished in the beginning period, which is conceivable. Early acknowledgment of Autism Spectrum Disorder will improve the general psychological wellness of that particular youngster. The machine learning methodology is applied to diagnose Autism Spectrum Disorder (ASD), and in this work, we have used machine learning techniques and Optimization on an ASD dataset. We have employed XGBoost algorithms to the dataset considered, and as a result, efficient outputs are obtained. This will be incredible for the use of doctors to help them recognize Autism Spectrum Disorder at an early prior stage.

Keywords— ASD, Machine Learning, XGBoost..

I. INTRODUCTION

Autism Spectrum Disorder is a neurons-formative issue portrayed by tenacious shortfalls in social cooperation and correspondence. Individuals affected with autism possess dreary conduct, interest, and activities. ASD prevails in many ways; it can be from mild ASD to a severe form. It depended on hardness of indications. These will be considered that there is a need for earlier recognition of ASD and also stressed the fact that early intervention programs are very beneficial for the overall well-being of a child. Past research works have demonstrated that in children as youthful as the year and a half, ASD can be distinguished dependably by utilizing 2-phase screening systems. Studies likewise expressed that the average age from the start finding is between 3 to 6 years.

This space between possible age at first diagnosis (18 months) and the average age at early diagnosis is a gap that should be closed as much as possible. As stated before, physicians can use specific reliable formal screening instruments to expand the accuracy of the gauge concerning the developmental status of children. However, only a minority of physicians use those tools that are available. One way to improve the accuracy would be routine formative screenings for all children. These kinds of regular checks might prove to be costly. An alternative approach is developing a decision support system, which can be used by physicians to help them estimate the developmental status of children more accurately. It is possible to use structured data

to predict ASD in an earlier stage. Using structured data, we propose a decision support based approach to help physicians improve the accuracy of their estimates regarding the developmental status of children and specifically in the case of ASD. The structured data that is used is gathered from the 2011-2012 National Survey of Children's Health (NCSH), which was conducted in the United States of America. This survey includes children from the age of 2 to 17 across every state in the United States of America and contains answers from essential guardians of these youngsters.

II. LITERATURE REVIEW

The information of grown-up individuals from the age between of 17 to 60 years was taken, which analyses the existence of ASD by applying machine learning algorithms. There was a full scope of questions in the dataset that has been utilized in their exploration. The CART, KNN, LR, SVM, LDA, and Naive Bayes algorithms have been used. They characteristics parameters of the dataset were converted into numerical values. Finally, they stated the result executed using the LDA algorithm, which obtained an excellent outcome of 72.20% accuracy, which was considerably higher than the other algorithms [9]. And In another work, the executions were made utilizing three diverse classification techniques, Naive Bayes, R Random Forest, and KNN on the ASD dataset. In this work, the result obtained by the random forest technique has been demonstrated to be more accurate than naive Bayes and KNN algorithms[13]. In another action, a similar report on the prediction analysis of a few learning classifiers was carried out [14].In work [11], it was explored which model gives the high precision, review rates, and excellent accuracy execution for the given AI issue. This examination closes that the Deep Learning classifier accomplished the most noteworthy characterization execution as far as exactness, review rates, and accuracy. Random Tree classifier is the least performed classifier in all measurements. The work [1] utilized an example of the dataset, and it contained different sorts of inquiries identified with ASD. They have been used five algorithms in classification models like linear discrimination analysis, SVM, KNN, and Decision Tree. It was observed that to improve the accuracy, and they utilized cross-validation techniques to find the best classification method. This Decision tree indicated 100% accuracy in every k value, and it was determined to be the best classifier.

Revised Manuscript Received on May 07, 2020.

Ishwaria A, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India. Email: aishubuvan5511@gmail.com

Ajaypradeep N, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India. Email: ajaypradeep.n@vit.ac.in

Dr.Sasikala Ra, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India., Email: sasikala.ra@vit.ac.in

In work [2], the result obtained through the linear discrimination analysis algorithm gave the best outcome than the K-Nearest Neighbor algorithm in the exactness. The F measure value had determined as 0.90 from the straight separation examination method and 0.89 for the k-Nearest Neighbor algorithm. In straight separation examination in the F-measure value given a 1.95 percentages preferable achievement measure over the K-Nearest Neighbor algorithm. In another work, the performance of the four distinct kinds of SVM kernels was considered utilizing Weka programming. They took four sorts of SVM kernels, which are the normalized polynomial portion, inclusive kernel, polynomial piece, and radial basis work. The kernels had been applied for classifications procedures of the data. The benchmark's sensitivity, F-measure precision, and confusion matrix, accuracy in every piece were calculated. Through this exhibition correlation, the best outcome of precision rate with an accuracy of 100% was obtained.

III. XGBOOST ALGORITHM PRINCIPLE

The XGBoost algorithm is a tree-based algorithm with regularizations which can be added to its optimized objective technique:

Step 1: Initialize the model.

In the initial step, the model should be instated with the functions of $F_0(x)$. $F_0(x)$ ought to be a function that limits the MSE.

$$F_0(x) = \underset{y}{\operatorname{argmin}} \sum_{i=1}^n L(y_i, y)$$

Step 2: Calculate the pseudo-residual.

Based on the learner it will compute the ideal value,

$$\underset{y}{\operatorname{argmin}} \sum_{i=1}^n L(y_i, y) = \underset{y}{\operatorname{argmin}} \sum_{i=1}^n L(y_i - y)^2$$

Finally, it will update in this model:

$$F_m(X) = F_{m-1}(X) + y_m h_m(X)$$

In the main objective of XGBoost algorithm is to analyze the negative edge of the misfortune function in rehashed iterations in advancing the exact loss function followed by utilizing the direct inquiry strategy to create the ideal feeble patient. The XGBoost algorithm actualizes the patient by enhancing the organized misfortune work, and it doesn't utilize the straight pursuit strategy, the algorithm uses the principal subsidiary and the second derivative of the misfortune function. The presentation of algorithm is improved by weighted quintile, pre-arranging, scanty network identification.

Step 3: XGBoost loss function

In this regularization, two coefficients have been controlled the complexity and given the output to the model.

$$\begin{aligned} \psi(y, F(X)) &= \sum_{i=1}^N \psi(y_i, F(x_i)) + \sum_{m=0}^T \Omega(f_m) \\ &= \sum_{i=1}^N \psi(y_i, F(x_i)) + \sum_{i=1}^T \gamma L_m + \frac{1}{2} \|\omega_m\|^2 \end{aligned}$$

In the point,

$$F_{m-1} = F_{m-1}(X)$$

XGBoost have been taken the Taylor approximation to the loss function. It will be prepared the moderate learners by limiting the loss function

$$\begin{aligned} (\psi_m) &= \sum_{i=1}^N (y_i, F(x_i) + f_m(x_i) + \Omega(f_m)) \\ &= \sum_{i=1}^N \psi(y_i, F_{m-1}(x_i)) + g_i f_m(x_i) + \frac{1}{2} h_i f_m(x_i) + \Omega(f_m) \end{aligned}$$

Step 4: Determining the optimal output value of each leaf node.

The tree has been created in the mth iteration with L leaf nodes ($l_1, l_2, l_3 \dots l_m$). It was the example Index number that falls on jth leaf nodes. At that point, q is the result of m'th spanning-tree, ie..,

$$l_j = \{i | q(x_i = j)\}$$

Which mapped test x into the relating leaf node. In this advancement capacity can be expressed,

$$= \sum_{i=1}^N [(\sum_{i \in l_j} g_i) \omega_j + \frac{1}{2} (\alpha + \sum_{i \in l_j} h_i) \omega_j^2]$$

The solution is

$$= \sum_{j=1}^L f(\omega_j) + \gamma L$$

This solution was achieved by applying the minimization loss of values in the leaf nodes.

$$\psi_m(q) = -\frac{1}{2} \sum_{j=1}^L \frac{(\sum_{i \in l_j} g_i)^2}{\lambda + \sum_{i \in l_j} h_i} + \gamma L$$

In this above results, it can minimize the loss in the tree from the mth generation model.

$$F_m(X) = F_{m-1}(X) + f_m(X)$$

The optimal o/p values of each node have been determining by the 1st derivatives g and 2nd derivatives h of the loss functions.

Step 5: Integration of parting conditions

Assume that beginning root node list set of all examples in the nodes which are represented to by (I) and the index set of the splits left and right child tests are represented to by (I), and there was decreased of m'th tree loss functions after parting.

$$\Delta \psi = \frac{1}{2} \left[\frac{(\sum_{i \in l_l} g_i)^2}{\lambda + \sum_{i \in l_l} h_i} + \frac{(\sum_{i \in l_r} g_i)^2}{\lambda + \sum_{i \in l_r} h_i} - \frac{(\sum_{i \in I} g_i)^2}{\lambda + \sum_{i \in I} h_i} \right] - \gamma$$

From the value equation, the separating point and the separating attributes are optimized in the separating of each node. This is the optimal m'th iteration. It has the T+1 weak learner obtained after T iteration, and T+1 models were added to integrate weak learners. XGBoost uses the shrinkage method to reduce overfitting and the model integration

$$F_m(X) = F_{m-1}(X) + \eta f_m(X), 0 < \eta \leq 1$$

Shrinkage is used to reduce overfitting in XGBoost tectonic tree model, the method of randomly selecting a certain number of feature subsets in a random forest is also used in determining the optimal splitting point, and the overfitting can also be suppressed.

The larger feature subset in smaller deviation of each weak learner, however, the larger invariance. We have to utilize cross-validation to weight decisions in a particular proportion.

IV. EXPERIMENTAL METHOD

A. Experimental datasets:

This experiment utilized standard datasets that can be freely downloaded from UCI repository that identified with autism screening of toddlers, which was used for investigation, particularly in deciding medically introverted attributes and improving the order of ASD cases. This dataset has a record of ten behavioral features (Q-Chat-10), useful in identifying the ASD cases from controls in behavior science.

B. Proposed Architecture

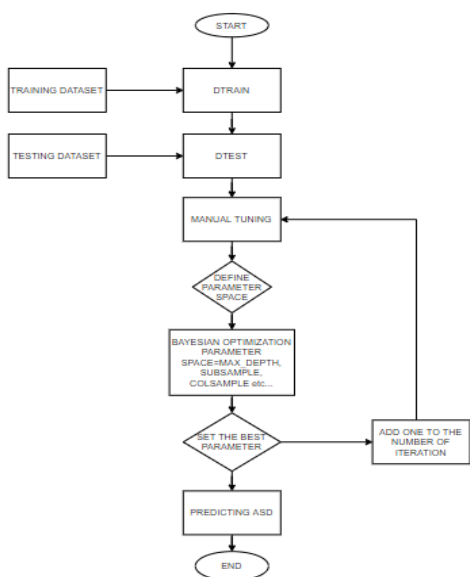


Fig:1- Flowchart of proposed method

C. Methodology:

In this study, we utilized XGBoost with Hyper-parameter tuning. For the hyper-parameter of XGBoost, it is tuned with Bayesian Optimization. It used to set up the model with selected parameters. Bayesian Optimization is utilized with the Bayesian theorem with Gaussian distribution to discover the probability of the given value in the given condition. These Optimization strategies have both hyper-parameter techniques, and standard classifiers are considered as a kind of perspective that focuses on the investigation. In these Hyper-parameter tuning, Bayesian Optimization is the best strategy when contrasted with Grid search or Random search. At last, the model will show signs of improvement with compelling parameters.

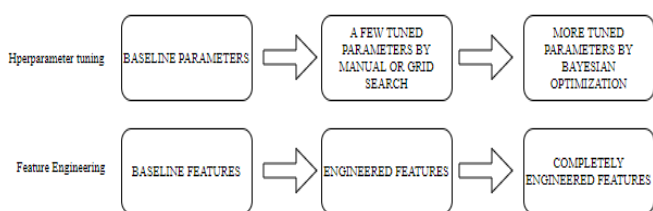


Fig:2- Block Diagram of Hyperparameters.

D. Bayesian Optimization:

Bayesian Optimization is a way to deal with enhancing target works that set aside an extended effort to assess. It is most appropriate for advancement over persistent spaces of under measurements and endures stochastic clamor in work assessments. It manufactures a surrogate for the goal and evaluates the vulnerability in that surrogate utilizing a Bayesian machine learning technique Gaussian process regression and afterward uses an acquisition function characterized by this surrogate to choose where to test.

In this building, a probabilistic model for the performance metric $f_{\mu}(\theta)$. This introduces computational overhead. As a result, applying Bayesian Optimization only really make sense if,

- The number of hyperparameters is very high.
- It is computationally costly to evaluate Optimization only really make sense if
- The number of hyperparameters is very high.
- It is computationally costly to evaluate $f_{\mu}(\theta)$ for a single point θ .
- It has three steps:
- Update the current belief of the loss surface of the learner f .
- Choose that maximizes some utility over the current belief.
- Evaluate the performance of learner f with parameter θ .

E. Gaussian process:

The performance function f_{μ} is modeled using Gaussian Process(GPs).

- GPs are a generalization of a Gaussian distribution to distribution over functions, instead of random variables.
- Just as a Gaussian distribution is determined by its mean function and its variance function.
- A GPs is a function that, as instead to restoring a scalar $f(x)$ it resets the mean and variance of an ordinary distribution over the potential values of f at x .
- GP is the Gaussian distribution over function instead of random variables.

To find the best values, it uses the Acquisition function.

Acquisition function: It is used to formalize what constitutes a "best guess."

$$EI(\theta) = E \left\{ \max_{\theta} [0, f_{\mu}(\theta) - f_{\mu}(\theta)] \right\}$$

$$(\theta)_{NEW} = \text{ARGMAX } EI(\theta)$$

$$EI(\theta) = \begin{cases} (\mu(\theta) - f(\theta))\phi(z) + \delta(\theta)\phi(z) & \delta(\theta) > 0 \\ 0, & \delta(\theta) = 0 \end{cases}$$

$$z = \frac{\mu(\theta) - f(\theta)}{\delta(\theta)}$$

F. K-Fold Cross Validation:

K-Fold cross-validation technique was the simulation of information that will be arbitrarily partitioned into training as 80% and testing as



20%.

This is utilized because of to locate the best parameters of the model. Right now, overlap cross-validation is used, and the preparation will part of ten pieces of the same size. The nine sections are utilized as training, and one will be as testing so that this procedure will be at the same time rehased into ten cycles. Lastly, the average will be taken as this precision accuracy to form scores.

G. Analysis of prediction results:

To look at the prediction impacts of the XGBoost model dependent on ASD, this work utilizes Random forest and Linear Discriminant Analysis as the examination algorithm to look at the XGBoost dependent on ASD. From the data, each of the three algorithms was utilized in the entire expectation process, and it can be seen that the RMSE error of XGBoost with Bayesian Optimization. The RMSE is used, and to break down the scattering of anticipated qualities among the three algorithms of the RMSE numerical worth depends on the ASD XGBoost. In Fig:1, XGBoost has less RMSE compare to another ML algorithm.

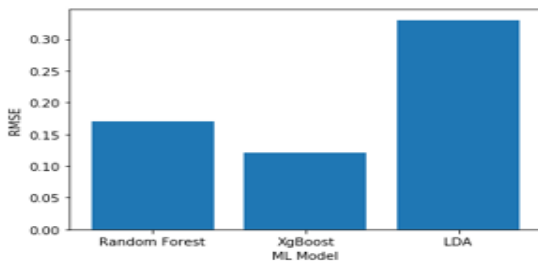


Fig-3- Comparing with other ML Models

V. CONCLUSION

In this work, XGBoost with Bayesian optimization has been used for Predicting ASD among children. The RMSE of the XGBoost model is 0.12, and the error rate is much lower than the 0.17 of the random forest and 0.33 of the LDA. This method was effectively validated to significantly improve the prediction accuracy. In more advanced techniques, including running multiple function evaluations in parallel, multi-fidelity and multi-information source optimization, expensive-to-evaluate constraints, random environmental conditions, multi-task Bayesian optimization, and the inclusion of derivative information. Since it is hard to diagnose autism at an early stage, the discussed methodologies can be a good solution to identify and diagnose autism in an early stage. When it comes to real time diagnosis, the discussed methods will help a large scale of kids, by early intervention of the disease.

REFERENCES

1. Lawi, A., & Aziz, F. (2018). Comparison of Classifications Algorithms of the Autism Spectrum Disorder Diagnosis. 2018 2nd East Indonesia Conference on Computer and Information Technology
2. Altay, O & Ulas M, Prediction of the autism spectrum disorder diagnosis with linear discriminant analysis classifier and K-nearest neighbor in children.6th International Symposium on Digital Forensic and Security (ISDFS)2018.
3. TALABANI, H., & AVCI, E. (2018). Performance Comparison of SVM Kernel Types on Child Autism Disease Database. (2018) International Conference on Artificial Intelligence and Data Processing.

4. Che, M., Wang, L., Huang, L., & Jiang, Z. (2019). An Approach for Severity Prediction of Autism Using Machine Learning, International Conference on Industrial Engineering and Engineering Management.
5. Akter, T., Satu, M. S., Khan, M. I., Ali, M. H., Uddin, S., Lio, P., Moni, M. A. (2019). Machine Learning-Based Models for Early Stage Detection of Autism Spectrum Disorders.
6. Souza P. V. de C., & Guimaraes A. J. Using fuzzy neural networks for improving the prediction of children with autism through mobile devices IEEE Symposium on Computers and Communications (ISCC) 2018.
7. Lu, H., & Liao, D (2018). Classify autism and control based on deep learning and community structure on resting-state fMRI—2018 Tenth International Conference on Advanced Computational Intelligence.
8. Chen, S., Wu, J., & Chen, X. (2019). Deep Reinforcement Learning with Model-Based Acceleration for Hyperparameter Optimization. IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI).
9. Buyukoflaz, F, Ozturk A (2018). Early autism diagnosis of children with machine learning algorithms,26th Signal Processing, and Communications Applications Conference.
10. Halibas, A. S., Reazol, L. B., Delvo, E. G. T., & Tibudan, J. C. (2018). Performance Analysis of Machine Learning Classifiers for ASD Screening, International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT).
11. Kundu, M. R., & Suranjan Das, M. (2019). Predicting Autism Spectrum Disorder in Infants Using Machine Learning, Journal of Physics.
12. AVCI E & TALABANI H, (2018). Performance Comparison of SVM Kernel Types on Child Autism Disease Database. International Conference on Artificial Intelligence and Data Processing(2018).
13. Tyagi, B., Mishra, R., & Bajpai, N. (2018). Machine Learning Techniques to Predict Autism Spectrum Disorder.

AUTHORS PROFILE



Ishwaria A received the B.E., Degree from Periyar Maniammai Institute of Science and Technology. Currently Studying M.Tech., Computer science and Engineering in VIT University, TamilNadu, India.



Ajaypradeep.N received his B.E., Degree from Ganathipathy Tulsis Jain Engineering College, Tamilnadu, India in the year 2014 and M.E., Degree from Global Institute of Engineering and Technology, Tamilnadu, India in the year 2016. He is currently a research scholar under School of Computer Science and Engineering in VIT University, Tamilnadu, India. His areas of interest include Machine Learning, Data Analytics, Mobile Computing and Deep Learning.



Dr.Sasikala.R received the B.E. degree from Kongu Engineering College, Tamilnadu, India in 1994, the M.E degree from Govt. College of Technology, Coimbatore, Tamilnadu in 2003, and the Ph.D. degree from the Anna University, Chennai, Tamilnadu in 2011. She is currently an Associate Professor, Department of Computational Intelligence in School of Computer Science and Engineering at VIT University, Vellore, Tamilnadu, India. Her research interests include: Cloud Computing, Big Data, Wireless Networks, Body Sensor Networks and Optical Switching Networks. She has published many research papers in various conferences and reputed journals. She is a life member of Computer Society of India and Indian Society for Technical Education.