Support Vector Machine for Classification of Autism Spectrum Disorder based on Abnormal Structure of Corpus Callosum

Jebapriya S[1], Shibin David[2], Jaspher W Kathrine[3], Naveen Sundar[4]
Department of Computer Science and Engineering
Karunya Institute of Technology and Sciences, Coimbatore, India

***Abstract*—Autism Spectrum Disorders (ASD) is quite difficult to diagnose using traditional methods. Early prediction of Autism Spectrum Disorders enhances the in general psychological well- being of the child. These days, the research on Autism Spectrum Disorder is performed at a very high pace than earlier days due to increased rate of ASD affected people. One possible way of diagnosing ASD is through behavioral changes of children at the early ages. Structural imaging ponders point to disturbances in various mind regions, yet the exact neuro-anatomical nature of these interruptions stays misty. Portrayal of cerebrum structural contrasts in children with ASD is basic for advancement of biomarkers that may in the long run be utilized to enhance analysis and screen reaction to treatment. In this examination we use machine figuring out how to decide a lot of conditions that together end up being prescient of Autism Spectrum Disorder. This will be of an extraordinary use to doctors, making a difference in identifying Autism Spectrum Disorder at a lot prior organize.**

Keywords*—Autism Spectrum Disorder (ASD); ASD screening data; ABIDE; machine learning

I. Introduction

Autism Spectrum Disorder (ASD) is a neuro-developmental issue that can be described by distinct issues that may arise during social collaboration, correspondence and conduct. There are numerous other mental disorders like Autistic disorder and Asperger's disorder which have similar classification of symptoms. Depending upon the severity of symptoms, it prevails in numerous forms from extremely gentle ASD to exceptionally serious ASD. ASD must be analysed immediately so as to follow the advancement of the youngster and give legitimate treatment. An underlying screening is done in order to check the development of the child. ASD cannot be analysed by utilizing a solitary screening. A second screening test must be performed when the child is 2-3 years of age. It very well may be precisely analysed after the second screening test process. In the on-going years, researchers are working on it to anticipate it precisely within 18 months of the child. There are a lot of formal screening tools accessible for doctors to expand the exactness for assessing the formative status of the children. Anyway, just a significant number of doctors utilize those accessible tools. And furthermore, this needs a periodical screening way to track the development status of the child.

To address this issue, computer-aided learning for individuals with mental imbalance was created [1], [2]. Furthermore it was concluded that it would be progressively valuable if this would assist in identifying three noteworthy territories like social and relational abilities, unbending nature of reasoning and relational abilities. Despite the fact that it is hard to get a correct number of Autism cases and it is generally recognized that the predominance has been expanding in the course of recent years. There are signs that some ASD range issue might be ascribed to a mix of certain hereditary susceptibilities; for example, introduction to mercury at basic formative stages and diminished capacity to discharge mercury [3]. There are a few investigations that have appeared even environmental toxicity may play a major role in Autism. Analysts are yet dealing with these clutters to comprehend the shrouded manifestations and subtleties of Autism.

Autism spectrum disorders (ASD) is a disability in the socio - development which was first identified by Kanner. Kanner characterized ASD by the features of difficulties of social behaviours, limited repetitive interests, difficulties in interaction and behaviours [4]. In [5], it is stated that ASD is a disability related to development which progresses from childhood to adult and will exist as long as they live. Some of the common distinctive features of ASD are lack in language and communication skills, inadequacy in social interaction, show of inappropriate behaviours, etc.

In DSM 5, the grouping of pervasive developmental disorders as specified by DSM-IV-TR is modified by placing autism spectrum along with social, communication and restricted or repetitive behaviours [6]. In DSM-IV-TR criteria, pervasive developmental disorders are classified as five separate disorders such as, autistic disorder, childhood disintegrative disorder, Rett's disorder, Asperger's disorder and pervasive developmental disorder. Children diagnosed with ASD has been increasing as the years are progressing [7].

A. Problem Statement

Machine learning has been utilized to anticipate instances of youngster misuse utilizing organized information and literary data. This has nonetheless, not been done frequently and scarcely ever been finished for ASD and building up a choice emotionally supportive network that helps doctors with the recognition of ASD, has hardly been done, which is demonstrated by the absence of writing on the utilization of it.

II. Research Methods

A. Literature Survey

The literature review mainly focused on how the customary techniques were utilized to anticipate Autism Spectrum Disorder and what was the outcome obtained from the analysis. The study was finished utilizing two sorts of datasets M-Chat which is utilized for little children and rs-fMRI which is utilized for all age gatherings [8, 9].

a) Identifying the risk caused by ASD: The seriousness of ASD was resolved utilizing Modified Checklist for Autism in toddlers, otherwise called the M-CHAT. This is a screening apparatus which comprises of inquiries that must be replied by the guardians. The objective is to find the preeminent classifier for an autism dataset through feature relevance analysis. The classification algorithm is also used for predicting the threat level of autism [10, 11]. Among various classification algorithm applied, for example, BVM, CVM and MLR created high exactness of 95.21% utilizing Runs Filtering and it likewise precisely grouped the test dataset. This is valuable for age group of 16-30 years yet the forecast was not exact in all cases.

b) Auxiliary Imaging using Voxel-based Morphometry: The examination demonstrated that utilizing a Multivariate Pattern Analysis(MPA) and voxel-based morphometry (VBM) to order structural magnetic resonance imaging data obtained from 24 children and young people with mental imbalance IQ coordinated neurotypical participants was pertinent just to small data sets [12]. Multivariate Pattern Analysis (MPA), which is a pattern recognition technique that is solely based on machine-learning, can be used to group information by isolating between at any rate two classes. In MPA, the groups are distinguished with about 90% of accuracy based on gray matter in the medial prefrontal cortex, posterior cingulate cortex (PCC), and bilateral medial temporal lobes which are all regions within the default mode network (DMN) [13].

c) Machine Learning Techniques: There are many machine learning algorithms that can be used for classification. The three popular classification algorithms used are Random Forest, Naive Bayes and Support Vector machines .Apart from these algorithms we can use the java implementation of the C4.5 algorithm known as the J48 algorithm [14]. Using these assorted algorithms makes sure that the outcomes are highly reliable and it additionally helps us in finding whether the algorithm is usable for not for the classification task. When these algorithms were applied to the dataset that was isolated into two classes-ASD or No ASD, the results are as in Table I.

Be that as it may, utilizing the extra data we can discover how seriously the individual is influenced with ASD. This is finished by detaching into four groups. They are No ASD, Mild ASD, Moderate ASD and Severe ASD [14]. On utilizing similar characteristics and the previously mentioned machine learning algorithms the outcome got are as in Table II.

So as to build this we apply the 1-away strategy to the J48 calculation as it is the best one. Thus, the exactness is expanded from 54.1% to 90.2%.

TABLE. I. Results Obtained using Two Classes

Algorithm Applied	Accuracy	Precision	Recall
Naïve Bayes	0.865	0.866	0.865
Support Vector Machine	0.833	0.835	0.833
J48 (Decision tree)	0.871	0.871	0.871
Random Forest	0.851	0.854	0.851

TABLE. II. Results Obtained using four Classes

Algorithm Applied	Accuracy	Precision	Recall
Naïve Bayes	0.512	0.479	0.512
Support Vector Machine	0.493	0.475	0.493
J48 (Decision tree)	0.541	0.524	0.541
Random Forest	0.507	0.489	0.507

d) Using deep learning algorithms and Resting state functional magnetic resonance imaging (rs-fMRI): The classification of brain imaging data is stricter when using deep learning algorithms than using supervised methods. On utilizing significant neural framework a mean game plan exactness of 70% and a precision somewhere in the range of 66% and 71% in individual folds was acquired [15,16]. An expansion of 5% in classification accuracy was acquired while using deep learning classification method instead of Support Vector Machine [17]. In spite of the fact that the ABIDE dataset contains sensitive varieties, the deep learning methods envelop such assortments and yield better outcomes over machine learning algorithms. The neural patterns obtained from the classification show an anti-correlation of brain function between posterior and anterior areas of the brain or cerebrum work among foremost and back regions of the brain.

B. Data Exploration

The Modified Autism Checklist in Toddlers (M-CHAT) is a validated developmental screening tool for children aged 16 to 30 months. It is intended to identify children who may benefit from a more thorough evaluation of development and autism. This helps to find the best autism dataset classifier by analysing feature relevance and classification algorithm. Among the various classification algorithms used, algorithms such asBVM, CVMand MLR which produced an accuracy of 95.21 % using Runs Filtering method [18]. This method accurately classified the test dataset.

The Autism Brain Imaging Data Exchange (ABIDE) initiative has totalled practical and auxiliary brain imaging information gathered from research facilities around the globe to quicken our comprehension of the neural bases of autism. Every gathering was made through the accumulation of datasets freely gathered crosswise over in excess of 24 global brain imaging research centers and are being made accessible to examiners all through the world, predictable with open science standards. Since the data set contains more than 1112 records, before the data can be used for our machine learning process, the data must be cleaned. The data will be explored after cleaning the ABIDE data to determine if co-occurring conditions are present in the data. If there are any clusters that could help us predict ASD, it will also be investigated [19]. Support Vector Machine algorithm will be used on the data set, to determine if there are attributes that seem strongly

correlated to ASD. In this algorithm, each data item is plotted as a point in n-dimensional space with each feature being a coordinate value. 'n' denotes the number of features selected for identification [20].

III. Proposed Work

A survey was done on various methods of predicting Autism Spectrum Disorder (ASD) using different machine learning algorithms. Based on the survey a conclusion is drawn that the algorithm Support Vector machine (SVM) can be used for identifying patterns from autistic brain images [21]. Here we have used the region called "Corpus Callosum" to identify the differences between the autistic and non-autistic brain images. The Corpus Callosum is a fibre bundle which connects the left and right hemisphere of the brain. Using SVM algorithm, there were striking differences noted in that particular region where the thickness of Corpus Callosum was either too thick or thin compared to those brain images without ASD [21], [22]. Furthermore it is found that the autistic brain had a decreased white matter volume and larger ventricles. Using these observations will ensure our result outcomes are more reliable with higher accuracy rates. The proposed architecture is shown in Fig. 1.

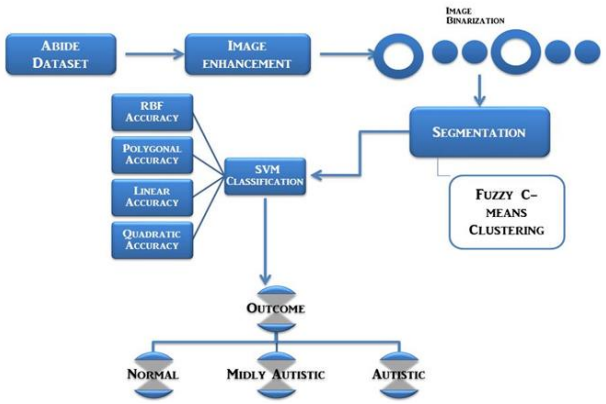


Fig. 1. Architecture of the Proposed Work.

A. Data Sets

The ABIDE dataset contains 1112 records. This includes data from 539 individuals with ASD and data from 573 typical controls (ages 7-64 years). The Data Processing Assistant for Resting-State fMRI (DPARSF), the Configurable Pipeline for the Analysis of Connectomes (CPAC), the Neuro imaging and the Connectome Computation System (CCS) are used for Functional pre-processing. This large pre-processed data is used as an input for the classification algorithm.

IV. Results

Below, the results from machine learning are described. We divide our data set into three different classes, namely Normal, Mildly Autistic and Autistic. A total of 100 iterations are performed to classify the individuals falling under the three classes.

A. Machine Learning Results using Three Classes

We applied the Support Vector Machine classification algorithm to the data set when it was divided into three classes, either the individual was Normal, or the individual was Mildly Autistic or Autistic. This was performed by using the attributes resolved in the data set step mentioned above. We have used the MATLAB software for accuracy calculation.

B. Enhancement of Image

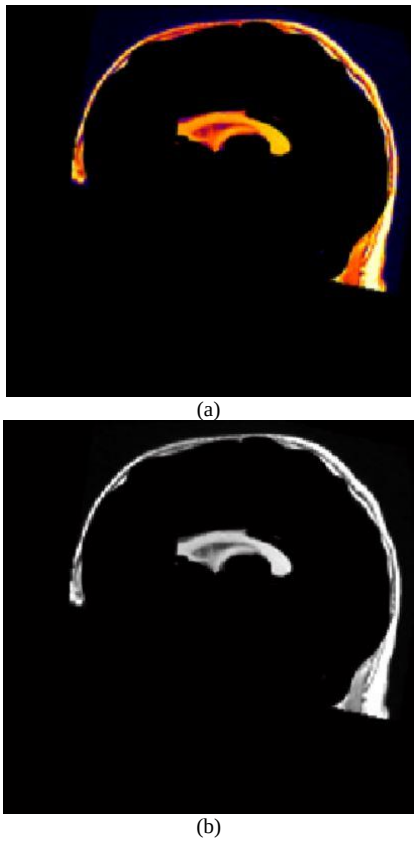
Image enhancement is a process that is widely utilized in numerous image processing applications, to amplify the quality of images. In MATLAB software, the image is enhanced by converting the image into gray scale. The difference is seen in the below given images (Fig. 2(a), Fig. 2(b)).

C. Image Binarization

Image binarization is the process of converting a pixel image into a binary image. Here, we use two main functions, one to normalize the gray scale image by defining a threshold value and the other to convert the indexed image to black and white intensity. The binary image after thresholding is shown in Fig. 3.

D. Segmentation

Image segmentation is the process of splitting images into multiple fragments. This division into fragments is mostly based on the characteristics of the pixels of the image. We have applied fuzzy c-means (FCM) clustering to produce one or more clusters of the given binary image. After FCM segmentation, the following clusters (Fig. 5) are generated out of which one is chosen.



(a)

(b)

Fig. 2. (a) MRI Scan of An individual before Enhancement. (b) MRI Scan of an Individual after Enhancement.

Fig. 3. Binary Image of an MRI Scan.

E. Accuracy Evaluation

A support vector machine (SVM) is a supervised learning algorithm that creates a hyper plane in between data sets to classify the belongingness of the data to its appropriate class. The maximum accuracy of different types of SVM with 100 iterations to classify the images into the aforementioned three classes has been evaluated. The confusion matrix generated is given in Fig. 4.

The accuracy in percentage for different types of SVM is mentioned in Table III and Fig. 6.

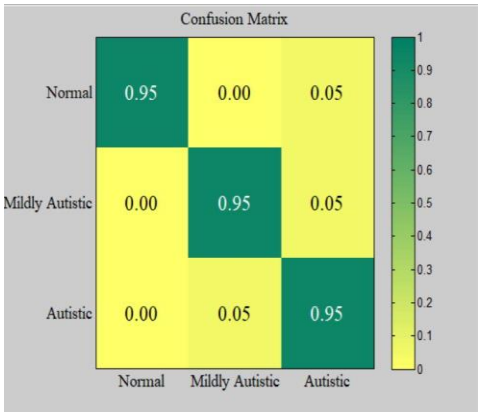


Fig. 4. Confusion Matrix for Classifier using three Classes.

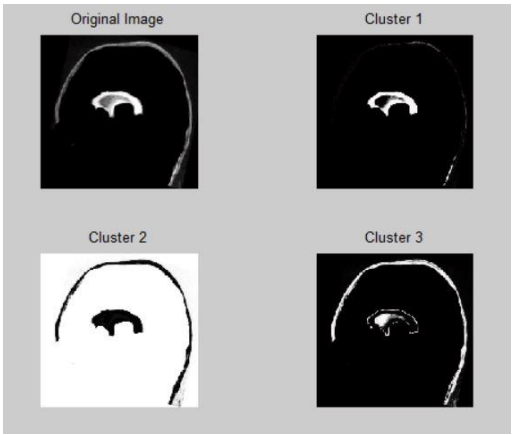
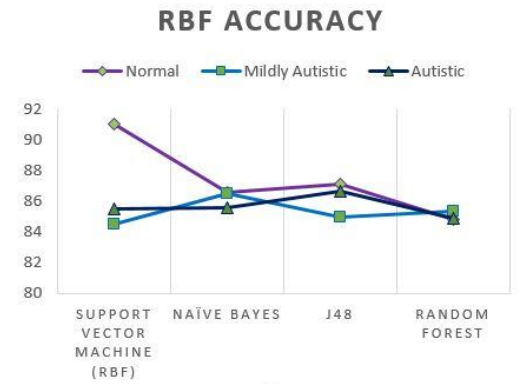
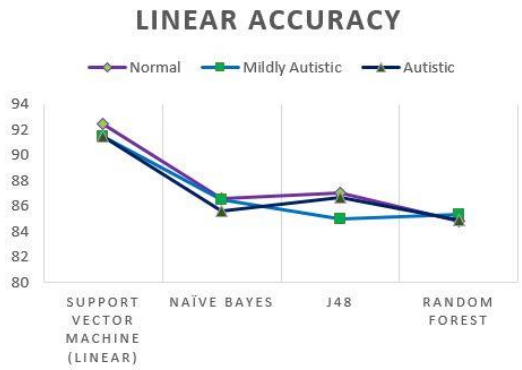


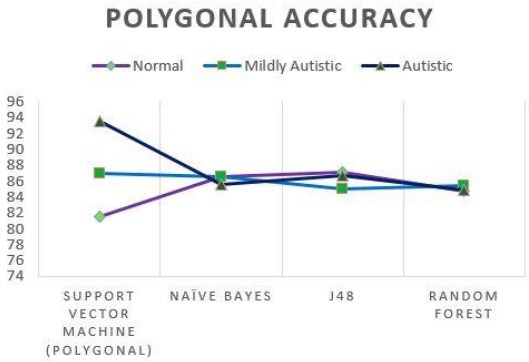
Fig. 5. Clusters Generated after FCM Segmentation.



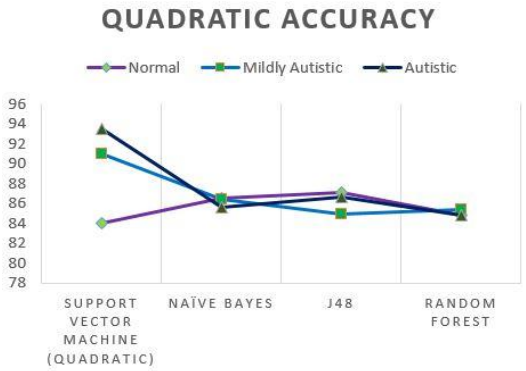
(a)



(b)



(c)



(d)

Fig. 6. (a) Rapid Basis Function (RBF) Accuracy versus Regular Machine Learning Techniques, (b) Linear Accuracy versus Regular Machine Learning Techniques. (c) Polygonal Accuracy versus Regular Machine Learning Techniques. (d) Quadratic Accuracy versus Regular Machine Learning Techniques.

TABLE. III. MAXIMUM ACCURACY ACQUIRED FROM VARIOUS TYPES OF SVM

Types of SVM	Normal	Mildly Autistic	Autistic
RBF Accuracy	91%	84.5%	85.5%
Linear Accuracy	92.5%	91.5%	91.5%
Polygonal Accuracy	81.5%	87%	93.5%
Quadratic Accuracy	84%	91%	93.5%

V. DISCUSSION AND CONCLUSION

The objective was to identify the conditions that demonstrate to be prescient of ASD. This data can be utilized by physicians to enable them to confirm a complete formal screening for ASD. Complex arrange parameters were utilized to plan and analyze discriminate examination along with bolster vector group of classifiers with a most extreme reachable exactness of 94.7% utilizing four highlights and a second request polynomial bit in SVM. The investigation has endeavored to characterize the chemical imbalance range scatter and creating subjects utilizing administered learning systems as depicted in Figure 6. For future work, the focus is towards the investigation of likelihood by utilizing profound learning approaches for the programmed acknowledgment of SMM practices inside and crosswise over subjects.

REFERENCES

[1] Kathleen M. Carroll, Bruce J. Rounsaville, “Computer-assisted Therapy in Psychiatry: Be Brave—It’s a New World”, Current Psychiatry Reports, 2010 Oct; 12(5): 426–432.

[2] Jorn Moock, “Support from the Internet for Individuals with Mental Disorders: Advantages and Disadvantages of e-Mental Health Service Delivery”, Frontiers in Public Health. 2014; 2: 65.

[3] Amy E. Kalkbrenner, Rebecca J. Schmidt, Annie C. Penlesky, “Environmental Chemical Exposures and Autism Spectrum Disorders: A Review of the Epidemiological Evidence”, Current Problems in Pediatric and Adolescent Health Care. 2014 Nov; 44(10): 277–318.

[4] Rachel Cooper, “Diagnostic and Statistical Manual of Mental Disorders (DSM), Encyclopedia of Knowledge Organization, ISKO, available at: https://www.isko.org/cyclo/dsm.htm

[5] Vihang N. Vahia, “Diagnostic and statistical manual of mental disorders 5: A quick glance”, Indian Journal of Psychiatry. 2013 Jul-Sep; 55(3): 220–223.

[6] Murat Gök, “A novel machine learning model to predict autism spectrum disorders risk gene”, Neural Computing and Applications, 2018, pp. 1-7.

[7] Centers for disease control and prevention (CDC), available at: https://www.cdc.gov/ncbddd/autism/data.htm

[8] Anibal Slon Heinsfeld , Alexandre Rosa Francop, R. Cameron Craddockt,, Augusto Buchweitzp,, Felipe Meneguzzio “Identification of Autism Spectrum Disorder using Deep Learning and the ABIDE Dataset“,2017.

[9] Felix D.C.C Beacher ,Eugenia Radulescu ,Ludovico Minati, Simon Baron-Cohen, Michael V.Lombardo, Meng-Chuan Lal, Anne Walker, Dawn Howard, Marcus A.Gray ,Neil A. Harrison, Hugo D. Critchley “Sex Differences and Autism: Brain Function during Verbal Fluency and Mental Rotation”,2012.

[10] Lucina Q. Uddin, Vino Menon, Christina B. Young, Srikanth Ryali, Tianwen Chen, Amirah Khouzam, Nancy J. Minshew, and Antonio Y. Hardan; “Multivariate searchlight classification of structural MRI in children and adolescents with autism”,2017.

[11] LamyaaSadouk ,TaoufiqGadi,andEl Hassan Essoufi; “A Novel Deep Learning Approach for Recognizing Stereotypical Motor Movements within and across Subjects on the Autism Spectrum Disorder”,2018

[12] M. S. Mythili and A. R. Mohamed Shanavas; “An improved autism predictive mechanism among children using fuzzy cognitive map and feature extraction methods (feast)”,2016

[13] Christine Ecker, Susan Y Bookheimer, Declan G M Murphy; “Neuroimaging in autism spectrum disorder: brain structure and function across the lifespan”,2015.

[14] Bram van den Bekerom “Using Machine Learning for Detection of Autism Spectrum Disorder”,2017.

[15] R.Geetha Ramani, K.Sivaselvi “Autism Spectrum Disorder Identification Using Data Mining Techniques”,2017.

[16] M. S. Mythili, A. R. Mohamed Shanavas; “A Study on Autism Spectrum Disorders using Classification Techniques “, 2014.

[17] Tabtah, F. “Autism Spectrum Disorder Screening: Machine Learning Adaptation and DSM-5 Fulfillment.” Proceedings of the 1st International Conference on Medical and Health Informatics ,pp.1-6. Taichung City, Taiwan, ACM.,2017.

[18] Thabtah, F. “Machine Learning in Autistic Spectrum Disorder Behavioural Research: A Review.” To Appear in Informatics for Health and Social Care Journal. December, 2017 (in press),2017.

[19] Thabtah F, Kamalov F., Rajab K “A new computational intelligence approach to detect autistic features for autism screening.” International Journal of Medical Infromatics, Volume 117,2018.

[20] Christina Schweikert, Yanjun Li, David Dayya, David Yens, Martin Torrents, D. Frank Hsu “Analysis of Autism Prevalence and Neurotoxins Using Combinatorial Fusion and Association Rule Mining”, 2009.

[21] Koyamada S, Shikauchi Y, Nakae K, Koyama M, Ishii S (2015) “Deep learning of fMRI big data: a novel approach to subject-transfer decoding”, 2015.

[22] Plis SM et al. (2014) Deep learning for neuroimaging: a validation study frontiers in neuroscience 8,2014.