A REVIEW OF MACHINE LEARNING TECHNIQUES FOR FEATURE BASED CLASSIFICATION OF AUTISM SPECTRUM DISORDER

MANVI VERMA, DINESH KUMAR and TANVI ARORA

Sai College of Engineering and Technology
Badani, Pathankot, Punjab, India

CGC-College of Engineering
Landran, Mohali, Punjab, India
E-mail: tanviverma@rediffmail.com

Abstract

Autism spectrum disorder (ASD) is a type of neurological problem that causes difficulties in social communication and language problems and is associated with typical behavior problems. It can be diagnosed at any age. It is termed as a development order; typically the symptoms are visible in the first two years of the individual. It is very difficult to diagnose the ASD, as there is no standard laboratory test, like blood test, to diagnose if an individual has ASD. The diagnosis is generally done by a medical practitioner based upon the behavior, learning traits and age appropriate development of the child. But the diagnosis is a really challenging task, and it has been observed that some individuals do not receive a firm diagnose till late years and as a result they cannot get the early assistance that is required. With the advent of machine learning based methods, researchers have tried to develop classification approaches to diagnose ASD individuals based upon a set of features which are created using the behavior patterns of the ASD effected individuals. The machine learning based methods results in fast and reliable detection of the disorder, as the manual methods may be biased due to the state of the medical professional. In this work an effort has been made to compare the various machine learning based approaches proposed so far for the classification of individuals into Autistic or non-Autistic based upon the features.

1. Introduction

Autism spectrum disorder is a health condition affected by the development of the brain; it influences the social behavior of the individual

and also hinders the communication ability and the social interaction. It may also be associated with some typical specific behavior traits, which may be repetitive in nature. This disorder has been termed as a spectrum as it may influence each individual in a different way, leading to varied symptoms and severity [1]. The ASD covers a broad range of disorders which were earlier known by different names such as Autism, Asperger's Syndrome etc. But now all these conditions are collective covered under a single disorder named as Autism spectrum disorder.

The onset of ASD generally happens in the early years of an individual i.e. childhood when the child is to start interacting with the outside world. The autism symptoms generally start appearing in the first year of birth, but they are generally ignored thinking, that each child has its own time frame for achieving different milestones. But the symptoms are clearly visible at the age of around 2 years. Although till date no possible cure has been researched by the medical fraternity, but if the disorder is diagnosed early then early care and treatment can definitely make the lives of the individuals better and make a big difference as how they grow and live with this disorder during the latter years of their lives [2].

The symptoms of ASD in early childhood are generally less eye contact, less reaction to when the communication is carried out, some infants may develop normally for the initial few months, but later they drop the milestones that they have achieved in the early months. The firm symptoms of ASD are visible by the age of 2 years.

But it has been observed that each child that has been affected by ASD, will have varying degree of symptoms and behavioral patterns depending upon the degree of effect of the disorder. As it has been observed that few children have high intelligence level but have difficulty in communication while others can communicate normally but have low intelligence levels, whereas some may have difficulty in adjusting to the social environment and prefer being alone rather amongst the society.

Each individual having ASD may have an exclusive combination of symptoms, and the extent to which those symptoms are present is sometimes hard to evaluate. The diagnosis is generally made depending upon the degree of damage and how it reflects in the normal day to day life of the individual.

It has been observed that as an individual matures, some of them start interacting with others and reflect very mild behavioral disturbances. The once who are having very less problems can eventually lead near normal lives, but those who continue to suffer language and social problems tend to face lot of issues in their teenage.

The exact cause of the ASD is still not known, but it has been apprehended that the cause of this can be either due to the genes, that is mutation of the genes due to inheritance or may be due to spontaneous mutation. The genetic mutations effect the development of the brain or how the communication take place within the brain cells. The other factor that may lead to ASD is the environment, the research is still being conducted to study the effect of viral infections, complications during pregnancy, medical treatment or the pollution may be attributed to the onset of the ASD.

The diagnosis of ASD is bit tough, as there is not standard yard stick using which it can be diagnosed. The medical practitioner generally observes the conduct and growth of the child to diagnose this condition. The diagnosis is generally done in two phases, in the initial phase the normal developmental screening is carried out, if there are firm indicators to ASD, then the comprehensive evaluation is done in which genetic tests, neurological response test, hearing and vision tests are performed whereas in developmental screening test only the learning skills are monitored and evaluated.

Initially the whole diagnostic process was being done by the humans. The humans are not so versatile and they don't have much diligence, they tend to get tired and bored, so their diagnostic ability highly dependent upon their level of fatigue and mood. Therefore, there seems to be a need for taking the help of the computational ability of the computers to diagnose this multi facet disorder, which is influenced by many fuzzy factors.

With the advent of machine learning techniques efforts have been made by the researchers to automate the process of the diagnosis of ASD [3], by with the help of the classification technique, that classifies the individuals as effected by ASD or not depending upon the set of features that are observed during the diagnostic phase.

In this work an effort is being made to compare the various machine learning based methods developed to diagnose the ASD affected patients. The diagnostic procedure, advantages, and disadvantages of the major techniques proposed so far will be evaluated. The paper has been organized as: the first Section introduces the ASD, Section 2 presents the various techniques and methods proposed to automate the diagnostic process, Section 3 discusses the limitations and advantages of the methods proposed so far. Finally Section 4 concludes the paper by listing the research gaps and the future directions that can be used to enhance the automatic diagnostic methods for ASD.

2. Methods

In the year 1943 an article titled "Autistic Disturbances of Affective Contact", was published by Leo Kanner, in this the case studies of 11 children in the age group of two to eight years was presented, the author described the behavioral patterns and the environment in which the children were being brought up were studied along with the educational qualifications of the parents [4]. In the initial stages most of the physiologists diagnosed the cases of Autism based upon the observation patterns laid by Kanner.

An artificial neural network based classification technique for classification of Autism, using the gene signatures has been proposed [5]. The diagnose based upon the genetic factors is very complex as autism has heterogeneous and complex dependence upon the genes, as it is not dependent upon a single gene mutation, different individuals have different genetic mutations in the case of Autism. Therefore there is a need to analyze the genetic data at the genomic level, to discover the presence of ASD. The artificial neural networks have the capability to learn and efficiently classify the large pool of features. In the proposed method the researchers have used a multilayer perceptron based neural network using the back propagation based training on the microarray genetic information of the subjects who were being scanned for the possible ASD. The proposed method was able to efficiently accomplish its task with an accuracy of 73%.

The classification of ASD based upon supervised learning of brain connectivity measures extracted from EEG scanning [6]. In this work the patterns of phase synchronization of 128-channel EEG signals are captured

from normal and ASD effected children. A total of 24 children were scanned, consisting of 12 children from each group. The EEG's were captured when the children were having happy, frightened or neutral feelings. In this work the minimum and maximum phase synchronized states were taken, to study the connectivity of the features of the brain connectivity. In this work the researchers have classified the subjects based upon the support vector machines and discriminant analysis based upon polynomial kernel. The proposed method is able to classify the subjects with 94.7% accuracy, with the sensitivity of 85.7% and specificity of 100%.

A two phase method for categorization of Autism, was proposed by [7], in this method in the first phase a neural network based approach is used to classify the subjects as effected by Autism or not, and in the second phase a fuzzy system based approach is used to find the degree of Autism.

A machine learning based approach has been developed to classify the children who are eight years old, based upon the data collected by the Autism and developmental disabilities monitoring network (ADDM) from the population of the United States. The analysis is based upon the words and phrases used in the evaluation sheets of the children in the specified age group [8]. They have used a random forest classifier for the purpose of classification and it was able to predict with an accuracy of 86.5%, with 84.0% sensitivity, 89.4% predictive value positive.

A support vector based classification approach has been applied to diagnose the Autism [9], in this a dataset of 2500 records has been considered, that consist of observation and interview items. The proposed model is capable of achieving sensitivity between 85.6%-94.3% and specificity of 80.9%-89.3%, using a set of 10 features. The features for the proposed study have been selected using a greedy backward-elimination method and the classification has been carried out using the linear kernel and radial bias kernel based SVM classifier.

A performance comparison of few prominent machine learning based classification algorithms namely Adaboost, Random Forest, J48 and KNN, using the multiomics features [10] has been done using the standard dataset of NCBI and SFARI. The performance has been analyzed using the classification accuracy of the different classifiers, and the J48 classifier

outperformed all the classifiers under consideration with an accuracy of 92.4%.

A SVM based classifier has been used to predict Autism based on behavior and development features for likely to be effected by Autism [11]. A decision tree based tool has been developed to support the medical practitioners and the researchers to diagnose the ASD [12]. A deep learning based approach has been developed using the brain imaging dataset [13], to identify the presence of ASD in subjects using the Autism Brain Imaging Data Exchange dataset. The proposed method is able to detect the presence of ASD with an accuracy of 70%. The performance of Support Vector Machine, K nearest neighbor and Random Forest classifiers was measured using the data of 104 adolescent subjects using the scan data, using 10 fold cross validation [14], to detect the presence of ASD. The random forest method was able to classify the ASD with the accuracy of 100%.

The fMRI time series data was taken ABIDE dataset to classify the subjects as Autistic and non-autistic using the SVM classifier [15], they also analyzed the different features and then did the classification based upon the optimal feature set. The proposed method was able to classify the data with 66% accuracy.

An ASD screening mobile application has been developed based upon the feature selection and rule based machine learning method, using the Weka tool [16]. The proposed method selects the features that have the class determination capabilities and they are not correlated to each other, thus avoiding the redundancy. They have tested three methods on three different datasets.

A random SVM cluster based method has been proposed to classify the ASD patients [17], with classification accuracy of 96.15%, using the brain features of the subjects under study. A machine learning based method [18] has been developed to classify the ASD affected children, by using the parents' answers to the questionnaire and short videos that were captured and tagged to observe the behavior of the subjects.

An efficient ASD prediction application for mobiles have been developed based upon the machine learning techniques. The mobile application is not dependent upon the age of the subject being diagnosed, it uses the

combination of Random Forest and Classification and Regression Trees based classifier and Random forest and ID3 classifier [19]. The method has been tested on two different datasets namely AQ-10 Dataset and a self-collected dataset of 250 subjects. The proposed model gave encouraging results for both the datasets.

The personal characteristic data was extracted from the ABIDE dataset, and evaluated with nine state of the art machine learning algorithms, to classify the ASD behavior of the subjects [20]. Out of the nine classifiers neural network based classifier was able to classify with an accuracy of 62%. In yet another study four different machine learning algorithms were evaluated to grade the ASD subjects with mild, moderate, severe and no Autism [21], the decision tree based classification method provided an accuracy of 100% on training dataset and an accuracy of 96% on testing dataset.

An SVM based classification of autistic and non-autistic subjects were carried out using the voxel-based morphometric features and an accuracy of over 80% was achieved [22]. A comparative analysis of four different machine learning algorithms namely decision trees, random forest, SVM and neural networks, was carried out on ASD data collected from the UCI repository [23] and accuracy of almost 100% was achieved for all the classifiers. XG Boost Classifier and Gradient Boosting Classifiers [24] were used to classify the ASD affected individuals with an accuracy of 97.1%. Thus giving a hope that machine learning based methods can accurately classify the ASD subjects with good accuracy. A novel method termed as Ensemble Classification for Autism Screening (ECAS), which has been observed to be a very influential method in creating multiple classifiers and thus helping to accurately predict the autistic features in an individual.

3. Discussion

The sooner a patient is diagnosed with ASD, better are the prospects for the effected, as early intervention can help the effected lead an improved life. But it has been observed, that the parents refrain from consulting the medical practitioners, thinking that each child has its own way of learning or achieving the milestones. In majority of the cases the early symptoms of ASD

are ignored. It is only when the child becomes eligible to enter the school, either the teachers or the parents realize that there is something abnormal in the child, which is not getting him or her to be at pace with the peers. The second reason for late medical intervention is the lack of specialized doctors in the field of ASD, as it has no cure, but only the quality of life can be improved. Mostly physiologists are consulted to help the effected subjects. The routine general practitioners are generally not well trained to firmly confirm the presence of ASD and the experts are fewer in number. Therefore there seems to be a strong need to develop a cheap automated testing system that can be implanted at each pediatric clinic, whereby the parents can access the traits of their children by evaluation them based upon the questions that are being asked.

Although in the recent last decade much effort has been made by the researchers, to develop a machine learning based techniques, so as to classify the Autistic and non Autistic patients. Most of the techniques have been developed based upon the recent state of the art classifiers, considering the features of either brain activity, gene expression or the behavioral traits. Majority of the approaches are able to classify the subjects with around 70% of accuracy, which is quite low. As categorizing an individual as affected by ASD, is a big question mark on the future prospects of the individual's life. It generally lower downs the morale of the individual as well. Therefore utmost reliable and accurate system needs to be developed to do such a challenging classification. Still there are certain studies that claim to achieve 100% classification accuracy, but they have used a small sample size to test their claims.

With the proliferation of handheld computing devices or for that matter the smart mobile phones, the availability and affordability of the collection of data has increased. Most of the machine learning methods are dependent upon the datasets only, if accurate and large amount of datasets are available, then surely, encouraging and reliable results can be obtained by using machine learning techniques to classify the ASD and non ASD effected subjects with nearly hundred percent accuracy. The competence of knowledge extracted through machine learning approaches seems to provide hopeful diagnosis for the ASD research accomplishments.

4. Conclusion

The individuals affected by ASD, tend to show communication impairment, lack of social interaction, behavioral problems, restraining interest and certain sort of repetitive activities. For the accurate diagnosis of this disorder, all the above symptoms must be perceptible before 36 months of age. In some cases the parents and the primary care providers do observe these features as early as less than 12 months, but a firm diagnosis cannot be done, till the subject attains the 24 months of age, as before that age the symptoms have not established fully. Moreover in few cases the ASD affected subjects have learning disabilities as well whereas others have average intelligence. In this paper most of the research works proposed by the researchers regarding classification of the ASD effected and not effected subjects have been studied. It is hoped soon, an efficient and robust technique will be developed to do the diagnosis.

References

[1]   P. Munda, State of the art 2015 State of the art (2015), 28-32.

[2]   C. M. Corsello, Early intervention in autism, Infants Young Child 18(2) (2005), 74-85, doi: 10.1097/00001163-200504000-00002.

[3]   D. Bone et al., Applying Machine Learning to Facilitate Autism Diagnostics: Pitfalls and promises, J Autism Dev Disord 45(101) (2016), 1121-1136 doi: 10.1007/s10803-014-2268-6.Applying.

[4]   L. Kanner, Autistic disturbances of affective contact, Nerv. Child J. Psychopathol. Psychother. Ment. Hyg. Guid. Child 2 (1943), 217-250.

[5]   K. T. Quach, Application of Artificial Neural Networks in Classification of Autism Diagnosis Based on Gene Expression Signatures, 2012, doi: 10.1364/AO.56.00B222.

[6]   W. Jamal, S. Das, I. A. Oprescu, K. Maharatna, F. Apicella and F. Sicca, Classification of autism spectrum disorder using supervised learning of brain connectivity measures extracted from synchrostates, J. Neural Eng. 11(4) (2014), 1-27, doi: 10.1088/1741-2560/11/4/046019.

[7]   R. Ahuja and D. Kaur, Neuro-Fuzzy Methodology for Diagnosis of Autism 5(2) (2014), 2171-2172.

[8]   M. J. Maenner, M. Yeargin-Allsopp, K. N. Van Braun, D. L. Christensen and L. A. Schieve, Development of a machine learning algorithm for the surveillance of autism spectrum disorder, PLoS One, 11(12) (2016), 1-11 doi: 10.1371/journal.pone.0168224.

[9]   F. Hauck and N. Kliewer, Machine Learning for Autism Diagnostics: Applying Support Vector Classification, Int'l Conf. Heal. Informatics Med. Syst., pp. 120-123, 2017.

[10]   G. Anurekha and P. Geetha, Performance Analysis of Supervised Approaches for Autism Spectrum Disorder Detection, Int. J. Trend Res. Dev. (IJTRD), Eng. Technol. (2017), 16-18.

[11]   G. Bussu et al., Prediction of Autism at 3 Years from Behavioural and Developmental Measures in High-Risk Infants: A Longitudinal Cross-Domain Classifier Analysis, J. Autism Dev. Disord, 48(7) (2018), 2418-2433, doi: 10.1007/s10803-018-3509-x.

[12]   I. L. Cohen, The Autism Spectrum Disorder Decision Tree (ASD-DT) for the PDDBI: Rationale and Usage for Screening and Intervention Planning, p. 12, 2018.

[13]   A. S. Heinsfeld, A. R. Franco, R. C. Craddock, A. Buchweitz, and F. Meneguzzi, Identification of autism spectrum disorder using deep learning and the ABIDE dataset, NeuroImage Clin. 17 (2018), 16-23, doi: 10.1016/j.nicl.2017.08.017.

[14]   A. Demirhan, Performance of Machine Learning Methods in Determining the Autism Spectrum Disorder Cases, Mugla J. Sci. Technol. (2018), 79-84, doi: 10.22531/muglajsci.422546.

[15]   M. Mathur and T. Lindberg, Autism Spectrum Disorder classification using Machine Learning techniques on fMRI, 2018.

[16]   F. A. Thabtah, Detecting Autistic Traits using Computational Intelligence & Machine Learning Techniques, 2018.

[17]   X. A. Bi, Y. Wang, Q. Shu, Q. Sun and Q. Xu, Classification of autism spectrum disorder using random support vector machine cluster, Front. Genet. 9 (2018), 1-10, doi: 10.3389/fgene.2018.00018.

[18]   H. Abbas, F. Garberson, E. Glover and D. P. Wall, Machine learning approach for early detection of autism by combining questionnaire and home video screening, J. Am. Med. Informatics Assoc. 25(8) (2008), 1000-1007, doi: 10.1093/jamia/ocy039.

[19]   K. S. Oma, P. Mondal, N. S. Khan, M. R. K. Rizvi and M. N. Islam, A Machine Learning Approach to Predict Autism Spectrum Disorder, 2nd Int. Conf. Electr. Comput. Commun. Eng. ECCE 2019, pp. 7-9, 2019, doi: 10.1109/ECACE.2019.8679454.

[20]   M. N. Parikh, H. Li and L. He, Enhancing diagnosis of autism with optimized machine learning models and personal characteristic data, Front. Comput. Neurosci. 13 (2019), 1-5, doi: 10.3389/fncom.2019.00009.

[21]   C. S. Kanimozhiselvi, D. Jayaprakash and K. S. Kalaivani, Grading autism children using machine learning techniques, Int. J. Appl. Eng. Res. 14(5) (2019), 1186-1188.

[22]   M. Al Diabat and N. Al-Shanableh, Ensemble Learning Model for Screening Autism in Children, Int. J. Comput. Sci. Inf. Technol. 11(2) (2019), 45-62, doi: 10.5121/ijcsit.2019.11205.

[23]   T. L. Praveena and N. V. M. Lakshmi, Prediction of Autism Spectrum Disorder Using Supervised Machine Learning Algorithms 8 (2019), 142-145.

[24]   N. V. G. Raju, K. Madhavi, G. S. Kumar and G. V. Reddy, Prognostication of Autism Spectrum Disorder (ASD) using Supervised Machine Learning Models 4 (2019), 1028-1032.