A Review on Early Detection of Autism Spectrum Disorder using Eye Tracking and Deep Learning

Maria Sofia S[1], Nithin Mohanan[2]

Department of Computer Science

St. Thomas College (Autonomous), Thrissur

Kerala, India

Abstract:- **Autism Spectrum Disorder (ASD) is a neuro disorder characterized by different challenges like deficit in speech, language development, nonverbal communication and social interaction. ASD if detected at an earlier stage can bring about an improvement in their social behavior. Therefore, there is a need for automated and accurate tools for detection of ASD. This literature review examines two methods for the early detection of autism in children. The first method is the detection of autism using eye tracking. The second one presents a deep learning model for the classification of children as either healthy or potentially having autism with 94.6% accuracy using deep learning technology. Patients have common facial deformities which can be used to detect autism. The model recommends that autism can be diagnosed using images of children.**

Keywords – Autism Spectrum Disorder, eye tracking, deep learning, feature extraction, classification, MobileNet.

I. INTRODUCTION

ASD is a permanent problem. There are some difficulties in early detection of autism. Parents are unable to detect the subtle symptoms that characterize autism. Also pediatricians and other professionals lack proper tools for the detection in early stages.

At present, the diagnosis is done in two levels: at first level children with series of signs of probability of ASD is found out and these children move to the second level where a series of diagnostic tests are carried out and the case is registered for clinical diagnosis.

The project put forward by Natalia I. Vargas-Cuentas, et al.[1] introduces a creative tool which can be used in poor remote communities where there is a lack of specialists for the early detection of autism. Furthermore, this tool can be used by massively and can be used by doctors not specialized in pediatrics. The project also aims at creating a database of patients diagnosed.

Mohammad-Parsa Hosseini et al.[5], think that facial recognition is the most promising way for diagnosis of autism. They analysed a study conducted by the scientists at University of Missouri and reached at a conclusion that autistic children share common facial features which are distinct from the children who are not diagnosed with autism. The study unearthed certain features of autistic children like unusually broad upper face, wideset eyes, shorter middle region of the face including cheek set and nose.

II. BACKGROUND

ASD can be detected from deficiencies in social and communicative behaviour. In diagnosis using standard tools like ADOS (Autism Diagnostic Observation Schedule) and ADI – R (Autism Diagnostic Interview – Revised) medical experts are required.

In ADOS direct observation of the child is involved and in ADI – R child's parents are interviewed. Studies conducted in US by Centers for disease control denotes that ADOS and ADI – R are used in less than 0.1% and 2.1% respectively in communities. Natalia I. Vargas-Cuentas, et al., analysed the study conducted by Klin, Schulz and Jones[2] and pointed out that only 30% use other instruments and parent questionnaires and 67% use non standardised and validated tools.

Natalia I. Vargas-Cuentas, et al. examined a study conducted by Jones W., Carr K and Klin A[3] and revealed that children with ASD did not have an attachment with the eyes of the adults who approached them, and they preferred to see the portion of the mouth. So there is a relationship between eyes and level of social deficit, which can be useful for the early detection of autism.

Patients with autism find difficulties in social and communication skills. Also they show repetitive behaviour. Autism is a genetic disorder, but the diagnosis is done using behavioral characteristics and facial deformities. Autistic patients have common facial features which allows the diagnosis of the disease from a single image. This characteristic of the patients is used by Mohammad-Parsa Hossieni et al. in their deep learning model. This model uses MobileNet and two dense layers for feature extraction and classification.

III. LITERATURE REVIEW

According to Natalia I. Vargas-Cuentas, et al. (2016) a large percentage of population is affected by ASD. It is very difficult for the families of children with autism to diagnose autism at early stages. Not only the individual is impacted by this situation but the families are also affected. Therefore there is a significant need for the tools to identify early diagnosis.

In the paper "Diagnosis of Autism using an Eye Tracking System" (2016) Natalia I. Vargas-Cuentas, et al. has presented a project that aims to propose a method for the early diagnosis of autism, which is noninvasive and efficient. This proposal increases the probability of early detection of ASD and therefore early interference. The project presents an advanced and simple tool which can be used in poor and remote communities where there is a shortage of experts who can diagnose ASD. The tool is also envisioned to be used as a resource by doctors who are not specialised in paediatrics. There is also an intention to develop a database of patients.

Autism can be diagnosed from defective social and communicative behaviour. Standardised tools like ADOS in which direct examination of the child is conducted and ADI-R, where the child's parents are interviewed. But these tools require medical experts on the subject.

Natalia I. Vargas-Cuentas, et al. used the study presented by Jones W., et al [3], conducted in 66 children. It was found that the children with ASD did not have a binding with the eyes of the adult who approached them, instead they were looking into the mouth of the speaker[4].

The Autism diagnosis tool developed by Natalia I. Vargas-Cuentas, et al., is based on eye tracking system and it aims to identify the early changes in the visual favourites of the child. The video they used composed of two scenes. In the first one, the child is taking part in different games and social activities. The second one is the nonconcrete motion forms and curves with eye – catching colours.

The project developed by Natalia I. Vargas-Cuentas, et al., aimed at creating a web server that contains all the information about the child which helps to create an electronic health record. The three primary processes used in the project are as follows:

a) *Pre-processing*: In this, techniques are used to improve the quality of the image. A filter is applied to smooth and enhance the edges and to remove noise from the image. The image is converted to grayscale and contrast correction is also done.

b) *Processing*: In this stage, face coding of the child is performed. A cascade algorithm is also used for face detection and detection of both the eyes.

c) *Feature extraction*: In this stage, visual edge detection of the iris is done and black pixels is counted to distinguish the direction of observation.

To compare and validate the results, Natalia I. Vargas-Cuentas, et al., used a tool named M-CHAT which was designed by Robins et. al.[9], in the United States. It consists of 23 questions with yes/no answers.

Natalia I. Vargas-Cuentas, et al., identified that the sample size of the project should be 8 case children and 24 control children. Control children are male and female children in the age group of 18 months and 7 years who have no previous diagnosis of mental or neurological disease. The case children are the children who have previous diagnosis of ASD and receiving therapy.

Evaluations performed by Natalia I. Vargas-Cuentas, et al., were at IMLA(Medical Institute of Language and Learning). In the next step, MCHAT was implemented. It was done to gather additional information like the environment in which the child grows, sex, age, weight etc. This procedure is done by parents. Finally algorithm was implemented. A video is presented to the child with two sections on same screen. Left side is a social scene and right side is colourful moving objects. A webcam records the child watching the 2 scenes and this becomes the input to the software and it is used to quantify the percentage of time the child is watching a scene or another.

The children were asked to watch five videos of one minute. Based on the observations the following results were obtained by Natalia I. Vargas-Cuentas, et al.

Children number	Watched videos	Control child	Case child
14	5 videos	11	3
4	4 videos	4	0
5	3 videos	4	1
7	2 videos	4	3
1	0 videos	0	1

Table 1: Number of children per number of watched videos

As shown in Table 1, a higher number of clinical controls and cases could watch all the five videos displayed by the software. If any child could not see any of this video, then he belonged to case child.

The results were also compared with M-CHAT test. It was observed that 87.5% of case children were diagnosed with M-CHAT also.

The visual favourites of the child regarding social and abstract scenes helps to determine the risk of autism in children. Natalia I. Vargas-Cuentas, et al., have suggested that the children may be in the company of parents while watching the video, so that they feel secure to watch the videos.

In their work, Mohammad-Parsa Hosseini et al. have used the data set from Kaggle which comprises of over 3000 images of both children with and without autism. The data set was split into training, testing and validation subgroups. It was again sub divided into autistic and non-autistic folders. Both the autistic and non-autistic training groups contained 1,327 facial images each. The autistic and non-autistic testing group consisted of a total of 280 images, each having 140 images. The validation category had a total of 80 images – 40 images with autism and 40 images without autism.

Mohammad-Parsa Hosseini et al. reviewed the work done by Wen-Bing Horng and Associates(2001) [6] which categorized the facial images into four: babies, young adults, middle-aged adults and old adults. Their study classified the data using two back propagation neural networks – one emphasized on geometric features and the other on wrinkle features. One of problem with their work was that the age cut-offs for varying levels of adults do not have any distinguishing features which was necessary for the study. In order to avoid this complexity, Mohammad-Parsa Hosseini et al. decided to classify the data as simply autistic and non-autistic.

Mohammad-Parsa Hosseini et al. also reviewed the study by Shan(2012) [7], which used Local Binary Patterns(LBP) to depict faces. They applied SVM and were able to achieve a success rate of 94.81% in gender determination of the subject. The main innovation in this study was that they used only real life images for the classification. Until then, mostly ideal images, most of which were frontal, with clean background, occlusion free, were used for the classification. Comparatively Mohammad-Parsa Hosseini et al. also used real life images.

Mohammad-Parsa Hosseini et al. studied the work done by El-Baz et al.[8] which focused on the analysis of Cerebral White Matter(CWM) in individuals with autism. They used it to determine if classification can be done based only on brain images. In their work El-Baz et al., the CWM was first segmented from proton density MRI. Then CWM gyrification was extracted and quantified. This methodology used a cumulative distribution function of the

distance map of CWM gyrification to identify autistic and non-autistic persons. Though this study brought in successful results, the images were taken from deceased persons. So the success rate in living individuals is unknown. Mohammad-Parsa Hosseini et al. used real time images rather than costly MRI images.

Though there are a number of Convolutional Neural Networks (CNN), that can be used for image analysis, MobileNet[10] has proven to be very effective. It reduces cost and computation time. MobileNet has shown that thinner models has the same accuracy as the wider models and has significantly reduced the number of parameters for analysis. So Mohammad-Parsa Hosseini et al. decided to use MobileNet for analysis.

In their work, Mohammad-Parsa Hosseini et al., has used deep learning techniques to learn about autism. They have used facial features which are present in autistic children but not in non-autistic children for autism detection.

Mohammad-Parsa Hosseini et al. has used the data set from Kaggle, which contained images from online, Facebook and Google image searches. The images were cropped so that the face occupied most of the image. Before training, images were categorized into three: train, validation and test. Images are placed in each category manually. The duplicates in the dataset has been cleaned out to improve accuracy.

For this dataset Mohammad-Parsa Hosseini et al., have used MobileNet to perform deep learning. MobileNet is utilized followed by two dense layers. The first layer is used for distribution and customization of weights. This becomes the input to the second layer which is used for classification. For MobileNet, an alpha of 1 was used and depth multiplier of 1 was used. To make binary predictions from MobileNet, two fully connected networks are appended to the model. The first layer with 128 neurons, is connected to the prediction layer, which predicts whether the child is autistic or not. The training was completed with a test accuracy of 94.64%. The results are quite promising. There were many issues with the images in the data set like improper age ranges. Improvement in the dataset would result in more accurate results.

IV. CONCLUSION

The project put forward by Natalia I. Vargas-Cuentas, et al. had an accuracy of 87.5% , confirmed using MCHAT diagnosis. The algorithm proposed by V Mohammad-Parsa Hosseini et al. hopes to get an accuracy of more than 95%. Moreover success of this algorithm also helps to diagnose other diseases like Down's Syndrome, which changes the facial features. The diagnosis based on this algorithm can be done using an image of the face.

V. ACKNOWLEDGMENT

REFERENCES:

[1] Natalia I. Vargas-Cuentas, Daniela Hidalgo, Avid Roman-Gonzalez, Michael Power, Robert H. Gilman, Mirko Zimic, "Diagnosis of Autism using an Eye Tracking System", IEEE Global Humanitarian Technology Conference, pp. 13 – 16, October 2016.

[2] Klin, Shulz and Jones, "Social Visual Engagement in infants and Toddlers with autism: Early developmental transitions and a model of pathogenesis". Neuroscience & Biobehavioral Reviews, Elsevier, Vol. 50, pp. 189 – 203, March 2015

[3] Jones W., Carr K., & Klin A., "Absence of Preferential Looking into the eyes of approaching adults predicts level of social disability in 2-year-old toddlers with Autism Spectrum Disorder", Archives of General Psychiatry 65(8), pp. 946 – 54, August 2008.

[4] Warren J., & Klin A., "Heterogeneity and Homogeneity in across the Autism Spectrum: Role of Development", Journal of the American Academy of Child and Adolescent Psychiatry, Vol. 48, Issue 5, pp. 471 – 473, May 2009

[5] Mohammad-Parsa Hosseini, Madison Beary, Alex Hadsell, Ryan Messersmith and Hamid Soltanian Zadeh, "Deep Learning for Autism Diagnosis and Facial Analysis in Children", Frontiers in Computational Neuroscience, January 2022.

[6] Wen-Bing Horng, Cheng-Ping Lee and Chun-Wen Chen, "Classification of Age Groups Based on Facial Features", Tamkang Journal of Science and Engineering, Vol. 4, No. 3, pp. 183 – 191, 2001.

[7] Caifeng Shan, "Learning Local Binary Patterns for Gender Classification of Real-World Face Images", Pattern Recognition Letters 33, pp. 431 – 437, 2012.

[8] El-Baz A., Manuel F. Casanova, Georgy Gimel'farb, Meghan Mort, Andrew E. Switwala, "A New Image Analysis Approach for Automatic Classification of Autistic Brains", 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro (Arlington, VA: IEEE), April 2007.

[9] Diana Robins, Deborah Fein and Marianne Barton, "The Modified Checklist for Autism in Toddlers: an initial study investigating the early detection of autism and pervasive developmental disorders", 31(2), 2001.

[10] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobiyas Weyand, Marco Andreetto, Hartwig Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications", arXiv, April 2017.