Prediction of Autism Spectrum Disorder using Deep Learning: A Survey

Kiran Sankpal[1], Pradnya Sankpal[2], Rutuja Dudhagundi[3], Maryann Thomas[4], Prof. S.R. Nalamwar[5]

[1,2,3,4]*Undergraduate Student, Dept. of Computer Engineering, AISSMS COE, Pune, Maharashtra, India*
[5]*Professor, Dept. of Computer Engineering, AISSMS COE, Pune, Maharashtra, India*

---***---

Abstract - *Autism Spectrum disorder is a serious developmental disorder that impairs the ability to communicate and interact socially including speech as well as nonverbal communication. Autism patients may suffer many difficulties such as anxiety, depression, motor difficulties, and sensory problems. According to WHO, about 1 out of every 160 children are diagnosed with Autism. It is a chronic condition and may last for years or even lifelong. Diagnosis of Autism Spectrum Disorder requires a significant amount of time and cost. If diagnosed during early childhood, proper medication can be administered and it would prevent the child's condition to further deteriorate, improving the child's mental health. It would also help to reduce long term costs if diagnosed late. An accurate, easy to use, cheap and fast autism screening tool which would predict the presence of autism traits as well as the severity would be greatly useful to patients. In this project, our objective is to develop a user-friendly android mobile application which would predict the possibility of autism spectrum disorder. It would take input from the patient (or patient's parent) in two forms: symptoms in the form of radio buttons and a scanned image of the brain. This will be done with the use of deep learning algorithm: ADAM and other optimizing techniques to achieve accuracy.*

Key Words: Autism Spectrum Disorder, Deep Learning, Neural Networks, Multilayer Feedforward Neural Network, Gradient Descent

1. INTRODUCTION

Autism Spectrum Disorder (ASD) is a neurological medical condition which often goes unnoticed in toddlers until it has reached a severe stage. The main motivation of this project is to help the affected children and improve their lives for the better. We intend to use multi-layer feed forward artificial neural network to predict ASD.

1.1 NEURAL NETWORKS

In an Artificial Neural Network, there are large number of processors which operate in parallel and arranged in layers. A multilayer feed forward network comprises of a single input layer, several hidden layers, and an output layer. Raw input information is provided to the first layer which is analogous to nerves in the human brain. The next layer takes as input the output of the preceding layer instead of raw input data and finally, the last layer produces the output of the system.

1.2 Gradient Descent

We use Gradient descent algorithm to minimize error in the model. It is an optimization algorithm used to minimize cost function. It moves in the direction of steepest descent in an s manner as defined by the negative of the gradient. [1]

There are 3 variants of Gradient Descent:

1. Batch Gradient Descent- It helps in computing the cost function. Each and every training example is processed for all iterations of gradient descent. When the numbers of examples are tremendously high, the computation cost increases to a great extent and renders this variant unfeasible.

2. Stochastic Gradient Descent (SGD) - In this variant, a single training example is processed per iteration. Here, the parameters are updated with each iteration. It avoids performing redundant computations for large datasets unlike Batch Gradient Descent which processes all examples in each iteration and recomputes gradient for similar examples. Hence, SGD is much faster than Batch Gradient Descent.

3. Mini batch Gradient Descent- It is similar to Batch Gradient Descent. But it divides the entire training set into batches and processes each batch separately. It is faster than Batch Gradient Descent.[2]

We will use Adam algorithm which incorporates the concept of SGD in our system.

2. RELATED WORK

[3]. M Duda, R Ma, N Haber, DP Wall, "Use of machine learning for behavioural distinction of autism and ADHD", IEEE: In this study, the performance of six machine learning algorithms on data, using the 65 items in the SRS as features and the diagnosis of either ASD or ADHD as the prediction class. This subsampling technique was used to ameliorate the issue of significantly unbalanced classes, as well as to safeguard against any age or gender biases that may be inherent to the data. Behavioural diagnosis of both ASD and ADHD is a time-intensive process. In this analysis of archival data faces the problem of limited content of the data sets available.

[4]. Kazi Shahrukh Omar, Prodipta Mondal, Nabila Shahnaz Khan, Md. Rezaul Karim Rizvi, Md Nazrul Islam, “A machine learning approach to predict Autism Spectrum Disorder”: In this paper, a mobile application was created. Using Amazon Web Services, an API was created. This application divides the user into three different age groups to obtain accurate results. Various Machine Learning algorithms were tried and tested. The results showed that Random Forest-CART showed better performance than the Decision Tree-CART algorithm, while the combination of Random Forest-CART and Random Forest-ID3 algorithm showed better performance when compared with both the Random Forest-CART and Decision Tree-CART separately.

[5]. Bram van den Bekerom, "Using Machine Learning for Detection of Autism Spectrum Disorder": Prediction of Autism Spectrum Disorder in a child was carried out by using developmental delay, learning disability and speech or other language problems as attributes. Physical activity, premature birth and birth weight was also used to improve the accuracy. They used the 1-away method to predict the severity of Autism Spectrum disorder quite reasonably. The 1-away method improved the accuracy from 54.1% to 90.2%, which is a significant increase. Predicting the severity of ASD was possible with this kind of data set after applying the 1-away method, but the big increase in accuracy does warrant further research. This research could verify whether the 1-away method is actually useful in this case, or should not be used. It could also look into other attributes that are linked better to the severity of ASD.

[6]. Daniel Bone, Matthew S. Goodwin ,Matthew P. Black, Chi-Chun Lee, Kartik Audhkhasi, Shrikanth Narayanan, "Applying Machine Learning to Facilitate Autism Diagnostics: Pitfalls and Promises”: This paper critically evaluates and attempts to reproduce results from two studies that claim to drastically reduce to diagnose time autism using machine learning. Their failure to generate comparable findings to those reported by Wall and colleagues using larger and more balanced data under-scores several conceptual and methodological problems associated with these studies. It concludes with proposed best-practices when using machine learning in autism research, and highlight some especially promising areas for collaborative work at the intersection of computational and behavioral science.

[7]. G.Brightwell, C. Kenyon, H. Paugam-Moisy, “Multilayer neural networks: one or two hidden layers?”: In this research paper, they explain the number of hidden layers required by a multilayer neural network with threshold units to compute a function f from Rd to {O, I}.In dimension d = 2, Gibson characterized the functions computable with just one hidden layer, understood the assumption that there is no "multiple intersection point" and that f is only defined on a compact set. The restriction of f to the neighborhood of a multiple intersection point or of infinity is also considered which give necessary and sufficient conditions for it to be locally computable with one hidden layer. Adding the conditions to Gibson's assumptions is not sufficient to ensure global computability with one hidden layer by exhibiting a new non-local configuration, the "critical cycle", which implies that f is not computable with one hidden layer.

[8]. Simon S. Du, Jason D. Lee, Haochuan Li, Liwei Wang, Xiyu Zhai, “Gradient Descent Finds Global Minima of Deep Neural Networks”: This paper proves gradient descent achieves zero training loss in polynomial time for a deep over parameterized neural network with residual connections (Res Net).The authors have proposed the deep residual network (ResNet) architecture which helps training neural network. It was proved that the randomly initialized gradient descent converges to zero training loss at a linear rate. And finally it was proved that randomly initialized gradient descent achieves zero training loss.

3. PROPOSED SYSTEM

3.1 AQ10 Dataset

The AQ is a ten item constricted version of a 50 item measure which is designed to measure the degree of severity of ASD symptoms in an individual. It is a set of questions which are mostly behavioral in nature.[9]

3.2 Deep Learning Model

With deep learning algorithms, we can achieve the task of predication of ASD without human intervention. Neural networks (NNs) prove to be a great choice for achieving predication accuracy. For training the model, Stochastic Gradient Descent (SGD) is one of the best options available. In this study, we have used ADAM algorithm for implementing gradient descent.

3.2.1 Feed-forward Networks

A Feed Forward Network is a type of Neural Network having several layers of nodes which process and generate appropriate output. There may be any number of hidden layers between the input and output layer. More the number of hidden layers, more is the accuracy and precision. The signal moves only in one direction from input layer to output layer. There are three types of functions in feed forward network namely linear function, Heviside function and Sigmoid function. Sigmoid and Linear functions are mostly used in the input and hidden layers whereas Heviside function is used in the output layer.

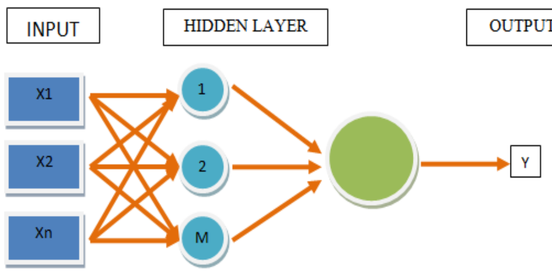

Fig -1: Multi-layer Feed-forward Neural Network

3.2.2 ADAM Algorithm

Adaptive Moment Estimation (Adam) [10]. is an algorithm in which adaptive learning rates for each parameter are computed. It is a combination of RMSprop and Momentum.

By using *V* (exponential moving average of gradients), it acts upon the gradient component, like in momentum. Like in RMSprop, it acts upon the learning rate component by dividing the learning rate α by square root of S, the exponential moving average of squared gradients.

$$w_{t+1} = w_t - \frac{\alpha}{\sqrt{\hat{S}_t + \epsilon}} \cdot \hat{V}_t$$

Where,

$$\hat{V}_t = \frac{V_t}{1-\beta_1^t}$$
$$\hat{S}_t = \frac{S_t}{1-\beta_2^t}$$

Are the bias corrections, and

$$V_t = \beta_1 V_{t-1} + (1-\beta_1)\frac{\partial L}{\partial w_t}$$
$$S_t = \beta_2 S_{t-1} + (1-\beta_2)\left[\frac{\partial L}{\partial w_t}\right]^2$$

With V and S initialized to 0. Where, t-time step, w-weight/parameter to be updated, α -learning rate, ∂L/∂w-gradient of L, the loss function to minimize, w.r.t. to w

The authors have proposed the default values of 0.9 for β1, 0.999 for β2, and 10−810−8 for ϵ. Their results prove that Adam works well in practice and when compared with other adaptive learning-method algorithms, it works better.

3.2.3 Additional Strategies for optimizing SGD

3.2.3.1. Shuffling and Curriculum Learning

In order to avoid biasing the optimization algorithm, we try to provide the training examples in an unordered manner. Therefore, after every epoch, it is better to shuffle the training data. Alternatively, for progressively harder problems, curriculum learning is preferred in which the training examples are provided in a meaningful order. This may lead to improved performance and better convergence. [11].

3.2.3.2. Early stopping

Error should always be monitored on a validation set during training. Training must be stopped if the validation error does not improve enough.

3.2.3.3. Gradient Noise

According to Neelakantan et al. [12] adding noise that follows a Gaussian distribution to each gradient update makes the network more sturdy to poor initialization. It helps train complex networks efficiently.

4. ARCHITECTURE

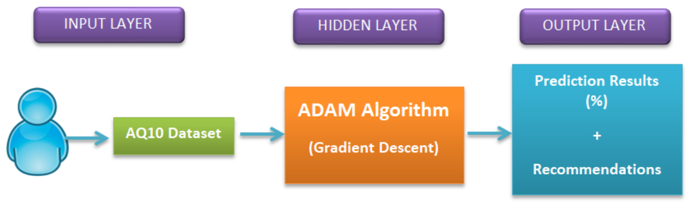

Fig -1: System Architecture

The user will enter the symptoms through the AQ10 dataset. This data will be tested on the training model which works on ADAM Algorithm. The model will then provide output in the form of percentage and will also provide recommendations according to the results.

5. FUTURE SCOPE

The proposed system deals with the prediction of Autism Spectrum Disorder. But this can be extended by providing Image Processing of MRI brain scans. With Image Processing, the patient can avoid tedious process of answering the AQ10 questionnare. Alongside Image Processing of MRI scans, we can also implement real time behavioral analysis with video screening.

6. CONCLUSION

In this paper we have proposed a system for the prediction of Autism Spectrum Disorder which will increase the accuracy of the percentage prediction result. It will also give faster results compared to previously created models due to the use of ANN (Deep Learning). It will also suggest recommendations which would prove to be helpful to the health of patients.

ACKNOWLEDGEMENT

We would like to express our gratitude to our project guide, Prof. S.R. Nalamwar for her able support and guidance. We would also like to thank God for providing us undivided strength and perseverance.

REFERENCES

[1] https://ml-cheatsheet.readthedocs.io/en/latest/gradient_descent.html [online]

[2] D. Kornack and P. Rakic, "Cell Proliferation without Neurogenesis in Adult Primate Neocortex," Science, vol. 294, Dec. 2001, pp. 2127-2130, doi:10.1126/science.1065467.

[3] M Duda, R Ma, N Haber, DP Wall, "Use of machine learning for behavioural distinction of autism and ADHD",Transl Psychiatry (2016) 6, e732; doi:10.1038/tp.2015.221

[4] Kazi Shahrukh Omar, Prodipta Mondal, Nabila Shahnaz Khan, Md. Rezaul Karim Rizvi, Md Nazrul Islam, "A machine learning approach to predict Autism Spectrum Disorder", 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), 7-9 February, 2019

[5] Bram van den Bekerom, "Using Machine Learning for Detection of Autism Spectrum Disorder", University of Twente

[6] Daniel Bone, Matthew S. Goodwin ,Matthew P. Black, Chi-Chun Lee, Kartik Audhkhasi, Shrikanth Narayanan, "Applying Machine Learning to Facilitate Autism Diagnostics: Pitfalls and Promises", ", J Autism Dev Disord

DOI 10.1007/s10803-014-2268-6

[7] G.Brightwell, C. Kenyon, H. Paugam-Moisy, "Multilayer neural networks: one or two hidden layers?", LIP, URA 1398 CNRS ENS Lyon, 46 allee d'Italie F69364 Lyon cedex, FRANCE

[8] Simon S. Du , Jason D. Lee, Haochuan Li , Liwei Wang, Xiyu Zhai, "Gradient Descent Finds Global Minima of Deep Neural Networks", arxiv:1811.03804v4[cs.LG] 28 May 2019

[9] 9https://www.midss.org/content/autism-spectrum-quotient-aq-and-aq-10[online]

[10] Kingma, D. P., & Ba, J. L. (2015). Adam: a Method for Stochastic Optimization. International Conference on Learning Representations, 1–13.\

[11] .Bengio, Y., Louradour, J., Collobert, R., & Weston, J. (2009). Curriculum learning. Proceedings of the 26th Annual International Conference on Machine Learning, 41–48. http://doi.org/10.1145/1553374.1553380

[12] Neelakantan, A., Vilnis, L., Le, Q. V., Sutskever, I., Kaiser, L., Kurach, K., & Martens, J. (2015). Adding Gradient Noise Improves Learning for Very Deep Networks, 1–11. Retrieved from http://arxiv.org/abs/1511.06807