

## Predictive Analysis for daASD Using Population based Incremental Learning

Yugandhar Bokka<sup>1,\*</sup>, R.N.V.Jagan Mohan<sup>2</sup> and M.Chandhra Naik<sup>3</sup>

<sup>1</sup>GIET University, Odisha

<sup>2</sup>Dept of CSE, SRKR Engineering College, Bhimavaram-534202, AP.India

<sup>3</sup>Dept of CSE, GIET University, Odhisha

Received 24 October 2020; Accepted 29 June 2021

### Abstract

Autism is a clinically defined behavioral syndrome that initially appears in childhood and reflects the underlying neuro-developmental irregularities. The main aim is to define the clinical trajectories of dyslipidemia-associated ASD (daASD) is to take the massive data integration has revealed an emerging ASD subtype characterized by dyslipidemia. However, the clinical course of this subtype remains unknown. To analysis of Electronic Health Record (EHR), data identifying the pattern sequence of medical events experienced by people with daASD. Within these data, individuals with daASD will be identified as those having at least one ASD diagnosis and dyslipidemia diagnosis or repeated blood lipid test results that are outside the normal range. Comparisons will be made between individuals with daASD, ASD with no dyslipidemia (ASD+/dyslipidemia-), dyslipidemia with no ASD (ASD-/dyslipidemia+), and individuals with no history of neuro-developmental disorders or abnormal lipid levels (ASD-/dyslipidemia-). Matching will be based on age, gender, ethnicity and address/zip code, as a proxy of socioeconomic status. The parents of each member of the above groups will be identified and their data will be collected. The direct comparisons of daASD with other ASD subtypes are required diagnosis for comorbid epilepsy/recurrent seizures and comorbid Fragile X syndrome using population based incremental learning. In this regard, Random Forest approach is on daASD for classification and semi-supervised learning is a process for daASD. The daASD experimental result is based on Electronic Health Records of Least Square Estimation Regression with Predict the Disease.

**Keywords:** Autism, dyslipidemia, Electronic Health Records, Least Square Regression, Random Forest Classification

### 1. Introduction

Autism is a neuro-developmental disorder [1-3,17]. It is including abnormal or unreciprocated interpersonal and emotional interactions, disordered language and communication, and repetitive and stereotypic behavior [18]. The few studies have evaluated young children with autism. Various scientists have characterized brain structural abnormalities early in the clinical course of autism as part of an ongoing longitudinal study of brain development. Obesity and diabetes are highly common among pregnant women in India. Still there is no study has examined the independent and combined effects of maternal pre-pregnancy obesity and motherly diabetes on the risk of Autism Spectrum Disorder (ASD) in parallel with other development disorders [4]. Motherly diabetes has been connected with increased risk of autism spectrum disorder in children. The relationship between motherly pre-pregnancy obesity and ASD has been inconsistent. No study yet has examined the joint effects of these two conditions. The combination of maternal obesity and diabetes was connected with more serious risk of ASD than either obesity or diabetes alone, especially as ASD co-occurred with brain disability. Human being overweight assigned as having a Body Mass Index (BMI) of 25 to 29.9. Obesity is outlined as having a BMI of 30 or greater. The normal kind of obesity, three levels reflect the increasing health risks for women that go along with increasing Body Mass Index. Obesity during pregnancy for women's puts at a risk of several serious health problems such as cardiovascular

diseases like gestational diabetes, preeclampsia and sleep apnea [5]. ASD have enlarged radically in which can't be entirely explained by changes in diagnostic practice. During a similar period, obesity and diabetes increases to epidemic levels. Presently amid women of generative age, more than a third are obese have pre-pregnancy diabetes and additional will develop gestational risks during pregnancy [15]. The emerging proof is a relation between motherly prenatal diabetes with the risk of ASD in children. This study on the connection between maternal obesity and the risk of ASD has produced inconsistent effect. Although obesity and diabetes are highly comorbid rarely have studies attempted to separate their independent and combined effects in ASD. In addition, although ASD frequently co-occurs with other developmental disorders and may have shared etiology with some developmental disorders[19-21]. Meta analysis has evaluated the association between antidepressant use during pregnancy and the risk of ASD in children, all reported a significant association [6-10]. Machine Learning provides an approach to efficiently feature derived from large data, there by generating clinically relevant predictions [11-14,18]. The aim of this study was to test the ability of ML models applied to Electronic Health Records to predict ASD early in life. To predict the risk of ASD in pregnancy women in a large population-based sample.

### 2. Related Work

In view of this paper is to consider for daASD by population based incremental learning (PBIL). An Electronic Health Records based PBILuses for Random Forest Classification. In

\*E-mail address: yug.599@gmail.com

ISSN: 1791-2377 © 2021 School of Science, IHU. All rights reserved.

doi:10.25103/jestr.143.23

addition to process of Meta-Analysis is a Semi-supervised learning method steps for predictive purpose.

**2.1. Population based Incremental Learning**

The Population based incremental learning are patient Electronic Health Records represented by real value in the range of [1, N] indicating the probability that any particular all Electronic Health elements appears in that women’s pregnancy Obese Health records. In this approach relation the implicit population defined by uni-variate distribution. In each generation is denoted value ‘t’.

The population P(t) u s then sort the decreasing order of fitness. S sort (f) (P(t)) with ‘u’ being the ASD best and ‘v’ being the non-ASD worst solution.

The probability of a function is defined by

$$P \propto (X_i) = P_{\alpha-1}(X_i) + \gamma (u_i - v_i)$$

$$\forall i=1, 2, 3 \dots N$$

Where  $\alpha$  is a Constant and it varies from 1, 2... N and  $r \in [1, 50]$  is a constant defined by  $\gamma = 50/N$ .

For instance, If  $\gamma=50/N$  for  $N=1, \gamma=50/1=50; \gamma=50/N$ , for  $N=2, \gamma=50/2=25; \gamma=50/N$ , for  $N=3, \gamma=50/3=16.5$  and soon.

**2.2. Random Forest Procedure of Clinical Trail Classification**

A random forest is a supervised learning based machine learning procedure is used for both classification and regression[16,22]. A random forest makes decision trees on randomly selected medical data samples, gets prediction from each tree and selects the best solution by means of elective. It also provides a respectable pointer of the feature significance. A random forest has a variety of applications, such as feature selection in which is used to classify and predict diseases.

The Figure-1 shown, there are two fragments. One is to find out the Children’s with ASD other one is Non-ASD whereas Pregnancy Women’s about their comorbid diagnosis of individual patients using decision tree procedure. The second part, after collecting all the Electronic Health Records (EHRs) of comorbid diagnosis of patients’ trajectory in the list of attributes compare with other attributes for ASD or Non-ASD. The whole process of being above said from pregnancy women’s and either ASD or Non-ASD in which identify the disease. Classifiers are also known as the forest. Figure-2 shows the individual decision trees are generated using an attribute selection indicator such as, Least Square estimation, Correlation Coefficients for each record. Each tree depends on an independent random sample. In a classification problem for each tree, compare the individual attributes with obese in both Non-ASD and ASD. To find out the ASD or Non-ASD each individual attribute with the value of correlation coefficient and estimate the parameter value with BMI value.

**2.3. Procedure of Meta-Analysis for Obesity with ASD**

Meta-Analysis is a statistical approach to combine the results from multiple studies to increase power, improve estimates of the size and resolve uncertainty when report disagrees. A meta-analysis is generally led by a methodical assessment, because this allows recognition and assessment of all the relevant facts i.e., in that way warning the risk of bias in précis approximations.

**The following common steps are as follows:**

1. Formulation of queries on Obesity with ASD trouble using PICO model (Population, Intervention, Comparison, and Outcome).
2. Exploration of Earlier Patient’s clinical trial literature.
3. Selection of studies (integration criteria) such as
  - 3.1. Derived from quality criteria for instance the necessity of randomization and unsighted in a clinical trial.
  - 3.2. Variety of exact studies on a well-specified subject matter for the treatment of interference to diminish Obesity in Pregnancy with ASD.
  - 3.3. Make a decision whether unpublished studies are incorporated to avoid publication bias (partiality).
4. Decide which dependent variables or précis measures are allowed.
5. Assortment of a meta-analysis model for fixed result or random belongings meta-analysis.
6. Inspect sources of between-study heterogeneity for this case using subgroup analysis or meta-regression.

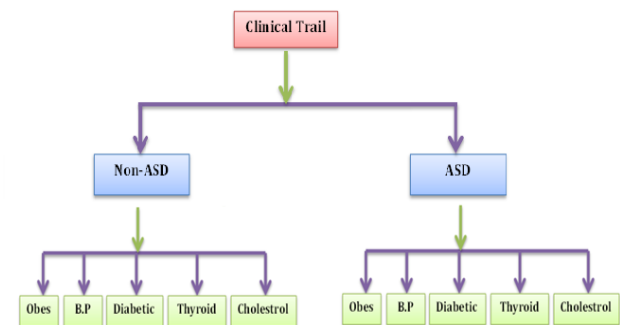


Fig. 1. Clinical Trail Decision Tree Patient Diagnosis

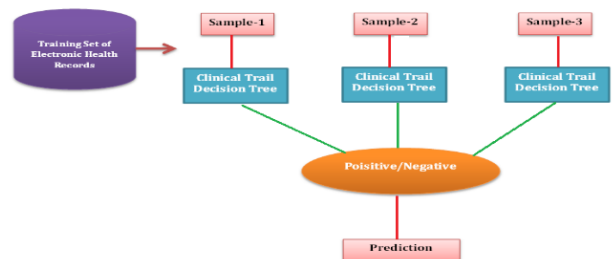


Fig. 2. Random Forest Classification of ASD/Non-ASD Disease Idf

The Official supervision for the behavior and coverage of Obesity with ASD meta-analysis is provided.

**3. Results and Discussion**

The experimental result analyses on electronic health records from India and Clalit Health Services, Israel’s largest health maintenance organization (HMO). It includes diagnoses, medications, lab results, procedures, and clinical measurements and importantly, Clalit data includes familial ties, enabling parent-child analyses. The Table-1 shows the sample of 150 pregnancy women’s has taken and observed how many are obese according to their BMI values. Then we have to divide them all into two categories, i.e. obese and non-obese. After that, we need to monitor all health parameters like Systolic Blood Pressure, Glucose (sugar), LDL and HDL of all pregnant women. Then, we have to compare all these parameters among obese and non-obese woman. It needs to observe if obesity is associated with ASD during pregnancy women’s.

**Table 1.** Clinical Trial data of patients like pregnancy women's

Patient ID	Age	(BMI) Obesity	SystolicBloodPressure	Glucose(Sugar)	LDL	HDL
P01501	17	58	177	118	172	18
P01502	17	49	182	99	183	33
P01503	18	57	190	94	164	36
P01504	17	48	169	98	161	34
P01505	19	57	185	116	178	36
P01506	20	57	190	98	189	31
P01507	20	50	178	113	179	30
P01508	18	50	163	91	174	39
P01509	20	51	176	119	184	31
P01510	19	58	167	111	169	33

a. To estimate High Blood Pressure determine the prediction equation, which is the **least square regression equation** of Y on X, is a Weight Status of Pregnancy Women's. Assume it to be  $Y=a_0+a_1X$  its normal equations are:

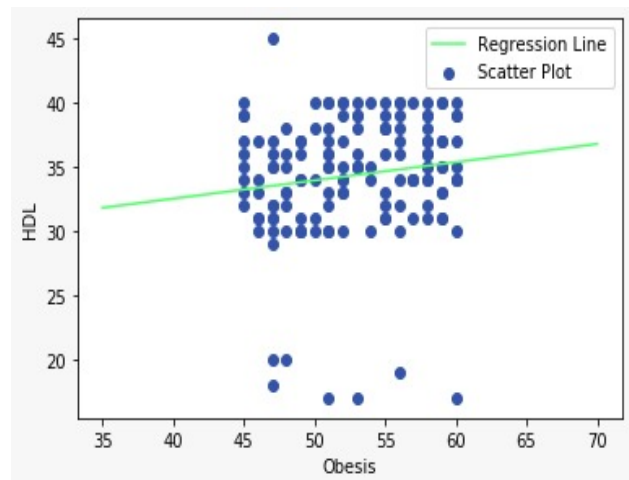
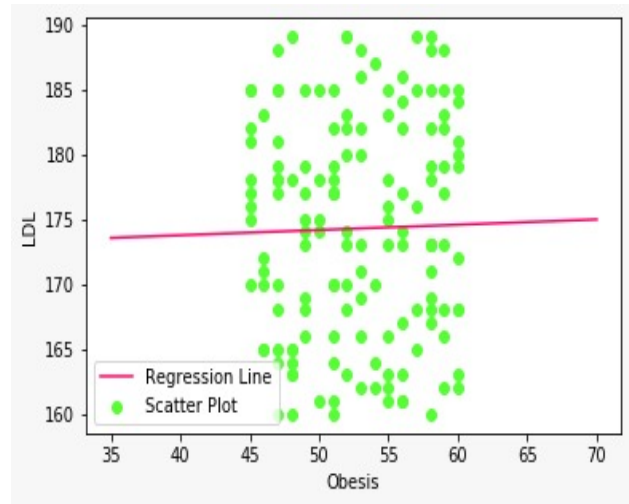
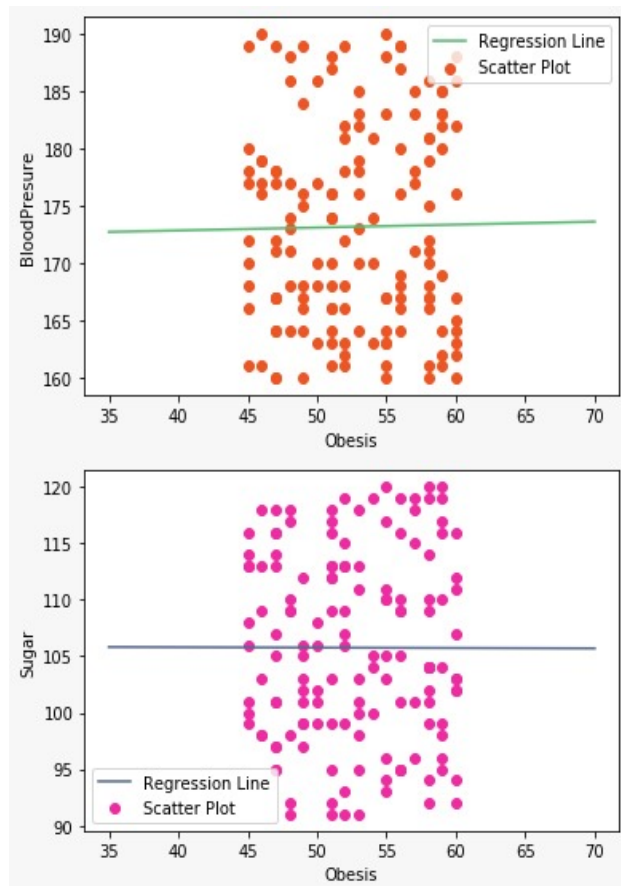
$$\sum Y = Na_0 + a_1 \sum X \tag{1}$$

$$\sum XY = a_0 \sum X + a_1 \sum X^2 \tag{2}$$

b. To find the association between Pregnancy Weight Status women and High Blood Pressure determine the correlation coefficient r by

$$r = \frac{N \sum XY - \sum X \sum Y}{\sqrt{[N \sum X^2 - (\sum X)^2][N \sum Y^2 - (\sum Y)^2]}} \tag{3}$$

The least-squares regression technique is a way commonly used in Regression Analysis is a mathematical approach. It is used to find the best-fit line that signifies the association between an independent variable like Obesity and dependent variables like Systolic Blood Pressure, Cholesterol like LDL and HDL, Glucose (Sugar) with the help of mathematical technique.



**Fig. 3.** Correlation between Obese and Blood Pressure, Sugar, LDL, HDL

The Figure 3 shown, a) The Correlation Coefficient is R from Obese with Blood Pressure and (Range between  $-1 \leq r \leq 1$ ). Pregnancy women with the Obese X and Blood Pressure Y are positively correlated. We may conclude that the obese patients with High Blood Pressure caused for ASD. b) Pregnancy women with the obese X and Highest Sugar Y are positively correlated. We may conclude that the obese patients with Sugar caused for ASD. c) Pregnancy women with the Obese High(X) and Low Cholesterol (Y) are positively correlated. We may conclude that the obese patients with Low Cholesterol caused for ASD. d) Pregnancy women's with the highest Obese X and High Risk Cholesterol Y are positively correlated. We may conclude that the obese patients with High Cholesterol caused for ASD.

**Table 2.** The Correlation Coefficient value for obese with BP, Sugar, LDL, HDL

	SystolicBloodPressure	Glucose(Sugar)	LDL	HDL
Obese	R=0.5	R=0.5	R=0.4	R=0.4

The random forest classifier produced the precision, recall and F1 score, these are known as efficient of predicting ASD traits.

**Table 3.** Random forest classification algorithm for ASD Prediction

Precession	Recall	F1 score
0.17	1.00	0.29

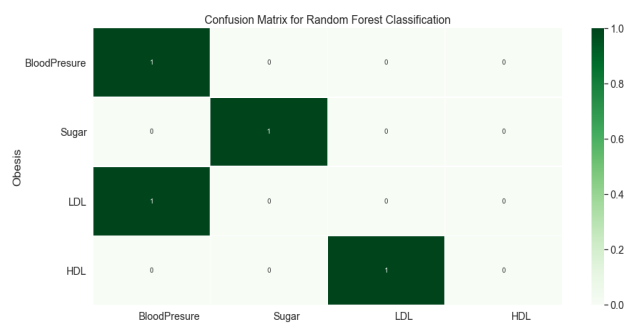


Fig. 4. Confusion Matrix of Predicted data for Obese and Blood Pressure, Sugar, LDL, HDL

#### 4. Conclusion

This study of work, it is considerable data integration has revealed an emerging ASD subtype characterized by dyslipidemia. To compile and then mine large-scale electronic health record (EHR) data to identify the prototypical sequence of medical events experienced by individuals with daASD. The Patient electronic health records are indicating the probability that women's pregnancy

approaches for population based incremental learning and meta-analysis is used. To find out the association between Pregnancy Weight Status women, High Blood Pressure, Sugar and Cholesterol determine the correlation coefficient that is Pregnancy women with the Weight Status X and High Blood Pressure, Sugar and Cholesterol Y are positively correlated. So that, the obese patients are not of High Blood Pressure they are considered or there may be considered as ASD. The Random Forest classification model is measured on precision, recall, accuracy for Autism.

#### Acknowledgment

As a follow up study, we will examine the prevalence of our findings in New Delhi families with autism seen at the National Centre for Autism, in collaboration with Dr. Nidhi Singhal, Director of Research and Training at Action for Autism (National Centre for Autism) in New Delhi. The Ministry of Science and Technology, Department of Science and Technology, **India-Israel Collaboration Joint Research Project**, supported this work.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License.



#### References

- Baio, Jon, et al. "Prevalence of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2014." *MMWR Surveillance Summaries* 67.6 (2018): 1.
- Rahman, Rayees, et al. "Identification of newborns at risk for autism using electronic medical records and machine learning." *European Psychiatry* 63.1 (2020).
- Thabtah, Fadi. "Autism spectrum disorder screening: machine learning adaptation and DSM-5 fulfillment." *Proceedings of the 1st International Conference on Medical and Health Informatics 2017*. 2017.
- Ng, Marie, et al. "Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the Global Burden of Disease Study 2013." *The Lancet* 384.9945 (2014): 766-781.
- Ama Moor, Vicky Jocelyne, et al. "Dyslipidemia in patients with a cardiovascular risk and disease at the University Teaching Hospital of Yaoundé, Cameroon." *International journal of vascular medicine* 2017 (2017).
- Zhou, Xi-Hong, et al. "Association between maternal antidepressant use during pregnancy and autism spectrum disorder: an updated meta-analysis." *Molecular autism* 9.1 (2018): 1-7.
- Man, Kenneth KC, et al. "Exposure to selective serotonin reuptake inhibitors during pregnancy and risk of autism spectrum disorder in children: a systematic review and meta-analysis of observational studies." *Neuroscience & Biobehavioral Reviews* 49 (2015): 82-89.
- Kaplan, Yusuf C., et al. "Prenatal selective serotonin reuptake inhibitor use and the risk of autism spectrum disorder in children: A systematic review and meta-analysis." *Reproductive Toxicology* 66 (2016): 31-43.
- Rais, Theodor B., and Alexandra Rais. "Association between antidepressants use during pregnancy and autistic spectrum disorders: a meta-analysis." *Innovations in clinical neuroscience* 11.5-6 (2014): 18.
- Willis, Brian H., and Richard D. Riley. "Measuring the statistical validity of summary meta-analysis and meta-regression results for use in clinical practice." *Statistics in medicine* 36.21 (2017): 3283-3301.
- Sahni, Nishant, Gyorgy Simon, and Rashi Arora. "Development and validation of machine learning models for prediction of 1-year mortality utilizing electronic medical record data available at the end of hospitalization in multicondition patients: a proof-of-concept study." *Journal of general internal medicine* 33.6 (2018): 921-928.
- Steele, Andrew J., et al. "Machine learning models in electronic health records can outperform conventional survival models for predicting patient mortality in coronary artery disease." *PloS one* 13.8 (2018): e0202344.
- Devika Varshini, G., and R. Chinnaiyan. "Optimized Machine Learning Classification Approaches for Prediction of Autism Spectrum Disorder." *Ann Autism Dev Disord*.2020; 1 (1) 1001 (2020).
- Cheng, Wenshuai, et al. "Varying coefficient models for analyzing the effects of risk factors on pregnant women's blood pressure." *2014 13th International Conference on Machine Learning and Applications*.IEEE, 2014.
- Morales, Daniel R., et al. "Antidepressant use during pregnancy and risk of autism spectrum disorder and attention deficit hyperactivity disorder: systematic review of observational studies and methodological considerations." *BMC medicine* 16.1 (2018): 1-14.
- Altay, Osman, and Mustafa Ulas. "Prediction of the autism spectrum disorder diagnosis with linear discriminant analysis classifier and K-nearest neighbor in children." *2018 6th International Symposium on Digital Forensic and Security (ISDFS)*. IEEE, 2018.
- Christensen, Deborah L., et al. "Predictence and characteristics of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2012." *MMWR Surveillance Summaries* 65.13 (2018): 1.
- Alwidian, Jaber, Ammar Elhassan, and Rawan Ghnemat. "Predicting Autism Spectrum Disorder using Machine Learning Technique." *International Journal of Recent Technology and Engineering* 8 (2020): 4139-4143.
- Heléne EK Sundelin, Olof Stephansson, Christina M Hultman, Jonas F Ludvigsson "Pregnancy outcomes in women with autism: a nationwide population-based cohort study" 157.47.76.174 on 17-Apr-2021.
- Zerbo, Ousseny, et al. "Maternal infection during pregnancy and autism spectrum disorders." *Journal of autism and developmental disorders* 45.12 (2015): 4015-4025.
- Caly, Hugues, et al. "Machine learning analysis of pregnancy data enables early identification of a subpopulation of newborns with ASD." *Scientific reports* 11.1 (2021): 1-14.
- Tyagi, Bhawana, Rahul Mishra, and Neha Bajpai. "Machine Learning Techniques to Predict Autism Spectrum Disorder." *2018 IEEE Punecon*.IEEE.