International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. #, No.# (Mon-20..)

Maternal Dyslipidemia During Pregnancy Correlates with Elevated Lipid Levels in One-Year-Old Infants

Yugandhar Bokka[1], R N V Jagan Mohan[2] and M Chandra Naik[3]

[1] CSE, Gandhi Institute of Engineering and Technology (GIET) University Gunupur, Odisha, India.
[2]CSE, Sagi Rama Krishnam Raju Engineering College, Bhimavaram, India.
[3]CSE, Gandhi Institute of Engineering and Technology (GIET) University Gunupur, Odisha, India.

E-mail address: bokka.yugandhar@giet.edu, mohan.rnvj@srkrec.edu.in, srichandra@giet.edu

Abstract: Developmental disorders like autism spectrum disorder (ASD) can result from differences in children's brains. Neurodevelopmental problems are exacerbated by a confluence of genetic, environmental, and prenatal risk factors associated with ASD. Repetitive behaviors and deficiencies in social communication are among the early indicators in children. Although gestational risk factors are not the cause of ASD, they can impact how children interact with one another. On the other hand, these risk factors might also favorably influence the development of ASD, and women who are pregnant can help with interventions. Later in life, the development of autism is linked to changes in lipid levels at birth. Dyslipidemia, characterized by abnormal cholesterol and triglyceride levels, is more common in individuals with ASD than in their healthy siblings or unrelated controls. However, the specific predictive value of blood lipid profiles and the key markers for dyslipidemia associated with ASD remain unclear. This paper explores the influence of infant lipid levels on the development of dyslipidemia associated with ASD, considering gestational risk factors for mothers. A machine learning model is constructed using combined parental and childhood lipid levels to predict ASD. The model is then validated using independent cohorts and tested against lipid profiles from infancy. Various statistical approaches designed for biomarker discovery in Electronic Health Records (EHR) data are applied to achieve these objectives.

Keywords: Maternal dyslipidemia, Gestational dyslipidemia, Lipid levels, Early childhood development
Infant health, Longitudinal study.

1. INTRODUCTION

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder [1]. A person with autism spectrum disorder (ASD) experiences difficulties with speech, communication, and social skills. According to some estimates, India has 18 million cases of autism. The numbers could be higher as many go unrecognized, especially in rural areas where access to health care may be inadequate. Most people with ASD will need support in daily life. Approximately fifty percent of children with autism develop of mental impairment. About fifty percent have a major speech problem, one fourth have at least seizures related to epilepsy, and some have abnormally enlarged brains [4]. According to research, those with ASD have different brain development than usual controls.

The causes of autism are uncertain. It is a combination of environmental, biological, and genetic factors. A higher risk of autism associated during pregnancy, because preterm labour, low birth weight, gestational diabetes, Obese, Dyslipidemia, complications during childbirth and use of acetaminophen in pregnancy. we will identify the risk factors during pregnancy and their children then to reduce to risk of autism at an early stage. Some Studies have looked at young children with autism. As part of an ongoing study of brain development, a group of scientists has identified brain structural anomalies early in the clinical course of autism [8]. Pregnancy-related risk factors are the primary causes of ASD issues.

In India, obesity and diabetes are very common among pregnant women. A study comparing the risk of autism to the risk of other developmental disorders has not yet been conducted [4]. Dyslipidemia is the term used to describe abnormally high or low lipid levels. An overly high level of one or more lipid types in a pregnant woman's blood is known as dyslipidemia [5,6,9]. Predicting lipid levels and getting involved early can both ensure the health of the

E-mail:author's email

unborn child and greatly reduce the incidence of ASD during pregnancy.

Nowadays, machine learning techniques play a major role in healthcare, particularly when it comes to using health information for identifying diseases. The term "machine learning" (ML) describes a number of statistical methods that allow computers build up knowledge from their experiences without needing explicit code. Clinical trial research can be improved in a number of ways by applying machine learning (ML) technologies. This work is to create an improved and more accurate machine learning model for determining the presence of autism spectrum disorder.

The following are the main contributions made by this work:

- Investigation of routine blood lipid profiles in infancy and their information is associated with daASD.
- To explore maternal lipid levels are associated with infancy lipid profiles age of one year below.
- Classification models are used to predict ASD in infancy by observing the dyslipidemia and also, we look into the link between maternal lipids and children lipid levels using various statistical approaches.
- And finally test the predictive power of both cases using independent cohort studies.

This paper is organized as follows: The introduction to our paper is contained in the "Introduction" section. The entire literature review is summed up in the "Literature Review" section. The workings and methodology of the system we have developed and its implementation are explained in the "Materials and methods" section. The results and insights are presented in the "Results and Discussion" section. Our conclusions are finally highlighted in the "Conclusion" section.

2. LITERATURE REVIEW

This section reviews recent investigations that used machine learning methods to diagnose and evaluate autism spectrum disorder. The primary objective is to investigate and identify limitations in order to provide a novel, improved, and more successful machine learning methods for the forecasting of autism spectrum disorder.

The authors [1] describe the application of a variety of machine learning models to improve performance and accuracy rates, so enabling early diagnosis prediction of ASD. Five classification models are being used, with random forest tree and decision tree having the highest accuracy. With the use of a deep neural network (DNN), adults with ASD are classified in this study [2]. Diagnoses for ASD were correctly recognized by the DNN with 99.40% accuracy. A dataset from a different study [3] included sixteen people with autism who executed a sequence of hand signals diagnosing autism using machine learning methods. The ASD screening dataset is taken into consideration for analysis [4].The adult, adolescent, and newborn autism screening datasets were used by the author. The ASD is recognized by these characteristics.

A collection of 21 items from the UCI laptop research repository used for ASD analysis. Swarm Genius, it has been determined that only 10 features of the ASD dataset 21 components are necessary to distinguish ASD patients from non-ASD patients[5]. Those with ASD and ADHD had the highest risk of substance use-related issues. Additionally, parents, half-siblings, and full siblings of ASD probands had higher chances of substance use-related issues[6]. This study uses a Naïve Bayesian classification strategy to aid in the early identification of autism. The Naive Bayesian classifier's efficient classification process comes from its features' lack of significant correlation[7]. Bayes' Rule is used to derive the posterior from the previous and likelihood, since the latter two are generally simple to compute using a probability model[8].

Using a huge amount of data, a classification model is utilised to predict heart disease[9]. The accuracy rates of this study's autism prediction were 92.26%, 93.78%, and 97.10% in children, adolescents, and adults. In regard to efficiency, the Random Forest-CART algorithm performed higher than the Decision Tree-CART method[10]. ML is helpful in developing reliable algorithms. The purpose of these algorithms is to enhance ASD screening and diagnostic instruments [11].

This study evaluated if identifying a subset of behavioral traits could to improve the finding of ASD using a machine learning system. These findings might facilitate the challenging task of differential diagnosis for doctors and help to simplify the complex diagnostic process of ASD[12]. In order to anticipate kidney injury, we created a stacked-long short-term memory network using a pattern-mixture model. Compared the suggested outcome to traditional algorithms such as long short-term memory model and gradient enhanced trees. These models perform better in predicting kidney injury 24 hours before it occurs than the machine learning model[13].

Deep learning, machine learning techniques were used in this work to classify autism spectrum disorder. In comparison to previous methods, the deep neural network performed better[14]. The way an individual with ASD behaves among other people is one of the most significant signs of the disorder [15]. Approximately 1% of people overall are affected by ASD, and this percentage is rising quickly [16]. Social, racial, and ethnic groups can all have members with ASD. Most children in the US do not receive an ASD diagnosis until they are four years old [17]. This study identified the syndromic ASD using supervised machine learning approaches. By employing SVM and decision trees, they achieved 98% and 94% accuracy, respectively [18]. KNN is a nonparametric approach that depends on how close the training set and sample feature are to one another [19]. The primary goals of this effort are to determine the most effective machine learning classifier and to determine the key variables linked to autism. A multilayer perceptron is constructed to predict the probability of autism [20]. For their analysis of autism risk, the authors employed a weighted decision tree prediction model [21]. Using the NSCH autism dataset, the SVM, Naive Bayes, Decision Tree, and Random Forest algorithms were used to identify a set of conditions and assess the severity of ASD[22]. The goal of this effort was to construct some classification models that would aid clinical paediatricians in the early diagnosis of autism levels [23].

The suggested study makes use of an ASD diseases classification model for adults with ASD who are patients and non-patients who are either categorised with or not have ASD disease[24]. According to the results, linear SVM produce better accuracy [25]. utilising classification models in this investigation. Logistic regression yields the most accuracy when determining if a child is at risk for in its early stages [26]. This study uses various classification models like decision trees and naïve bayes to predict cardiovascular illness more accurately. The Heart Failure Dataset shows good performance using the Gaussian Naïve Bayes algorithm[27].

3. MATERIALS AND METHODS

A. *Classification*

We have used various classification models such as logistic regression, random forest, support vector machines and naive Bayes. Classification model is a supervised learning, which attempts to predict class labels. In this scenario, it aims to identify the abnormal or normal lipid levels using a binary classification. For binary classification, logistic regression (LR) has been used to model the probability of fitting into a specific category. The equation(1) is showing the sigmoid function is used in this classification model, producing a value between 0 and 1.

$$\emptyset(z) = \frac{1}{1+e^{-z}} \quad (1)$$

Because random forests (RF) are based on several decision trees, they have advantage of reducing the likelihood of overfitting. Every tree predicts a class, and then used to identify which class each patient belongs to normal or abnormal lipid. Selecting the optimal split is crucial, Entropy is one metric used for evaluating this purity, a minimum value of zero denotes the purity in the sample, while a maximum value of one denotes the impurity. In the entropy Equation (2), Pi refers to the proportion of values falling into class level i.

$$Entropy = \sum_{i=1}^{c} -P_i * \log_2(P_i) \quad (2)$$

Support vector machine (SVM) is a one of the robust methods and is also helpful for classification. SVM divides data into groups of related items, using a linear boundary known as a hyperplane, show in equation (3). In general, a well-constructed hyperplane that best separates classes or whichever has the most distance between them will enable accurate classification, as represented by the class values, provided you have two classes.

$$W.X + b = 0 \quad (3)$$

The Naive Bayes (NB) technique looks for ways to characterize an event's likelihood of happening. The Bayes theorem is used to determine the probability that event A will occur; this process is also known as conditional probability because event B has already happened, showing in equation (4). As a result, the likelihood that a sample will belong to a particular class is determined.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (4)$$

Extreme Gradient Boosting (XGBoost) is a robust and renowned machine learning algorithm that has acquired understanding for its effectiveness and performance in a variety of real-world applications. XGBoost is an ensemble learning method that generates a strong predictive model by aggregating the predictions of several weak learners. It works by growing trees one after the other, rectifying the mistakes caused by the preceding

tree. Reducing the entire loss function is the learning goal. Regularization techniques have been implemented into XGBoost in order to manage tree complexity and avoid overfitting. The inherent ability of XGBoost to manage incomplete data minimize the requirement for prepping the data. Both DART (Dropouts meet Multiple Additive Regression Trees) and XGBoost (Extreme Gradient Boosting) are boosting algorithms; in particular, DART is merely an extension of XGBoost. Dropout regularisation is implemented by DART when training trees. In order to avoid overfitting, dropout entails randomly removing some of the trees throughout each iteration. Dropout is used by DART, where every tree has a chance of being dropped after every round of boosting.

B. Model Architecture

The recommended model, as seen in Figure 1, is an overview of a framework that includes concepts that are utilized by lipid levels for help in learn, comprehension, or assessment of the estimate of autism spectrum disorder.

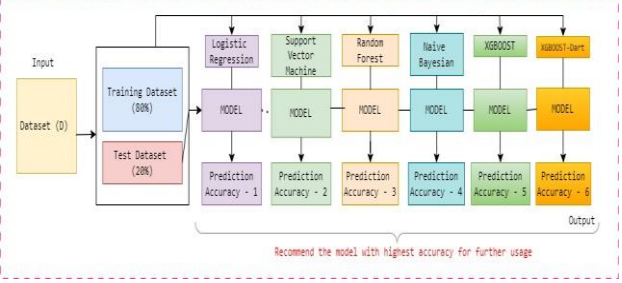


Figure 1. Model Architecture prediction of ASD.

The following five important steps are depicted in the proposed model: (1) Data is acquired from multiple laboratories. (2) A method of data splitting that divides data into both testing and training datasets (3) This model using various classification models such LR, SVM, RL, NB, XGB and DART-XGB (4) Metrics of evaluation are calculated with respect to parameters such as recall, accuracy, and precision.

C. Pseudo-code for proposed Model

Algorithm 1 XGBoost Classification Algorithm
Input: Dataset D
Output: Class Label (Normal / Abnormal)

```
Import Libraries
- pandas
- train_test_split from sklearn.model_selection
- xgboost as xgb
Load dataset
- data = ReadCSV('E:/Object-3/Mother_Child.csv')
- X = data.drop('M_target', axis=1)
- y = data['M_target']
Split data into training and testing sets
- X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
Initialize XGBoost classifier
- xgb_clf = xgb.XGBClassifier(objective="binary:logistic", n_estimators=100, use_label_encoder=False)
Train the model
- xgb_clf.fit(X_train, y_train)
Make predictions on the test set
- y_pred_xgb = xgb_clf.predict(X_test)
```

D. Dataset Description

In this study involved 150 mothers and their children in this study, showing below Table I. Blood lipids from infants under one year old and mother records of three trimesters during pregnancy are included. Clinical evaluation revealed no symptoms of sickness or illness in these infants. All of the newborns had weights adequate for their gestational ages and were born at term or almost term (finished at 37–41 weeks). Neonates or newborns who had any congenital defects, low birth weight or teeny for gestational age, premature delivery, or neonatal jaundice were thus excluded. Based on specific diagnostic metrics included in the collection, the dataset attempts to determine if a patient has autism or not.

TABLE I. D
ATASET FOR BLOOD LIPIDS WITH MOTHER AND THEIR INFANCY

M_BMI	*M_TC*	*M_TG*	*M_HDL*	*M_LDL*	*C_TG*	*C_HDL*	*C_LDL*	*C_TC*
30	270	217	71	192	98	41	113	124
31	310	476	71	213	94	42	96	155
31	236	158	86	199	57	42	86	148
28	261	202	90	138	66	42	112	131
29	379	413	82	206	73	41	96	128

E. Evaluation Metrics

A confusion matrix based on correctly and incorrectly predicted models is used to generate test data, which is used for evaluating how effectively the classification model performs, shown in Table II. The confusion matrix is then used to calculate the parameters of accuracy, sensitivity, and accuracy. Table III shows the lipids classified as Normal and Abnormal in confusion matrix.

TABLE II.
ONFUSION MATRIX FOR BINARY CLASSIFICATION

		Predicted Values	
		Positive	***Negative***
Actual Values	***Positive***	TP	FN
	Negative	FP	TN

TABLE III.
ONFUSION MATRIX FOR NORMAL AND ABNORMAL LIPIDS LEVELS

True positive (TP)	The affected patient's lipid levels have been shown to be abnormal.
False positive (FP)	The non-affected patient's lipid levels have been shown to be abnormal.
False negative (FN)	The affected patient's lipid levels have been shown to be normal.
True negative (FN)	The non-affected patient's lipid levels have been shown to be abnormal.

A binary classification test's accuracy is a statistical measure of how well it accurately detects or excludes out a condition. Equation (5) represents the percentage of correct results out of all the cases that were analyzed.

$$Accuracy = \frac{\sum_i cm_{i_i}}{\sum_i \sum_j cm_{i_j}} \quad (5)$$

Equation(6), where TP means true positive cases and FN for false negatives, is used to determine the true positive rate (Sensitivity), or the likelihood that an unusual patient would get a positive test result.

$$Sensitivity = \frac{\sum_i cm_{i_i}}{\sum_j cm_{i_j}} \quad (6)$$

Equation (7) represents the true negative rate, also known as specificity, which is the likelihood that a normal patient will have a negative test result.

$$Specificity = \frac{\sum_i cm_{j_i}}{\sum_j cm_{i_j}} \quad (7)$$

Equation (8) and (9) represents precision and F1-score.

$$Precision = \frac{\sum_i cm_{i_i}}{\sum_i \sum_j cm_{j_i}} \quad (8)$$

$$F1 - score = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} \quad (9)$$

The comparison of ROC and AUC results can also be used for evaluating how well each algorithm performs. The sensitivity and specificity of the classification method are plotted visually by ROC at various classification levels. AUC, on the other hand, calculates the area below the ROC curve, which represents the likelihood that a random positive data will be ranked higher by the classification algorithm than a random negative data.

4. RESULTS AND DISCUSSION

The results discussed here correspond to the study that details a thorough examination of the lipid profile in daASD. In the initial research study, the goal was to determine whether it was possible to distinguish between the two groups of infants who were 6 and 12 months old using statistical methods and also, we look into the link between maternal lipids and children lipid levels using various statistical approaches. Second, various machine learning algorithms (RL, SVM, RF, NB, XGB, and DART-XGB) were used to train the classification models that will identify patients at risk of ASD based on maternal lipids and infancy lipids compared to the control group.

First, we examine the routine blood lipids in infancy and also find their information is associated with daASD. The study includes one hundred fifty infants, of which 77 (451.3%) were under the age of six months and 73 (48.6%) older than six months; 68 were female and 82 were male. Table IV show the mean and standard deviation, as well as the interquartile-range (25th–75th percentile) and percentile normal distribution (25, 50, 75, and 90 percentile) for the lipid parameters—TC, TG, HDL, and LDL. The blood lipid values in both age groups—under 6 months and older than 6 months—are compared in Table V.

TABLE IV. I
NFANCY LIPID PARAMETERS NORMAL-DISTRIBUTED (N=150)

Parameters	**Mean ± SD (mg/dl)**	**Normal-distribution(mg/dl)**				**IQR (mg/dl)**
		Percentile (25%)	***Percentile (50%)***	***Percentile (75%)***	***Percentile (90%)***	
Total Cholesterol (mg/dl)	147.39 ± 16.009	136.60	147.39	158.19	167.91	21.60
Triglyceride (mg/dl)	80.07 ± 22.77	64.71	80.07	95.43	109.26	30.72
HDL-cholesterol (mg/dl)	47.5 ± 4.80	44.26	47.50	50.74	53.66	6.48
LDL-cholesterol (mg/dl)	102.78 ± 11.80	94.81	102.78	110.75	117.92	15.93

TABLE V. COMPARSION OF INFANCY LIPID PARAMETERS BETWEEN <6 AND >12 MONTHS OF AGE

Parameter	Age ≤6 Months (77/150)	Age >6 Months (73/150)	Significance (*P*)
Total cholesterol (mg/dl)	76.79±22.48 (mg/dl)	83.53±22.39 (mg/dl)	0.613
Triglyceride (mg/dl)	48.27±4.87 (mg/dl)	46.68±4.55 (mg/dl)	0.049 *
HDL-cholesterol (mg/dl)	102.93±11.30 (mg/dl)	102.61±12.24 (mg/dl)	0.036 *
LDL-cholesterol (mg/dl)	147.98±15.53 (mg/dl)	146.76±16.58 (mg/dl)	0.087

The total cholesterol measured at 76.99 mg/dl (± 22.48) for infants under 6 months of age and 83.53 mg/dl (± 22.39) for those over 6 months of age showed no statistically significant differences (P = 0.613). The LDL-cholesterol mean values for the age groups ≤6 months and >6 months were 147.98 mg/dl (± 15.53) and 146.76 mg/dl (± 16.58), respectively. These values did not vary statistically (P = 0.087).

But the HDL-cholesterol levels were discovered to be 102.93 mg/dl (±11.30) in the age group ≤6 months and 102.61 mg/dl (±12.24) in the age group >6 months, indicating a statistically significant difference (P = 0.036). Additionally, the measured triglycerides were statistically different (P = 0.049) at 48.27 mg/dl (±4.87) in the age group ≤6 months and 46.68 mg/dl (±4.55) in the age group >6 months.

There is a positive correlation between the lipid levels of the mother during pregnancy and the comparable lipid levels of the infant at 6 and 12 months after delivery. The features of maternal and infantile factors are displayed in Table VI. Pregnant women were, on average, 30.93 (SD 8.3) years old. At six months old (average 3.37 months (SD 1.6)) and twelve months old (average 8.71 months (SD 1.46)).

The lipid profiles of infants at 6 and 12 months old as well as those of pregnant women are displayed in Table VII. At 12 months, compared to 6 months, there is a large increase in total cholesterol and HDL levels and a significant decrease in triglycerides and LDL.

TABLE VI. BASIC FEATURES OF MATERNAL DURING PREGNANCY AND INFANCY

Features	During Pregnancy (n=150) Mean±SD	Infancy below 6 months (n=77) Mean±SD	Infancy above 6 months (n=73) Mean±SD
Maternal age (years)	30.93±8.32	3.376±1.68	8.71±1.46
Pre-pregnancy BMI (kg/m2)	26.44±4.35	-	-
Smoking	0.03±0.17	-	-
Drinking	0.02±0.16	-	-
High-Temperature	0.01±0.11	-	-
Total cholesterol (mg/dl)	242.28±58.59	76.79±22.48	83.53±22.39
Triglyceride (mg/dl)	230.40±121.42	48.27±4.87	46.68±4.55
HDL-cholesterol (mg/dl)	68.36±13.92	102.93±11.30	102.61±12.24
LDL-cholesterol (mg/dl)	143.00±44.78	147.98±15.53	146.76±16.58
Diabetics	152.01±17.69	-	-
BP(systolic)	135.06±7.29	-	-
BP(diastolic)	82.32±1.59	-	-
Obesity	0.26±0.44	-	-

TABLE VII. LIPID PARAMETERS OF MATERNAL DURING PREGNANCY AND INFANCY

Lipids Parameters	During Pregnancy (n = 150)	Below 6 months (n = 77)	After 6 months (n = 73)	P-value
Age	30.93±8.32	3.376±1.68	8.71±1.46	9.21
Total cholesterol (mg/dl)	242.28±58.59	76.79±22.48 (mg/dl)	83.53±22.39 (mg/dl)	0.08
Triglyceride (mg/dl)	230.40±121.42	48.27±4.87 (mg/dl)	46.68±4.55 (mg/dl)	0.59
HDL-cholesterol (mg/dl)	68.36±13.92	102.93±11.30 (mg/dl)	102.61±12.24 (mg/dl)	0.09
LDL-cholesterol (mg/dl)	143.00±44.78	147.98±15.53 (mg/dl)	146.76±16.58 (mg/dl)	0.86

The relationship between mother lipid levels and infant lipid levels is shown as a result of the data analysis using linear spline mixed models. The relationship between mother lipid levels and infant lipid levels from birth to age twelve is displayed in Table VIII. The p-values from the interaction between the maternal lipid parameters and the lipid parameters of the infant are the findings of the linear mixed model analysis. Because an

effect of interaction that is significant (P-value < 0.05) would suggest that the relationship between the mother's lipid levels and lipid levels of the infant vary according to the infant's age. A higher rise in the total cholesterol of infants was linked to higher maternal levels (0.09 ≤ P-value ≤ 0.90). The change in infant triglyceride levels was not substantially correlated with maternal lipid levels (0.40 ≤ P-value ≤ 0.95). Infancy HDL was correlated with higher maternal HDL levels (0.01 < P ≤ 0.73). A higher number of maternal triglycerides was linked to a higher LDL in infancy (0.01 ≤ P-value≤ 0.55).

TABLE VIII. [A-D] ASSOCIATION BETWEEN THE MATERNAL LIPID LEVELS AND INFANCY LIPID LEVELS

A: MATERNAL LIPID LEVELS AND INFANCY TC LEVEL

	Infancy TC (one year below age)			
Exposure(Maternal)	***β***	***CI Low***	***CI High***	***P-value***
TC(Total cholesterol) (mg/dl))	0.38	-0.06	0.83	0.09
TG(Triglyceride) (mg/dl))	-0.14	-1.20	0.91	0.78
HDL (mg/dl)	0.08	-0.04	0.21	0.21
LDL (mg/dl)	0.02	-0.42	0.47	0.90

B:MATERNAL LIPID LEVELS AND INFANCY TG LEVEL

	Infancy TG (one year below age)			
Exposure(Maternal)	***β***	***CI Low***	***CI High***	***P-value***
TC(Total cholesterol) (mg/dl))	-0.10	-0.42	0.21	0.51
TG(Triglyceride) (mg/dl))	0.11	-0.63	0.85	0.76
HDL (mg/dl)	0.03	-0.05	0.12	0.40
LDL (mg/dl)	0.009	-0.30	0.32	0.95

C: MATERNAL LIPID LEVELS AND INFANCY HDL LEVEL

	Infancy HDL (one year below age)			
Exposure(Maternal)	***β***	***CI Low***	***CI High***	***P-value***
TC(Total cholesterol) (mg/dl))	0.52	-0.98	2.03	0.49
TG(Triglyceride) (mg/dl))	-2.43	-5.94	1.08	0.17
HDL (mg/dl)	0.50	0.93	0.08	0.01
LDL (mg/dl)	0.26	-1.23	1.75	0.73

D: MATERNAL LIPID LEVELS AND INFANCY LDL LEVEL

	Infancy LDL (one year below age)			
Exposure(Maternal)	***β***	***CI Low***	***CI High***	***P-value***
TC(Total cholesterol) (mg/dl))	-0.23	-0.85	0.37	0.44
TG(Triglyceride) (mg/dl))	-1.80	-3.20	-0.39	0.01
HDL (mg/dl)	-0.05	-0.22	0.12	0.55
LDL (mg/dl)	0.20	-0.40	0.80	0.51

In the next step, four different classification techniques, including random forest (RF), logistic regression (RL), support vector machine (SVM), naïve Bayes (NB), XGBOOST and DART-XGBOOST were used to the dataset. These were used to predict 0 (normal lipids) or 1 (abnormal lipids). The assessment metrics listed in Table IX were then obtained after testing each of the trained models using the remaining 20% of the test data. Figure 2 shown the different evaluation metrics using different classification models.

TABLE IX. EVALUATION METRICS FOR EACH CLASSIFIER.

Classification	**AUC**	**Accuracy**	**Sensitivity**	**Specificity**	**Precision**	**F1-score**
LR	0.900	0.9000	0.6666	1.0	0.875	0.9333
SVM	0.420	0.7000	0.7777	0.66	0.700	0.8235
RF	0.85	0.933	0.7777	1.0	0.9130	0.9545
NB	0.900	0.9000	0.6666	1.0	0.875	0.9333
XGBOOST	0.920	0.9666	0.8888	1.0	0.9545	0.9767
XGBOOST with DART	0.900	0.9000	0.95	0.78	0.9090	0.9302

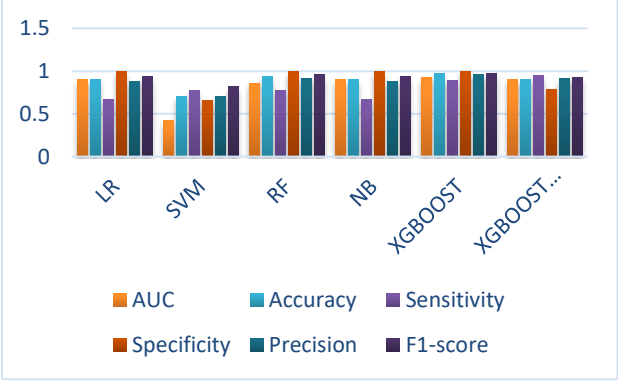
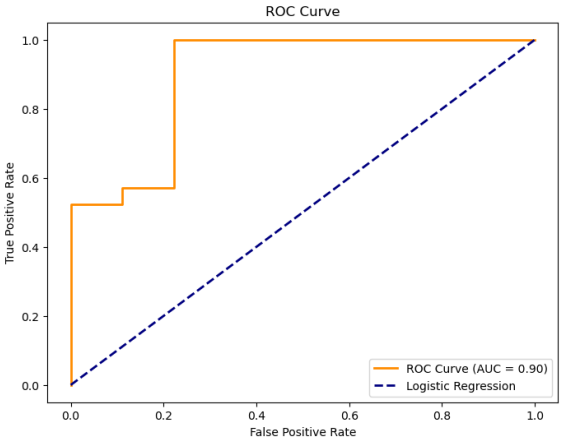


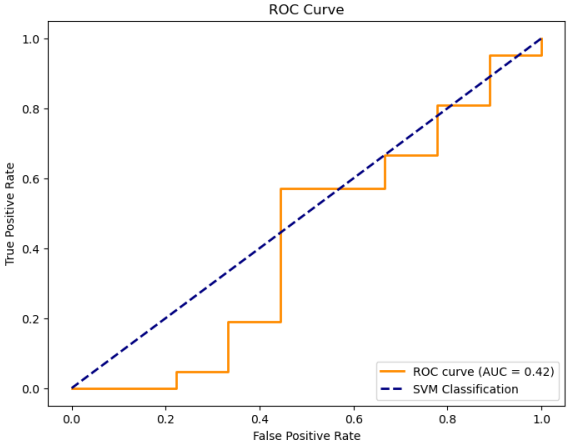
Figure 2. Comparision of different metrics of classification models

The AUC, accuracy and other metrics obtained during the models' implementation test are displayed in Figure 3. All classification methods, with the exception of SVM, are observed to perform better than 0.90 AUC and also identify the other matrices accuracy, sensitivity, specificity, precision, F1-sore. Compare all metrices, the best performing method is XGBOOST, it had accuracy is better than other models.

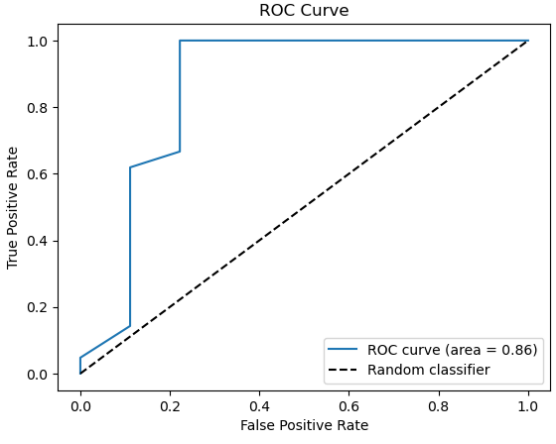
A: ROC Curve for Logistic Regression



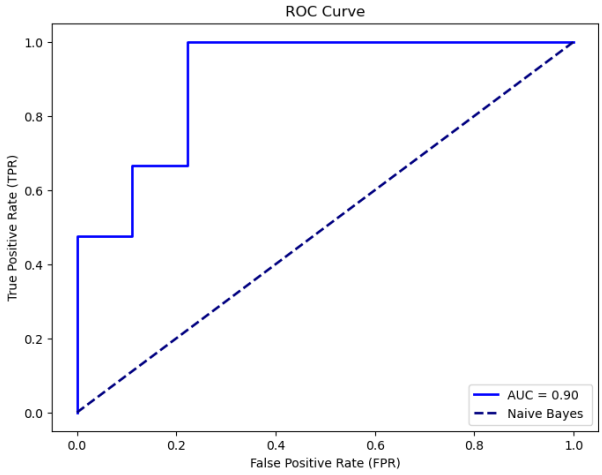
B: ROC Curve for SVM Classification



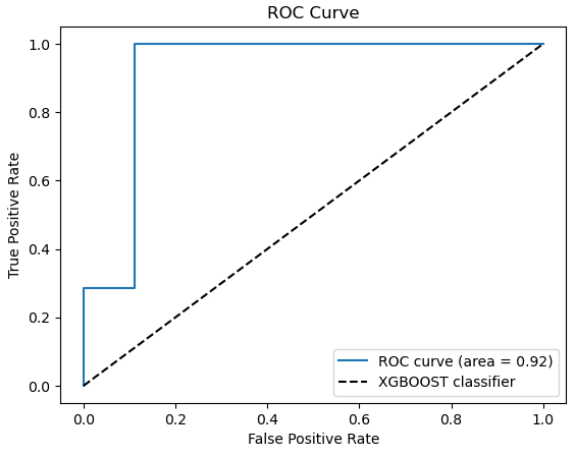
C: ROC Curve for Random Classifier



D: ROC Curve for Naïve Bayes Classifier



E: ROC Curve for XGBOOST



F: ROC Curve for DART XGBOOST Classifier

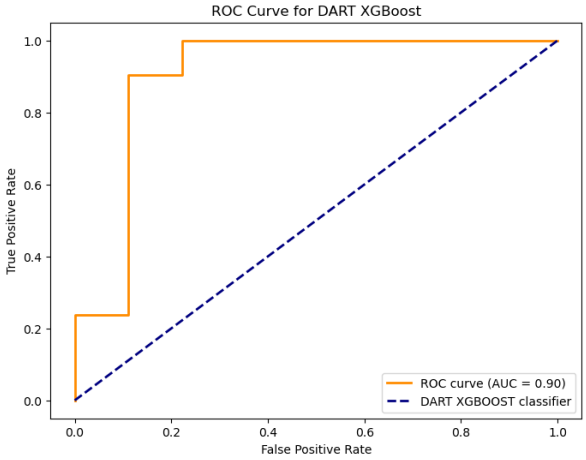


Figure 3. [A-B] ROC Curves of the different classification models

5. Conclusion and Future Work

In this study, we looked at routine blood lipid levels in infants and discovered that their data was linked to daASD. The study comprises one hundred fifty infants, of which 77 (451.3%) were younger than six months and 73 (48.6%) older than six months. The TC mean (± SD) in infants under the 6 months of age and infants older than 6 months of age did not differ statistically (P = 0.613). The LDL-cholesterol mean (±SD) in infants under 6 months and older children over 6 months did not differ statistically (P = 0.087). However, a statistically significant difference (P = 0.036) was observed in the HDL-cholesterol levels between the ≤6 months and >6 months age groups. As for the triglycerides, they were shown to be statistically different (P = 0.049) between the age groups of ≤6 months and >6 months.

Second, the results of the analysis of the data using linear spline mixed models are shown as the interaction between the lipid levels of the mother and the lipid levels of the infant. This study showed a positive association between the total cholesterol of the infant and the maternal level. The change in infant triglyceride levels did not correlate with maternal lipid levels. Maternal triglyceride level was linked to infancy LDL, and maternal HDL level was linked to infancy HDL. There was not any evidence that the lipid levels of mothers and infants were related. These results provide support to the theory that maternal womb effects could be the origin of the relationships. All models in this study provide accurate evaluations with an accuracy better than 0.90 and an AUC greater than 0.90. The XGBOOST approach give the best performance. These six models display the accuracy and AUC found during the models' use test data. All classification methods, with the exception of SVM, are shown to perform better than 0.9 AUC and accuracy. Finally, the lipid profile of children years later may be revealed by the maternal lipid levels during pregnancy. As a result, maternal lipid levels may serve as an early assessment of infant health effects. Observing these maternal lipid levels may provide an instance of opportunity for the initiation of early treatments with the goal of reducing the lipid levels of infants and possibly reducing their lifetime risk of ASD. It is necessary to conduct more research to determine the impact of genetics and environmental factors on a mother's lipid levels during pregnancy and the lipid levels of her infants ages later.

References

[1] Karthick, G., et al. "FORECAST THE AUTISM SPECTRUM DISORDER USING VARIOUS MACHINE LEARNING TECHNIQUES."

[2] M. F. Misman et al., "Classification of Adults with Autism Spectrum Disorder using Deep Neural Network," 2019 1st International Conference on Artificial Intelligence and Data Sciences (AiDAS), 2019,pp. 29-34, doi: 10.1109/AiDAS47888.2019.8970823.

[3] Baihua Li, Arjun Sharma, James Meng, SenthilPurushwalkam, and Emma Gowen. (2017) "Applying machine learning to identify autistic adults using imitation: An exploratory study." PloS one, 12(8): e0182652.

[4] A. Baranwal and M. Vanitha, "Autistic Spectrum Disorder Screening: Prediction with Machine Learning Models," 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), 2020, pp. 1-7, doi: 10.1109/ic-ETITE47903.2020.186.

[5] Vaishali, R., and R. Sasikala. "A machine learning based approach to classify Autism with optimum behaviour sets. (2018) " International Journal of Engineering & Technology 7(4): 18

[6] Butwicka, A., Långström, N., Larsson, H. *et al.* Increased Risk for Substance Use-Related Problems in Autism Spectrum Disorders: A Population-Based Cohort Study. *J Autism Dev Disord* **47**, 80–89 (2017). *https://doi.org/10.1007/s10803-016-2914-2*

[7] R. Reeta, G. Pavithra, V. Priyanka and J. S. Raghul, "Predicting Autism Using Naive Bayesian Classification Approach," 2018 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 2018, pp. 0109-0113, doi: 10.1109/ICCSP.2018.8524491.

[8] R. Bhuvaneswari and K. Kalaiselvi, "Naive Bayesian Classification Approach in Healthcare Applications", International Journal of Computer Science and Telecommunications, vol.3, January 2012.

[9] T. U. Mane, "Smart heart disease prediction system using Improved K-means and ID3 on big data," *2017 International Conference on Data Management, Analytics and Innovation (ICDMAI)*, Pune, India, 2017, pp. 239-245, doi: 10.1109/ICDMAI.2017.8073517.

[10] Syeda Roshni Ahmed et.al, Autism Spectrum Disorder Using Bernoulli's Naive Bayes

[11] Bone D, Bishop SL, Black MP, Goodwin MS, Lord C, Narayanan SS. Use of machine learning to improve autism screening and diagnostic instruments: effectiveness, efficiency, and multi-instrument fusion. J Child Psychol Psychiatry. 2016 Aug;57(8):927-37. doi: 10.1111/jcpp.12559. Epub 2016 Apr 19. PMID: 27090613; PMCID: PMC4958551.

[12] Küpper C, Stroth S, Wolff N, Hauck F, Kliewer N, Schad-Hansjosten T, Kamp-Becker I, Poustka L, Roessner V, Schultebraucks K, Roepke S. Identifying predictive features of autism spectrum disorders in a clinical sample of adolescents and adults using machine learning. Sci Rep. 2020 Mar 18;10(1):4805. doi: 10.1038/s41598-020-61607-w. PMID: 32188882; PMCID: PMC7080741.

[13] A pattern mixture model with long short-term memory network for acute kidney injury prediction, Journal of King Saud University - Computer and Information Sciences,Volume 35, Issue 4,2023,Pages 172-182,ISSN 1319-1578,https://doi.org/10.1016/j.jksuci.2023.03.007.

[14] Muhammad Kashif Hanif, Naba Ashraf, Muhammad Umer Sarwar, Deleli Mesay Adinew, Reehan Yaqoob, "Employing Machine Learning-Based Predictive Analytical Approaches to Classify Autism Spectrum Disorder Types", *Complexity*, vol. 2022, Article ID 8134018, 10 pages, 2022. https://doi.org/10.1155/2022/8134018

[15] Gök, Murat. "A novel machine learning model to predict autism spectrum disorders risk gene." *Neural Computing and Applications* 31.10 (2019): 6711-6717.

[16] Selvaraj, Shanthi, et al. "Autism spectrum disorder prediction using machine learning algorithms." *Computational Vision and Bio-Inspired Computing: ICCVBIC 2019*. Springer International Publishing, 2020.

[17] Christensen, Deborah L. "Prevalence 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 (2016).

[18] Pream, M. S. Sudha, and M. S. Vijaya, "Machine learning- based model for identification of syndromic autism spectrum disorder," in Integrated Intelligent Computing, Communication and Security Integrated Intelligent Computing, Communication and Security, pp. 141–148, Springer, Berlin,Germany,2019.

[19] Singh, Amanpreet, Narina Thakur, and Aakanksha Sharma. "A review of supervised machine learning algorithms." *2016 3rd international conference on computing for sustainable global development (INDIACom)*. Ieee, 2016.

[20] Jayalakshmi, V. Jalaja, V. Geetha, and R. Vivek. "Classification of autism spectrum disorder data using machine learning techniques." *International Journal of Engineering and Advanced Technology (IJEAT) ISSN* 8.6 (2019): 2249-8958.

[21] Mythili, M. S., and A. R. Mohamed. "An improved autism predictive mechanism among children using fuzzy cognitive map and feature extraction methods (FEAST)." *ARPN J Eng Appl Sci* 11 (2016): 1819-1828.

[22] Bram van den, "Using Machine Learning for Detection of Autism Spectrum Disorder", Journal of Autism and Developmental Disorders, 46(7):2317–2326, 2016.

[23] Kanimozhiselvi, C. S., D. Jayaprakash, and K. S. Kalaivani. "Grading autism children using machine learning techniques." *International Journal of Applied Engineering Research* 14.5 (2019): 1186-1188.

[24] Mashudi, Nurul Amirah, Norulhusna Ahmad, and Norliza Mohd Noor. "Classification of adult autistic spectrum disorder using machine learning approach." *IAES International Journal of Artificial Intelligence* 10.3 (2021): 743.

[25] Vaishali, R., and R. Sasikala. "A machine learning based approach to classify autism with optimum behaviour sets." *International Journal of Engineering & Technology* 7.4 (2018): 18.

[26] Vakadkar, K., Purkayastha, D. & Krishnan, D. Detection of Autism Spectrum Disorder in Children Using Machine Learning Techniques. *SN COMPUT. SCI.* **2**, 386 (2021). https://doi.org/10.1007/s42979-021-00776-5

[27] Reddy, V. Sai Krishna, et al. "Prediction on Cardiovascular disease using Decision tree and Naïve Bayes classifiers." *Journal of Physics: Conference Series*. Vol. 2161. No. 1. IOP Publishing, 2022.

Yuganadhar Bokka, Research Scholar, Department of CSE, Gandhi Institute of Engineering and Technology (GIET) University Gunupur, Odisha, India.

Dr. R.N.V.Jagan Mohan, Associate Professor of Computer Science and Engineering, Sagi Rama Krishnam Raju Engineering College, Bhimavaram. India.

Dr. M. Chandra Naik, Professor, Department of CSE, Gandhi Institute of Engineering and Technology (GIET) University Gunupur, Odisha, India.