International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 16, No.1 (Sep-24)

http://dx.doi.org/10.12785/ijcds/1601104

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]

[1,3]*CSE, GIET University, Odisha, India*
[2]*CSE, SRKR Engineering College, Bhimavaram, India*

Received 9 Mar. 2024, Revised 8 Jun. 2024, Accepted 9 Jun. 2024, Published 20 Sep. 2024

Abstract: Developmental disorders such as autism spectrum disorder (ASD) result from differences in brain development influenced by genetic, environmental, and prenatal factors. Early indicators of ASD include repetitive behaviors and social communication deficiencies. While gestational risk factors do not cause ASD, they can influence children's interactions and potentially affect ASD development. Changes in lipid levels at birth are linked to autism, with individuals with ASD often showing abnormal cholesterol and triglyceride levels compared to healthy controls. However, the predictive value of blood lipid profiles for ASD remains unclear. This study investigates the role of infant lipid levels in ASD development, considering maternal gestational risk factors. We developed a machine learning model using combined parental and childhood lipid levels to predict ASD. The model was validated with independent cohorts and tested against infant lipid profiles, employing various statistical approaches and multiple classifiers. Routine blood lipid levels were analyzed in 50 infants, with 77 youngers than six months and 73 older than six months. This analysis showed no statistical difference in total cholesterol or LDL cholesterol between infants under six months and older children over six months. However, significant differences were observed in HDL-cholesterol levels between the ≤6 and >6 month age groups. The analysis using linear spline mixed models showed a positive association between total cholesterol and maternal levels. The XGBoost model outperformed all other classifiers, achieving an AUC of 0.920, an accuracy of 0.9666, a specificity of 1.0, a sensitivity of 0.8888, an F1-score of 0.9767, and a precision of 0.9545. These findings suggest that specific lipid profiles at birth could serve as potential biomarkers for ASD.

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

1. Introduction

Autism Spectrum Disorder (ASD) is a neuro-developmental 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 [2]. 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 [3]. 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 [2]. 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 [4], [5], [6]. Predicting lipid levels and getting involved early can both ensure the health of

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

the 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. The goal of this work is to develop a machine learning model for autism spectrum disorder detection that is more accurate and improved.

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 "Materials and methods" section explains the system's operation and approach, as well as how it was implemented. 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 main goal is to look into and find limitations so that new, better, and more effective machine learning techniques may be developed for the prediction of autism spectrum disease.

The authors 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[1]. This paper uses an ASD screening dataset to analyze and predict probable cases in adults, children, and adolescents using machine learning methods like ANN, Random Forest, Logistic Regression, Decision Tree, and SVM[2]. 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[3]. Using a dataset of ASD diagnoses, this study examines how well machine learning models identify autism risk. Ten features are found by the model using a swarm intelligence-based firefly feature selection wrapper to differentiate ASD patients from non-ASD patients. Like the complete ASD diagnostic dataset, the results demonstrate an average accuracy of 92.12%-97.95% with optimal feature selections[4]. Autism spectrum disorder (ASD) is a type of autism that affects around 1% of the global population and can have long-term effects on social, cognitive, language, and speech development. Usually, the first two years following delivery are when symptoms start to show. Although early detection and therapy can help, ASD is primarily caused by genetic or environmental causes. Current diagnostic techniques, such clinical standardised testing, are expensive and time-consuming. Machine learning approaches are being utilised in addition to traditional methods to increase time and precision. To create prediction models based on outcomes, models like Support Vector Machines (SVM), Logistic Regression (LR), Random Forest Classifier (RFC), Naïve Bayes (NB), and KNN are applied to a dataset. The major goal is to expedite the diagnosis process by ascertaining whether a child is at risk for ASD in its early stages. The greatest accuracy for the chosen dataset is determined to be provided by logistic regression[5]. Big Data is a rapidly growing collection of data that requires processing and extraction of meaningful information. In the medical field, it is particularly useful in predicting heart disease, a major cause of death. The Hadoop Map Reduce platform is used for clustering, improving K-Means, and decision tree algorithms like ID3. This system aids in predicting parameters like chest pain, cholesterol, age, and resting Bp, improving clinical decision-making and treatment processes. Overall, Big Data is transforming the medical field[6]. This study focuses on the use of Deep Neural-Network (DNN) architecture in diagnosing Autism Spectrum Disorder (ASD) using adult screening data. The DNN model achieved 99.40% accuracy on the first dataset and 96.08% on the second, outperforming the Support Vector Machine (SVM) model. The results indicate that DNN can accurately identify ASD cases using adult screening data, demonstrating its potential in machine learning applications [7]. The purpose of this study was to see whether kinematic traits in the movement patterns of autistic people might be identified using machine learning. To replicate hand movements, the study analysed data from 16 ASC individuals and 14 age- and IQ-matched controls. The study showed that it is feasible to use machine learning techniques to high-dimensional data by identifying two ideal imitation circumstances and the nine most important kinematic parameters. The results point to the possibility of using machine learning to develop kinematic indicators that could aid in the diagnosis of autism[8].

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[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]. In this study, machine learning is used to enhance diagnostic and screening instruments for autism spectrum disorder (ASD). The scores from 1,264 and 462 people with and without ASD on the Social Responsiveness Scale and Autism Diagnostic Interview-Revised were used by the researchers. The robust classifier was used to construct the algorithms, which were later shown to be more efficient, customizable, and successful. With 89.2% sensitivity and 59.0% specificity, the screener algorithm was produced by combining ADI-R and SRS using machine learning[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 models 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, and 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]. A machine learning model has been developed to predict ASD risk genes using brain developmental gene expression data. The model uses Haar-wavelet-transform, discretization methods, and Bayes-network learning algorithm [15]. The study aims to predict autism spectrum disorder (ASD) in children using machine-learning algorithms. These algorithms have recently been used to improve the accuracy. The study compares and enhances the effectiveness of various feature selection algorithms. The Random Tree classification algorithm, based on Extra Tree calculation, outperforms the other methods. The results are evaluated using accuracy, recall, and precisionWith the use of the machine learning model, the study determines which characteristics are most important in helping to differentiate ASD in toddlers [16]. According to estimates, 14.6 out of every 1,000 eight-year-old children at 11 ADDM Network sites had autism spectrum disorder (ASD) in 2012. Compared to girls, boys were more common than girls. Compared to non-Hispanic black children and Hispanic children, non-Hispanic white children had a greater prevalence (15.5 per 1,000). There was significant variation in the prevalence across the sites, with greater rates in the monitoring sites where health and educational records were examined. Among children diagnosed with ASD, 82% had received a prior diagnosis or educational classification; this rate was lower for children of Hispanic descent (78%). With 43% having an evaluation by the age of 36 months, 40 months was the average age at the earliest documented complete examination[17]. Autism spectrum disorder (ASD) is a genetic disorder with various symptoms and disability levels. Mutations are key contributors to ASD, and developing targeted therapies is crucial. This chapter uses supervised machine learning techniques to identify syndromic ASD by classifying mutations. By employing SVM and decision trees, they achieved 98% and 94% accuracy, respectively[18].

Supervised machine learning is a method that uses external instances to predict future outcomes, focusing on categorizing data from prior information. It is commonly used in data science problems and has been compared to other techniques like rule-based, logic-based, instance-based, and stochastic techniques. The paper aims to evaluate the effectiveness of these algorithms in terms of accuracy, speed, complexity, and overfit risk[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, Decision Tree, Naïve-Bayes, 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 categorized with or not have ASD disease[24]. According to the results, CNN produce better accuracy[25]. utilizing 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]. 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[28]. The study looked at how a mother's pregnancy-related cholesterol levels affected the lipid levels in her offspring. The study discovered that children of mothers with high cholesterol had elevated triglyceride and very low-density lipoprotein (VLDL) levels in comparison to children of mothers without high cholesterol. There were no appreciable alterations in placental weight or umbilical cord length, although maternal hypercholesterolemia (MHC) also caused poor neonatal birthweight and placental efficiency. According to the study, on the second postpartum day, mothers' MHC increased the lipid levels of their newborns while decreasing the placental efficiency and baby birthweight. The mother's and the child's repeated assessments of their cholesterol levels were evaluated using the two-way ANOVA with Turkey's correction. Nevertheless, for the course of pregnancy, there

was no discernible difference in the groups' TG and VLDL levels[29]. Less wealthy newborn lipid profiles are associated with maternal dyslipidemia during pregnancy. Babies' lipid levels are impacted by the mother's dyslipidemia during pregnancy. examined expectant mothers at 24 and 36 weeks of pregnancy looked into kids who were three, six, and twelve months old. As non-parametric tests of independent samples, the Kruskal-Wallis and Wilcoxon-Mann-Whitney U tests were employed. Multivariate regressions were used in addition to the traditional correlation analyses to correct one or more independent variables. The Winkler score and the mother's BMI were selected as independent factors for the multivariate regression analysis, while the maternal lipid parameters were selected as dependent variables. This was done since the univariate regression studies revealed significant interdependencies between the two variables. The significance threshold was set at a value of less than 5%[30].

3. Materials and Methods

A. Classification

We have used various classification models such as logistic regression, support vector machines, random forest, and naive Bayes. A supervised learning technique called a classification model makes an effort to predict class labels. In this scenario, it aims to identify 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.

$$\phi(Z) = \frac{1}{1 + e^{-z}} \tag{1}$$

Random forests (RF) offer the advantage of having a lower chance of overfitting because they are based on several decision trees. 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.

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

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). As seen by the class values, proper classification can generally be achieved with two classes and a well-designed hyperplane that best divides them or whichever has the greatest distance between them.

$$W * X + b = 0 \tag{3}$$

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.

$$P(\frac{A}{B}) = \frac{P(\frac{B}{A}) * P(A)}{P(B)} \tag{4}$$

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.

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

The Algorithm 1 shows the pseudo code for proposed model.

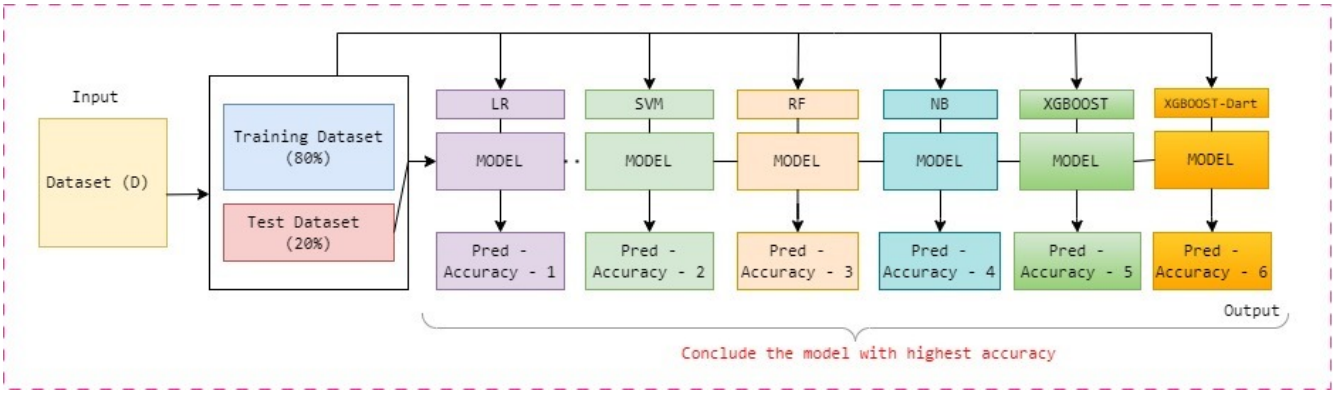


Figure 1. Model Architecture prediction of ASD

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 = pd.read_csv('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 predections 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.

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.

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.

$$Accuracy = \frac{\sum_i cm_{i_i}}{\sum_i \sum_j cm_{i_j}} \tag{5}$$

Equation (6) is used to find the true positive rate (Sensitivity), or the probability that a non-typical patient will have a positive test result. TP stands for true positive cases, and FN for false negatives.

$$Sensitivity = \frac{\sum_i cm_{i_i}}{\sum_j cm_{i_j}} \tag{6}$$

The probability that a normal patient would have a negative test result is represented by equation (7), which is often referred to as specificity or true negative rate.

$$specificity = \frac{\sum_i cm_{j_i}}{\sum_j cm_{i_j}} \tag{7}$$

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

$$Precision = \frac{\sum_i cm_{i_i}}{\sum_i \sum_j cm_{j_i}} \tag{8}$$

$$F1 - score = \frac{2}{\frac{1}{precision} + \frac{1}{Recall}} \tag{9}$$

TABLE I. Dataset 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

TABLE II. Confusion matrix for binary classification

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

TABLE III. Confusion 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

TABLE IV. Infancy Lipid Parameters Normal-distributed (n=150)

Parameters	Mean ± SD (mg/dl)	Normal-distribution(mg/dl) Percentile (25%)	Percentile (50%)	Percentile (75%)	Percentile (90%)	IQR (mg/dl)
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 comparison of ROC and AUC results can also be used to evaluate 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, looked into the link between maternal lipids and children's 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 are 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.

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.

The mother's lipid levels during her pregnancy and the baby's corresponding lipid levels six and twelve months after delivery are positively correlated. 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)).

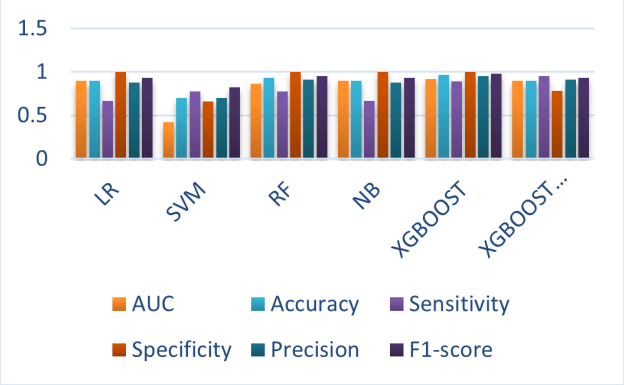


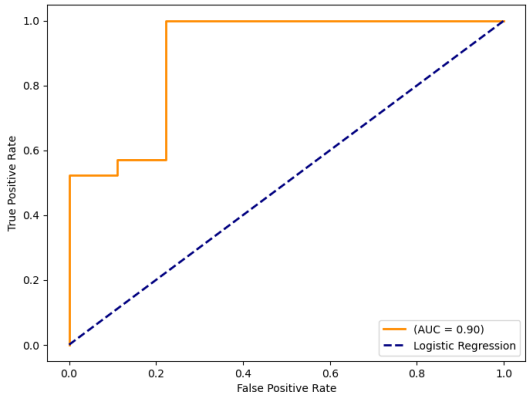
Figure 2. Comparison of different metrics of classification models

The lipid profiles of infants at 6 and 12 months old as well as those of pregnant women are displayed in Table VII. Triglycerides and low-density lipoprotein (LDL) levels significantly decline at 12 months, whereas total cholesterol and HDL levels significantly increase.

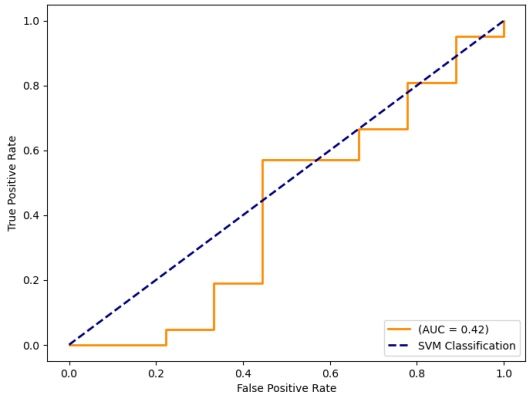
The relationship between mother lipid levels and infant lipid levels is shown as a result of the data analysis using linear spline mixed models. The relationship between mother lipid levels and infant lipid levels from birth to age twelve is displayed in Table VIII. The p-values from the interaction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

In the next step, four different classification techniques, including logistic regression (RL), support vector machine (SVM), random forest (RF), naive 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.

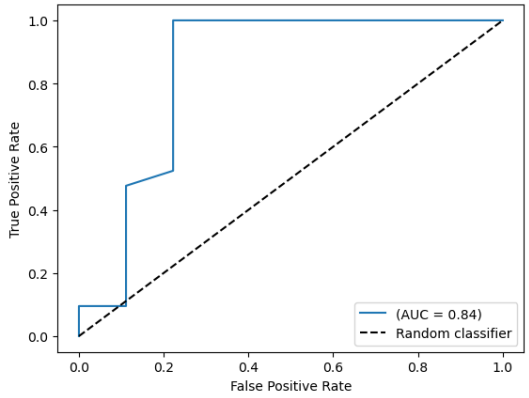
The AUC, accuracy and other metrics obtained during the models' implementation test are displayed in Figure 3. All classification methods, with the exception of SVM, are observed to perform better than 0.90 AUC and also identify the other matrices accuracy, sensitivity, specificity, preci-



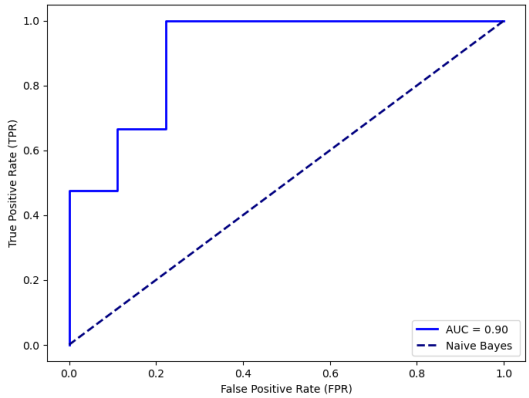
A: ROC Curve for Logistic Regression



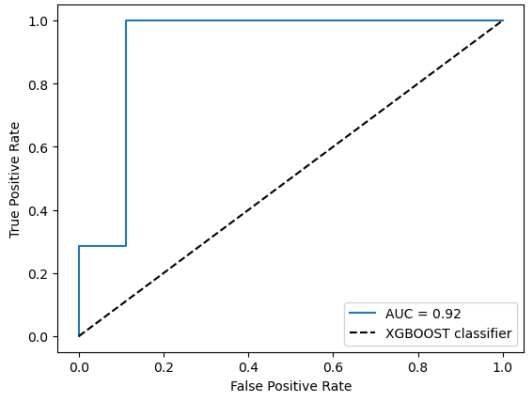
B: ROC Curve for SVM Classification



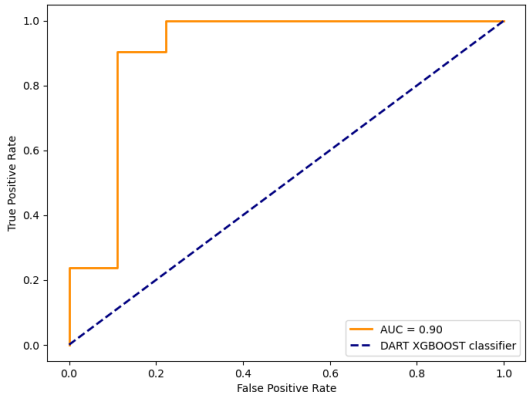
C: ROC Curve for Random Classifier



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

TABLE VI. Basic features of Maternal during pregnancy and infancy

Features	During Pregnancy (n=150) Mean±SD	Infancy below 6months (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 (mg/dl)	76.79±22.48 (mg/dl)	83.53±22.39 (mg/dl)	0.08
Triglyceride (mg/dl)	230.40±121.42(mg/dl)	48.27±4.87 (mg/dl)	46.68±4.55 (mg/dl)	0.59
HDL-cholesterol (mg/dl)	68.36±13.92(mg/dl)	102.93±11.30 (mg/dl)	102.61±12.24 (mg/dl)	0.09
LDL-cholesterol (mg/dl)	143.00±44.78 (mg/dl)	147.98±15.53 (mg/dl)	146.76±16.58 (mg/dl)	0.86

sion, F1-sore. Compare all metrices, the best performing method is XGBOOST, it had accuracy is better than other models.

In our proposed architecture, two classes of the dataset are considered. Existing work using screen datasets and proposed work using lipid profiles. Table. X shows the accuracies obtained by different classifiers during the experimentation. It clearly states that the XGBOOST model outperforms with an accuracy of 96.66%.

5. Conclusions 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. Routine blood lipid profiles may not accurately represent all lipid metabolism abnormalities linked to ASD, potentially missing crucial biomarkers. Expanding lipid

TABLE VIII. [A-D] Association between the maternal lipid levels and infancy lipid levels

A: MATERNAL LIPID LEVELS (mg/dl) AND INFANCY TC LEVEL (mg/dl)

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

B: MATERNAL LIPID LEVELS (mg/dl) AND INFANCY TG LEVEL (mg/dl)

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

C: MATERNAL LIPID LEVELS (mg/dl) AND INFANCY HDL LEVEL (mg/dl)

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

D: MATERNAL LIPID LEVELS (mg/dl) AND INFANCY LDL LEVEL (mg/dl)

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

TABLE IX. Model's Evaluation metrics for other classifiers

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

TABLE X. Model Accuracy (using Routine Blood lipids profiles) comparison with existing works

S.No	Authors	Methodology	Accuracy
1	Baranwal et.al. [2]	Decision Tree	0.8226
		Random Forest	0.8226
		Support Vector Machine	0.9677
		Logistic Regression	0.9032
		Naïve Bayes	0.9315
2	Vaishali et.al. [4]	Decision Tree	0.9110
		Support Vector Machine	0.9966
		K- Nearest Neighbour	0.8767
		Logistic Regression	0.9715
3	Vakadkar et.al [5]	Naïve Bayes	0.9479
		Support Vector Machine	0.9384
		Random Forest	0.8152
		K- Nearest Neighbour	0.9052
		Logistic-Regression	0.9000
		Support Vector Machine	0.7000
4	**Proposed Model**	Random Forest	0.9333
		Naïve Bayes	0.9000
		XGBOOST with DART	0.9000
		XGBOOST	**0.9666**

profiling and using advanced lipidomic techniques could help identify novel lipid biomarkers associated with daASD. 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] R. J. D. S. N. U. G Karthick, N Venkateswaran, "Forecast the autism spectrum disorder using various machine learning techniques," *European Chemical Bulletin*, p. 6030 – 6037, 2023.

[2] M. V. A Baranwal, "Autistic spectrum disorder screening: Prediction with machine learning models," *International Conference on Emerging Trends in Information Technology and Engineering*, pp. 1–7, 2020.

[3] R. Bhuvaneswari and K. Kalaiselvi, "Naive bayesian classification approach in healthcare applications," *International Journal of Computer Science and Telecommunications*, pp. 106–112, 2012.

[4] R. S. R Vaishali, "A machine learning based approach to classify autism with optimum behaviour sets," *International Journal of Engineering Technology*, pp. 1–6, 2018.

[5] K. Vakadkar, D. Purkayastha, and D. Krishnan, "Detection of autism spectrum disorder in children using machine learning techniques," *SN computer science*, pp. 1–9, 2021.

[6] 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)*, pp. 239–245, 2017.

[7] F. E. H. M. Z. S. H. H. M. H. MF Misman, AA Samah, "Classification of adults with autism spectrum disorder using deep neural network," *International Conference on Artificial Intelligence and Data Sciences*, pp. 29–34, 2019.

[8] J. M. S. P. E. G. Baihua Li, Arjun Sharma, "Applying machine learning to identify autistic adults using imitation: An exploratory study," *PLOS ONE*, pp. 1–19, 2017.

[9] R. Reeta, G. Pavithra, V. Priyanka, and J. Raghul, "Predicting autism using naive bayesian classification approach," *2018 International Conference on Communication and Signal Processing (ICCSP)*, pp. 0109–0113, 2018.

[10] V. I. M. T. V. Syeda Roshni Ahmed, Suvarna A B, "Autism spectrum disorder using bernoulli's naive bayes," *Journal of engineering sciences*, pp. 776–781, 2020.

[11] D. Bone, S. L. Bishop, M. P. Black, M. S. Goodwin, C. Lord, and S. S. Narayanan, "Use of machine learning to improve autism screening and diagnostic instruments: effectiveness, efficiency, and multi-instrument fusion," *Journal of Child Psychology and Psychiatry*, pp. 927–937, 2016.

[12] C. Küpper, S. Stroth, N. Wolff, F. Hauck, N. Kliewer, T. Schad-Hansjosten, I. Kamp-Becker, L. Poustka, V. Roessner, K. Schultebraucks *et al.*, "Identifying predictive features of autism spectrum disorders in a clinical sample of adolescents and adults using machine learning," *Scientific reports*, p. 4805, 2020.

[13] M. F. Begum and S. Narayan, "A pattern mixture model with long short-term memory network for acute kidney injury prediction," *Journal of King Saud University-Computer and Information Sciences*, pp. 172–182, 2023.

[14] M. K. Hanif, N. Ashraf, M. U. Sarwar, D. M. Adinew, and R. Yaqoob, "Employing machine learning-based predictive analytical approaches to classify autism spectrum disorder types," *Complexity*, pp. 1–10, 2022.

[15] M. Gök, "A novel machine learning model to predict autism spectrum disorders risk gene," *Neural Computing and Applications*, pp. 6711–6717, 2019.

[16] S. Selvaraj, P. Palanisamy, S. Parveen, and Monisha, "Autism spectrum disorder prediction using machine learning algorithms," *Computational Vision and Bio-Inspired Computing: ICCVBIC 2019*, pp. 496–503, 2020.

[17] D. L. Christensen, "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*, pp. 1–23, 2016.

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

[19] A. Singh, N. Thakur, and A. Sharma, "A review of supervised machine learning algorithms," *2016 3rd international conference on computing for sustainable global development (INDIACom)*, pp. 1310–1315, 2016.

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

[21] M. Mythili and A. Mohamed, "An improved autism predictive mechanism among children using fuzzy cognitive map and feature extraction methods (feast)," *ARPN Journal of Engineering and Applied Sciences*, pp. 1819–1828, 2016.

[22] B. van den Bekerom, "Using machine learning for detection of autism spectrum disorder," *Proc. 20th Student Conf. IT*, pp. 1–7, 2017.

[23] C. Kanimozhiselvi, D. Jayaprakash, and K. Kalaivani, "Grading autism children using machine learning techniques," *International Journal of Applied Engineering Research*, pp. 1186–1188, 2019.

[24] N. A. Mashudi, N. Ahmad, and N. M. Noor, "Classification of adult autistic spectrum disorder using machine learning approach," *IAES International Journal of Artificial Intelligence*, p. 743, 2021.

[25] S. Raj and S. Masood, "Analysis and detection of autism spectrum disorder using machine learning techniques," *Procedia Computer Science*, vol. 167, pp. 994–1004, 2020.

[26] K. Vakadkar, D. Purkayastha, and D. Krishnan, "Detection of autism spectrum disorder in children using machine learning techniques," *SN computer science*, pp. 1–9, 2021.

[27] V. S. K. Reddy, P. Meghana, N. S. Reddy, and B. A. Rao, "Prediction on cardiovascular disease using decision tree and naïve bayes classifiers," *Journal of physics: conference series*, vol. 2161, pp. 1–7, 2022.

[28] A. Butwicka, N. Långström, H. Larsson, S. Lundström, E. Serlachius, C. Almqvist, L. Frisén, and P. Lichtenstein, "Increased risk for substance use-related problems in autism spectrum disorders: A population-based cohort study," *Journal of autism and developmental disorders*, pp. 80–89, 2017.

[29] V. Jayalekshmi, R. Jagannath, S. Sreelekshmi, P. Rafeekha, R. Vidyalekshmy, and S. Ramachandran, "Maternal hypercholesterolemia during gestation is associated with elevated lipid levels of two-day-old neonates," *Molecular and Cellular Biochemistry*, pp. 363–371, 2024.

[30] A. Dathan-Stumpf, M. Vogel, N. Grafe, W. Kiess, and H. Stepan, "Relation between socioeconomic status and maternal serum lipids to infant lipid concentrations and anthropometry in the first year of life," *Archives of Gynecology and Obstetrics*, pp. 1009–1020, 2024.

Yugandhar 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, Department of CSE, Sagi Rama Krishna Raju Engineering College, Bhimavaram, India.

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