International Journal of Recent Engineering Research and Development (IJRERD)
ISSN: 2455-8761
www.ijrerd.com || Volume 09 – Issue 04 || Jul - Aug 2024 || PP. 09-13

# Infant Birth Weight Prediction using ML

**Umabharathi G, Abitha P,**
*UG Scholar, Department of CSE, A.R.J College of Engineering and Technology,
Mannargudi, Thiruvarur, Tamilnadu-614 001, India*

**Pathmapriya P,**
*Assistant Professor, Department of CSE, A.R.J College of Engineering and Technology,
Mannargudi, Thiruvarur, Tamilnadu-614 001, India*

**Abstract:** Accurate estimation of infant birth weight is essential for effective neonatal care and maternal health management. In this study, we present a data driven approach utilizing machine learning techniques, specifically the linear regression algorithm, to predict infant birth weight. Our diverse dataset incorporates maternal health indicators, demographic factors, and prenatal characteristics to develop robust predictive models. Key features such as maternal age, gestational age, prenatal care visits, maternal BMI, and other pertinent factors are considered in our model, aiming to accurately forecast birth weight. By leveraging the linear regression algorithm, we aim to enhance prenatal care and inform clinical decision-making, thereby contributing to improved maternal and infant health outcomes. Our study demonstrates the potential of machine learning in providing accurate birth weight estimates, which can assist healthcare professionals in better managing pregnancies and ensuring the well-being of both mothers and infants. This approach offers a valuable tool for optimizing prenatal care strategies and addressing potential risk factors, ultimately leading to better health outcomes for mothers and their newborns.

**Keywords:** Machine learning, linear regression algorithm, BMI.

## 1. Introduction

Predicting infant birth weight is essential for ensuring optimal neonatal care and maternal health. Accurate estimation of birth weight enables healthcare providers to identify potential risks during pregnancy and plan appropriate interventions to mitigate them. In this paper, we present a data-driven approach to predicting infant birth weight using machine learning techniques. Our study focuses on employing the linear regression algorithm, a widely-used method in predictive modeling, to develop robust predictive models. By analyzing a diverse dataset comprising maternal health indicators, demographic factors, and prenatal characteristics, we aim to provide accurate forecasts of birth weight. The features considered in our model include maternal age, gestational age, prenatal care visits, maternal BMI, and other pertinent factors that have been shown to influence birth weight. Through this research, we seek to enhance prenatal care and inform clinical decision-making, ultimately contributing to improved maternal and infant health outcomes.

## 2. Literature Review

Najmus Sakib Borson, et.al:(2020), Correlation analysis of demographic factors on low birth weight and prediction is described as according to a national survey in Bangladesh, a south Asian country, approximately 22.6 percent of the new born babies are born with low birth weight (below 2.5 kg or 2500 grams) . There are some key factors regarding low birth weight which are clinically recognized but apart from the clincical perspective some other health and demographic factors also play a vital role in this phenomenon which can be directly or indirectly associated. The purpose of this study is to utilize the potential machine learning algorithms to construct a predictive model for low birth weight given some health and demographic data related to neonatal health condition in the context of Bangladesh. For the predictive analysis, algorithms like Logistic Regression, Naïve Bayes, Random Forest, K-Nearest Neighbor, Support Vector Machine, Neural Network models have been used in the study. The findings of this study can be a guideline for the health professionals as well as the researchers for analyzing low birth weight infants which can help the people in mass to understand and take necessary precautions to avoid of any such event where a child is born with weight less than the average.

Dr. Suvarna Nandyal, et.al:(2021), An Approach for Fetal Weight Estimation using Machine is described as women's safety is important in society. So here's a model that will assist them during their pregnancy. Uneducated women have a lower level of pregnancy health awareness. Knowing the fetus weight at correct time of pregnancy is much important. According to the World Health Organization (WHO), the range of Low Birth Weight (LBW) is less than 2500g, the range of High Birth Weight (HBW) is greater than 4000g, and the range of Normal Weight is between 2500g and 4000g.According to weight of fetus it may undergo the disease of which some may be life taking ones. Hence to find out these traumas as soon as possible, we are building a Machine Learning model wherein we use two algorithms as methods one is Convolution Neural Network(CNN) and Multiclass SVM algorithm, which predicts the estimated weight and disease if suffering according. This helps clinicians to identify risks in fetus whose effect is usually to carring mother. As a result we found that multiclass SVM outperforms the best compared to CNN with approximate accuracy with 95%.

Wasif Khan, et.al:(2022), Infant birth weight estimation and low birth weight is described as accurate prediction of a newborn's birth weight (BW) is a crucial determinant to evaluate the newborn's health and safety. Infants with low Birth Weight (LBW) are at a higher risk of serious short- and long-term health outcomes. Over the past decade, machine learning (ML) techniques have shown a successful breakthrough in the field of medical diagnostics. Various automated systems have been proposed that use maternal features for LBW prediction. However, each proposed system uses different maternal features for LBW classification and estimation. Therefore, this paper provides a detailed setup for BW estimation and LBW classification. Multiple subsets of features were combined to perform predictions with and without feature selection techniques. Furthermore, the synthetic minority oversampling technique was employed to oversample the minority class. The performance of 30 ML algorithms was evaluated for both infant BW estimation and LBW classification. Experiments were performed on a self-created dataset with 88 features. The dataset was obtained from 821 women from three hospitals in the United Arab Emirates. Different performance metrics, such as mean absolute error and mean absolute percent error, were used for BW estimation. Accuracy, precision, recall, F-scores, and confusion matrices were used for LBW classification.

## 3. Proposed System

The proposed system is designed to harness a comprehensive dataset encompassing various maternal and prenatal factors crucial for accurate infant birth weight prediction. This dataset includes essential parameters such as maternal age, weight gain during pregnancy, gestational age, and prenatal care history. Leveraging advanced machine learning techniques, particularly linear regression, our system aims to develop robust predictive models capable of forecasting infant birth weight with high precision. By analyzing the relationship between these maternal and prenatal factors and birth weight, our system provides healthcare professionals with valuable insights into potential birth outcomes, enabling them to tailor prenatal care strategies effectively. Through the integration of machine learning algorithms and the collection of pertinent maternal and prenatal data, our proposed system seeks to revolutionize neonatal care and maternal health management. By accurately predicting infant birth weight, healthcare providers can proactively identify and address potential risks during pregnancy, thereby improving outcomes for both mothers and infants. This data-driven approach empowers clinicians to make informed decisions and implement personalized interventions, ultimately leading to enhanced maternal and neonatal health outcomes. The system begins by collecting a comprehensive dataset comprising various maternal and prenatal factors essential for predicting infant birth weight. This dataset undergoes thorough preprocessing, including handling missing values, removing outliers, and standardizing numerical features to ensure data quality. Next, feature extraction is performed to identify relevant predictors such as gestational parity, maternal age, height, weight, smoking status, and BMI. Once the dataset is prepared, the linear regression algorithm is employed for model training. This involves fitting the model to the training data, where it learns the relationships between the input features and the target variable – infant birth 20 weight. The dataset is then split into training and testing sets to evaluate the model's performance. The trained model parameters are stored in a database for future use. In the testing phase, users can input details such as gestational age, parity, maternal age, height, weight, smoking status, and BMI for a specific pregnancy. These input data are fed into the trained model, which predicts the baby's weight. This predictive capability enables healthcare professionals to assess potential birth outcomes and guide appropriate prenatal care strategies based on individual patient characteristics, ultimately leading to improved maternal and neonatal health outcomes.

## 4. System Architecture

**Data Collection and Pre-processing:** Gather a dataset comprising maternal and prenatal factors, including maternal age, weight gain during pregnancy, gestational age, and prenatal care history.

**Feature Extraction:** Identify relevant predictors such as gestation, parity, maternal age, height, weight, smoking status, and BMI from the dataset.

**Model Training:** Fit a linear regression model to the training data using the identified features as independent variables and infant birth weight as the dependent variable.

**Training and Splitting of Dataset:** Split the dataset into training and testing sets to evaluate the model's performance. The training set is used to train the model, while the testing set is kept separate for evaluation.

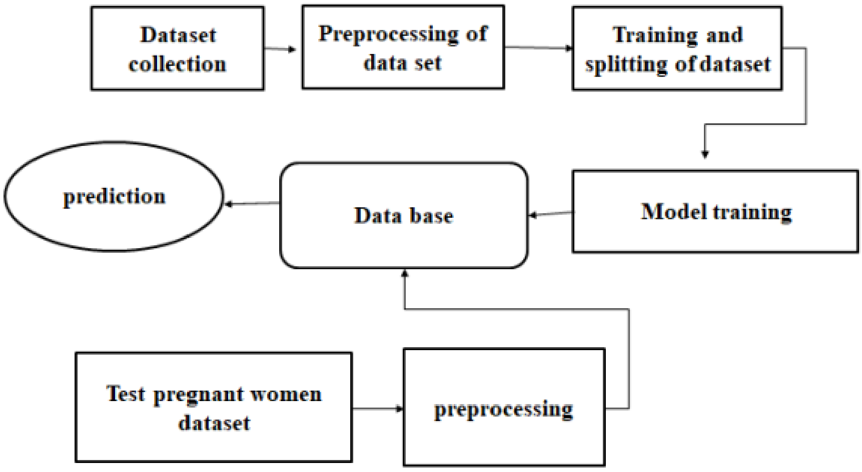


Figure 1: System Architecture

**Test Input Data**: Users input details such as gestation, parity, maternal age, height, weight, smoking status, and BMI for a specific pregnancy.

**Prediction:** The trained linear regression model utilizes the input data to predict the baby's weight. The model applies the learned coefficients to the input features to calculate the predicted birth weight.

**Output:** The system provides the predicted baby weight based on the input data, allowing healthcare professionals to assess potential birth outcomes and guide prenatal care strategies.

## 5. System Implementation

**Data Collection and Preprocessing:**

In the data collection and preprocessing module, we gather a diverse dataset comprising maternal and prenatal factors such as gestation, parity, age, height, weight, smoking status, and BMI. This dataset is obtained from various sources, including healthcare records and surveys. Preprocessing involves handling missing values, removing outliers, and standardizing numerical features to ensure data quality and consistency. This step is crucial for preparing the dataset for further analysis and modelling.

**Feature Extraction:**

The feature extraction module identifies relevant predictors from the pre-processed dataset. Key features such as gestation, parity, maternal age, height, weight, smoking status, and BMI are extracted to serve as independent variables for the linear regression model. These features are chosen based on their known associations with infant birth weight and their potential to influence the outcome.

**Linear Regression Model Training:**

In the linear regression model training module, the extracted features are used to train the predictive model. The linear regression algorithm is applied to the dataset, where it learns the relationships between the input features and the target variable, which is infant birth weight. The model is trained to minimize the difference between the predicted and actual birth weights by adjusting the coefficients.

**Training and Splitting of Dataset:**

Before training the model, the dataset is split into training and testing sets. The training set is used to train the model, while the testing set is kept separate for evaluation purposes. This ensures that the model's performance can be accurately assessed on unseen data and helps prevent overfitting.

**Database Storage:**

Once the model is trained, the trained model parameters are stored in a database for future use. This database serves as a repository for the model, allowing easy access and retrieval when making predictions on new input data.

**User Interface:**

The user interface module provides a platform for users to input details such as gestation, parity, age, height, weight, smoking status, and BMI for a specific pregnancy. This interface allows users, such as healthcare professionals or expectant mothers, to interact with the system and obtain predictions for infant birth weight.

**Testing and Analysis:**

In the testing and analysis module, the trained model is evaluated using the testing set to assess its performance and accuracy. Various metrics, such as mean squared error or R-squared, may be used to evaluate the model's performance and identify any areas for improvement.

**Prediction Module:**

The prediction module utilizes the trained linear regression model and the input data provided by the user to predict the baby's weight. The model applies the learned coefficients to the input features to calculate the predicted birth weight, providing valuable insights into potential birth outcomes.

## 6. Result

Developing a machine learning model to predict infant birth weight involves a systematic approach from data collection to model deployment. Initially, comprehensive datasets encompassing maternal health records, prenatal care details, and potentially genetic factors are gathered and preprocessed. This includes handling missing data, normalizing numerical features, and encoding categorical variables. Feature selection is crucial, focusing on factors like maternal age, weight gain during pregnancy, gestational age, and other clinically significant variables.

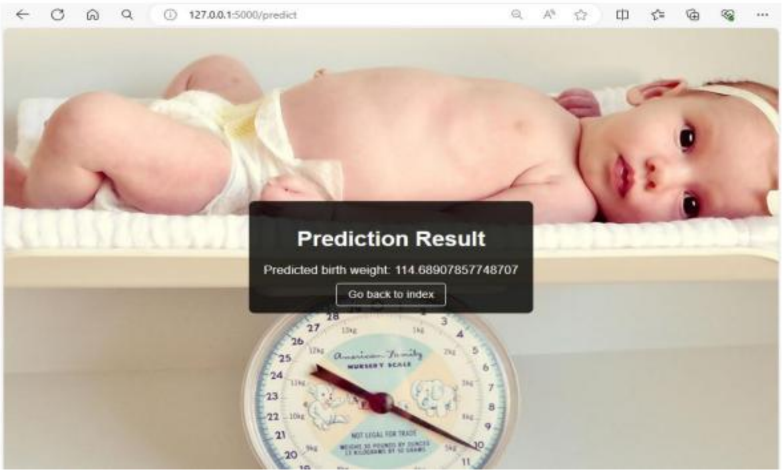


Figure 2: Result Image

Various regression models such as linear regression, decision trees, random forests, or gradient boosting models are evaluated for their predictive performance using metrics like Mean Absolute Error (MAE) or Root Mean Squared Error (RMSE). Through rigorous training and validation on split datasets, models are fine-tuned using techniques like grid search to optimize hyperparameters. Once a suitable model is selected based on validation performance, it is deployed to predict birth weights for new data. Continuous monitoring and periodic retraining ensure the model remains accurate and relevant in predicting infant birth weights effectively in clinical and research settings.

## 7. Conclusion

In conclusion, our study presents a data-driven approach to predicting infant birth weight using machine learning techniques, specifically linear regression. Through the analysis of a diverse dataset comprising maternal and prenatal factors, including maternal age, weight gain during pregnancy, gestational age, and prenatal care history, we have developed robust predictive models. Our findings demonstrate the effectiveness of the linear regression algorithm in accurately forecasting birth weight based on these factors. By leveraging this predictive capability, healthcare professionals can proactively identify and address potential risks during pregnancy, ultimately leading to improved neonatal care and maternal health outcomes. The system's ability to provide personalized predictions based on individual patient characteristics allows for tailored prenatal care strategies, ensuring that interventions are targeted and effective. Moreover, the simplicity and interpretability of the linear regression algorithm make it accessible and practical for use in clinical settings. Overall, our research highlights the importance of data-driven approaches in enhancing prenatal care and informing clinical decision-making. By leveraging machine learning techniques like linear regression, we can empower healthcare providers with valuable insights into potential birth outcomes, ultimately improving the well-being of both mothers and infants.

## 8. References


[1]. B. Y. Ooi, W. L. Beh, W. Lee, and S. Shirmohammadi, “A Parameter-Free Vibration Analysis Solution for Legacy Manufacturing Machines ’ Operation Tracking,” IEEE Internet Things J. (Early Access), 2020, doi: 10.1109/JIOT.2020.2994395.

[2]. G. Nota, F. D. Nota, D. Peluso, and A. T. Lazo, “Energy efficiency in Industry 4.0: The case of batch production processes,” Sustain., vol. 12, no. 16, 2020, doi: 10.3390/su12166631.

[3]. R. A. Mendoza and A. J. Del Cid Z., “Improvement of control and monitoring of industrial processes using PLC by implementing asynchronous bidirectional communications and UMI,” 2019 IEEE 39th Cent. Am. Panama Conv. CONCAPAN 2019, vol. 2019- Novem, 2019, doi: 10.1109/CONCAPANXXXIX47272.2019.8977063

[4]. L. Chen, L. Liu, M. He, and D. Liu, “Gearbox fault diagnosis based on VMD and acoustic emission technology,” I2MTC 2019 - 2019 IEEE Int. Instrum. Meas. Technol. Conf. Proc., vol. 2019-May, pp. 1–6, 2019, doi: 10.1109/I2MTC.2019.8826954.

[5]. B. Ooi, Z. Kong, W. Lee, S. Liew, and S. Shirmohammadi, “A Collaborative IoT-Gateway Architecture for Reliable and Cost Effective Measurements,” IEEE Instrum. Meas. Mag., vol. 22, no. 6, pp. 11–17, 2019.

[6]. C.L. Cutland, E.M. Lackritz, T. Mallett-Moore, A. Bardají, R. Chandrasekaran, C. Lahariya, M.I. Nisar, M.D. Tapia, J. Pathirana, S. Kochhar et al., "Low birth weight: Case definition & guidelines for data 57 collection analysis and presentation of maternal immunization safety data", Vaccine, vol. 35, no. 48, pp. 6492, 2017.

[7]. P. Hidalgo-Lopezosa, A. Jimenez-Ruz, J. Carmona-Torres, M. HidalgoMaestre, M. Rodríguez-Borrego and P. Lopez-Soto, "Socio demographić factors associated with preterm birth and low birth weight: A crosssectional study", Women and Birth, 2019

[8]. Ismail, L., Materwala, H., Tayefi, M., Ngo, P. & Karduck, A. P. Type 2 diabetes with artificial intelligence machine learning: Methods and evaluation. Arch. Computat. Methods Eng https://doi.org/10.1007/S11831-021-09582-X (2021).

[9]. Al Haddad, A. et al. Mother and Child Health Study: protocol for a prospective cohort study investigating the maternal and early life determinants of infant, child, adolescent and maternal health in the United Arab Emirates. BMJ Open 9, e030937 (2019).

[10]. 10.MIAO FENG, LI WAN, ZHI LI, LINBO QING, AND XIAORONG QI Using Machine Learning to Estimate Fetal Weight via Ultrasound . Accepted on June 22, 2019, date of publication July 2, 2019, date of current version July 18, 2019.