International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
ISSN: 2457-0362
www.ijarst.in
IJARST

Predicting Low Birth Weight: A Machine Learning Approach

M. Anitha[1], K. Hareesh[2], G. Mounika[3]

#1 Assistant Professor & Head of MCA Department, SRK Institute of Technology, Vijayawada.
#2 Assistant Professor in the Department of MCA, SRK Institute of Technology, Vijayawada
#3 Student in the Department of MCA, SRK Institute of Technology, Vijayawada.

ABSTRACT Predicting baby birth weight is an important component of prenatal care since it allows for early intervention and personalised healthcare for pregnant mothers and their infants. This project introduces "Birth Weight Predictor," a user-friendly web application developed using Flask and powered by machine learning that estimates birth weight depending on maternal characteristics. The application makes use of a large dataset that includes maternal health variables such as age, weight, height, medical history, habits, and more. Machine learning methods are used to analyse these maternal characteristics and predict birth weight accurately. The following are some of the application's key features: Maternal Data Input: Users can enter maternal data such as age, weight, height, and other pertinent parameters to get a personalised birth weight prediction Sophisticated machine learning methods, such as voting classifier with Random forest, Boosting algorithm, and logistic regression, are used to process the maternal data, allowing the model to uncover relevant patterns and associations for accurate predictions. Flask Web Interface: The user-friendly Flask-based web interface makes birth weight projections accessible and instructive for both healthcare providers and pregnant parents. This software is a useful tool for healthcare providers, expectant parents, and researchers, as it provides early insights into prospective birth weight outcomes. It contributes to improving prenatal care, minimising problems, and maintaining the well-being of both moms and newborns by leveraging the power of machine learning and Flask.

1.INTRODUCTION

Birth weight affects its possibilities of endurance. Low birth weight (LBW) is turning out to be more an issue, especially in arising nations. A significant reason for neonatal passing is low birth weight, under 2500 g . Infants brought into the world at a low birth weight are multiple times bound to bite the dust than children brought into the world at a typical birth weight

It's likewise a decent mark of a kid's future unexpected issues. Low birth weight influences one out of each and every seven infants, representing around 14.6 percent of the children conceived around the world. Anticipating birth weight is a critical part of pre-birth care and has a few significant ramifications for both maternal and neonatal wellbeing. Proof shows that the worldwide commonness of LBW dropped by 1.2 percent every year somewhere in the range of 2000 and 2015, implying that progress is deficient to satisfy the World Wellbeing Gathering's low birth weight focus of 30% by 2025 [1]. LBW is as yet a serious general wellbeing worry across the world [1], putting children and infants at an expanded gamble of death and dismalness. Thus, one of the principal points of the 'A World Fit for Kids' drive is to diminish low birth weight as a critical commitment to the Thousand years Improvement Objective.

Birth weight (BW) assumes a significant part in the endurance and wellbeing of babies, and precise BW expectation will assist medical services specialists with pursuing ideal choices. Babies with a BW of ≤ 2500 g are considered as low BW (LBW) newborn children. Low BW in newborn children can happen as a result of different reasons like maternal eating regimen, close pregnancy stretches, contaminations, high equality, preterm conveyance, and financial elements. Contrasted and typical BW newborn children, LBW babies are at a higher gamble of perinatal passing at a proportion of 8:11. Besides, LBW newborn children have a more prominent possibility having serious improvement issues like low IQ (level of intelligence), mental hindrance, visual and hearing debilitation, neonatal hypothermia, neonatal hypoglycemia, long haul handicaps, and untimely death2,3. Recognizing LBW newborn children before birth may considerably diminish such dangers contrasted and distinguishing such babies after birth. Consequently, exact and convenient conclusion of LBW newborn children is fundamental for clinical experts to diminish the gamble factors for moms and babies by giving suitable mediations and working on the general guess.

As of late, to give clinical experts better guess and conclusion support, AI (ML) calculations have turned into a standard decision for proficient clinical applications like BW assessment and classification1,3. Nonetheless, there are a few difficulties related with making such ML-based frameworks. ML-based frameworks require quality data4 for preparing and assessment; in any case, making such a great dataset is troublesome on the grounds that most clinical information are not freely accessible inferable from copyright and protection regulations. Moreover, a few records in these datasets contain missing records, which is very normal in clinical related data5,6 and influences the general presentation of a ML-based framework.

Datasets with high aspects present one more test for information mining and order assignments. Regularly, high-layered datasets incorporate an enormous number of insufficient or pointless factors that can adversely influence the ML model's presentation. To resolve this issue and work on the general execution, highlight determination calculations are utilized to choose applicable and significant elements from the dataset7. A few procedures are accounted for in literature7,8 that select an ideal list of capabilities for enough addressing the dataset to work on in general execution. The datasets utilized in current LBW grouping review are exceptionally class imbalanced, i.e., the quantity of information focuses accessible for various classes contrasts. Class unevenness extensively corrupts the productivity of a characterization framework. Generally, to resolve this issue, the minority class is oversampled by copying the haphazardly chosen tests and the larger part class is undersampled. The engineered minority oversampling procedure (SMOTE)9 is a notable information adjusting strategy, which oversamples the minority class by making orchestrated examples in view of the similitudes between sets of the current minority instances4,9. The Destroyed is a straightforward yet effective calculation that outflanks cutting edge generative ill-disposed networks (GANs)10. Along these lines, in this review, Destroyed is embraced for information adjusting. LBW and typical birthweight (NBW) can be ordered in view of the highlights gave to different classifiers, for example, support vector machines (SVM), calculated relapse (LR), gullible Bayes (NB), and irregular woodland (RF). Past investigations have assessed the presentation of numerous ML models utilizing heterogeneous datasets and different execution measurements. In any case, supposedly, no review has given a definite assessment of different ML models utilizing various execution measurements on a few subsets of highlights.

The new Habitats for Infectious prevention and Avoidance (CDC) yearly report shows that children with low birthweight (LBW), characterized as a baby weight of <2500 g by the World Wellbeing Association [1], represented 8.24% of all births in the US in 2020 [2]. Alarmingly, LBW was the subsequent driving reason for neonatal demise in the US after intrinsic deformities in 2019 and 2020 [3]. Besides, children with LBW have a higher gamble of short-and long haul unfriendly wellbeing impacts than those with a typical birthweight, for example, heart and lung entanglements and related ongoing infections [4]. Enormous monetary weights have additionally been forced on the groups of the infants with LBW and medical services payers [5]. Past examinations have found that maternal socioeconomics [6], previous medical issue [7], social determinants [8], and pre-birth care level [9] are related with LBW. Hence, exactly distinguishing which pregnant patients might be at the most serious gamble of having a child with LBW in the predisposition or early pregnancy stages is basic to save neonatal lives and lessen possibly avoidable clinical uses through direct clinical and wellbeing strategy mediations. Our benchmarking models and component significance examination results can possibly support this underlying clinical screening and ID of high-risk birthing individuals and the advancement of arrangements that further develop medical services quality and put resources into networks of most an open door. As of late, with the dramatic development in the amount and aspect of medical care information, AI (ML) techniques have been acquainted with handle perplexing and high-layered information [10,11]. Various examinations have shown that ML calculations accomplished great execution on track expectation [12,13], result assessment [14], and risk factor investigation [15,16] in the field of accuracy medication. In spite of the fact that ML is utilized as another computational strategy to investigate different medical conditions, the primary test of ML applications in the wellbeing area is the normal issue of imbalanced information that excessively center around the soundness of minority bunches in manners that might produce antagonistic occasions in the entire populace [17]. An imbalanced class circulation represents a test to the presentation of ML models. These models prepared on imbalanced informational collections can create misdirecting results for the planned expectation errands. These models frequently yield sub-par order results and may wrongly regard interesting minority models as commotion [18]. Furthermore, the utilization of worldwide execution measurements, for example, expectation exactness to direct the educational experience can make a predisposition toward the greater part class, prompting an absence of familiarity with interesting occasions regardless of whether the expectation model accomplishes high generally speaking precision [18,19]. In the pre-birth medical services field, ML models face similar difficulties with imbalanced informational collections for right expectations due to the lower recurrence of unfortunate results contrasted and ordinary results. Thus, distinguishing key benchmarks is significant to directing legitimate ML use in perinatal consideration, maternal wellbeing, and other wellbeing areas. This is an extensive information hole that we meant to fill in this review.

2.LITERATURE SURVEY

2.1 S. A, A. A, M. Koppala and R. Kaladevi, "Infant Birth Weight Estimation Using Machine Learning," 2023 7th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2023, pp. 1019-1023, doi: 10.1109/ICOEI56765.2023.10125609.

The burgeoning concern over low birth weight necessitates innovative approaches, and the integration of machine learning (ML) algorithms has emerged as a promising avenue. This literature review explores the landscape of ML applications in infant birth weight estimation, emphasizing key contributing factors and comparing various algorithms. Notably, the study by S. A, A. A, M. Koppala, and R. Kaladevi introduces a model showcased in the 7th International Conference on Trends in Electronics and Informatics (ICOEI), aiming not only for accurate predictions but also to recommend tailored meal programs for mothers based on algorithmic estimations. The envisioned impact extends beyond precision, addressing broader maternal and infant health concerns. Despite the progress, challenges in data quality, interpretability, and ethical considerations underscore the need for ongoing research to refine these models for practical implementation in healthcare settings.

2.2 M. W. L. Moreira, J. J. P. C. Rodrigues, V. Furtado, C. X. Mavromoustakis, N. Kumar and I. Woungang, "Fetal Birth Weight Estimation in High-Risk Pregnancies Through Machine Learning Techniques," ICC 2019 - 2019 IEEE International Conference on Communications (ICC), Shanghai, China, 2019, pp. 1-6, doi: 10.1109/ICC.2019.8761985.

The study by M. W. L. Moreira et al., presented at the 2019 IEEE International Conference on Communications (ICC), delves into the critical issue of low fetal birth weight in high-risk pregnancies, a condition with significant implications for newborn health and infant mortality rates. Recognizing the potential of artificial intelligence, particularly machine learning (ML), to anticipate fetal health issues, the paper investigates multiple ML techniques for predicting whether a fetus will be born small for its gestational age. Notably, the proposed hybrid model, labeled the bagged tree, demonstrates commendable accuracy and area under the receiver operating characteristic curve, with values of 0.849 and 0.636, respectively. The significance of early problem detection in fetal development lies in the prospect of extending gestation days through timely interventions, potentially improving birth weight and subsequently reducing neonatal morbidity and mortality. This research contributes to the growing body of literature leveraging ML for proactive and effective strategies in high-risk pregnancies, emphasizing the potential impact on maternal and infant health outcomes.

2.3 F. L. De Morais et al., "Predicting Low Birth Weight Using Machine Learning Models," 2023 18th Iberian Conference on Information Systems and Technologies (CISTI), Aveiro, Portugal, 2023, pp. 1-7, doi: 10.23919/CISTI58278.2023.10211576.

In the pursuit of enhancing prenatal care and mitigating the risks of low birth weight and associated adverse pregnancy outcomes, F. L. De Morais et al. contribute valuable insights through their study presented at the 18th Iberian Conference on Information Systems and Technologies (CISTI) in 2023. Emphasizing the pivotal role of prenatal care in reducing mortality and morbidity risks, the researchers aim to evaluate the efficacy of machine learning models in predicting the likelihood of low birth weight outcomes in pregnancies. Leveraging a dataset from the Brazilian Live Births Information System (SINASC), the study focuses on pregnant women, prenatal care, and newborns. Three tree-based machine learning models, with a particular emphasis on attributes highlighted in existing literature, were employed. Notably, the Adaboost model emerged as the most promising, achieving the highest metrics in the test dataset with an f1-score of 60.65% and a sensitivity of 51.34%. The study underscores the significance of attributes such as age, education, maternal occupation, and multiple gestations in influencing the prediction process. This research augments the growing body of literature on the application of machine learning in prenatal care, offering a potential avenue for early identification of at-risk pregnancies and targeted interventions to improve maternal and infant health outcomes.

4.PROPOSED WORK

1. Data Cleaning:

Missing Values:

- Identify and handle missing values in the dataset. Common methods include imputing missing values with mean, median, or mode for numerical variables and using the most frequent category for categorical variables.

Outliers:

- Detect and handle outliers in the dataset. You can use statistical methods like Z-score or IQR to identify outliers and either remove them or transform them.

Inconsistencies:

- Check for inconsistencies in the data, such as typos or errors. Correct or remove inconsistent data points.

2. Exploratory Data Analysis (EDA):

- Explore the distribution of variables using histograms, box plots, and other visualization techniques.
- Analyze correlations between variables, especially predictors and the target variable (childwt).
- Look for potential relationships and patterns in the data.

3. Data Splitting:

- Split the dataset into a training set and a testing set. Common splits are 80-20 or 70-30 for training and testing, respectively.

4. Model Building:

- Choose a machine learning algorithm suitable for predicting birth weight based on maternal factors. In this case, you've selected a voting classifier with Random Forest, Boosting algorithm, and Logistic Regression.

5. Model Evaluation:

- Evaluate the model's performance using metrics such as F1 score, precision, and recall. These metrics are particularly important for a binary classification problem like predicting low birth weight cases.

6. Model Optimization:

- Fine-tune the model and hyperparameters to improve prediction accuracy. You can use techniques like grid search or randomized search for hyperparameter tuning.

7. Prediction and Interpretation:

- Make predictions on unseen data (testing set) and interpret the results. Understand the impact of maternal factors on birth weight based on the model's predictions.

8. Deployment:

- Implement the model in a Flask framework for deployment. This involves creating an API that takes input data, performs predictions, and returns results.

3.1 IMPLEMENTATION

3.1.1 DATA COLLECTION:

- These features seem to cover a wide range of factors that can influence birth weight and fetal health.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
IJARST
ISSN: 2457-0362
www.ijarst.in

- **Maternal Characteristics:**
- age: Age of the mother.
- height: Height of the mother.
- weight: Weight of the mother.
- parity: Parity refers to the number of previous live births.
- **Pregnancy Characteristics:**
- gestation: Duration of pregnancy.
- smoke: Whether the mother smokes during pregnancy.
- **Fetal Monitoring:**
- accelerations: Fetal heart rate accelerations.
- fetal_movement: Fetal movement.
- uterine_contractions: Uterine contractions.
- light_decelerations: Mild decreases in fetal heart rate.
- severe_decelerations: Severe decreases in fetal heart rate.
- prolongued_decelerations: Prolonged decreases in fetal heart rate.
- **Cardiotocogram (CTG) Features:**
- abnormal_short_term_variability: Short-term variability in fetal heart rate.
- mean_value_of_short_term_variability: Mean value of short-term variability.
- percentage_of_time_with_abnormal_long_term_variability: Percentage of time with abnormal long-term variability.
- mean_value_of_long_term_variability: Mean value of long-term variability.
- **Histogram Features:**
- histogram_width: Width of the fetal heart rate histogram.
- histogram_min: Minimum value in the fetal heart rate histogram.
- histogram_max: Maximum value in the fetal heart rate histogram.
- histogram_number_of_peaks: Number of peaks in the fetal heart rate histogram.
- histogram_number_of_zeroes: Number of zeroes in the fetal heart rate histogram.
- histogram_mode: Mode of the fetal heart rate histogram.
- histogram_mean: Mean of the fetal heart rate histogram.
- histogram_median: Median of the fetal heart rate histogram.
- histogram_variance: Variance of the fetal heart rate histogram.
- histogram_tendency: Tendency of the fetal heart rate histogram.

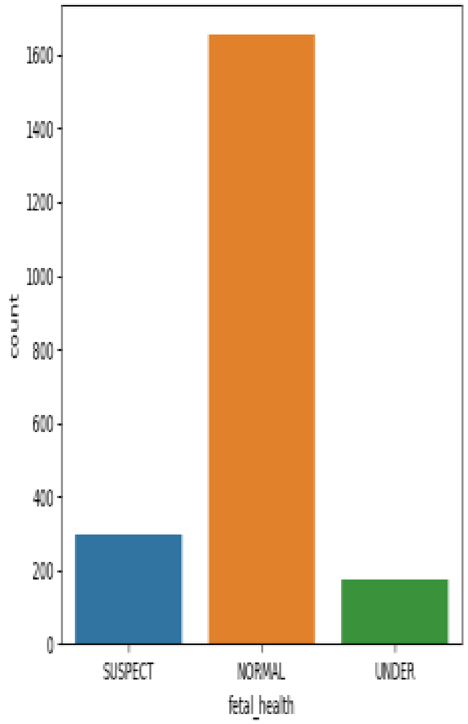
3.1.2 Data Preprocessing:

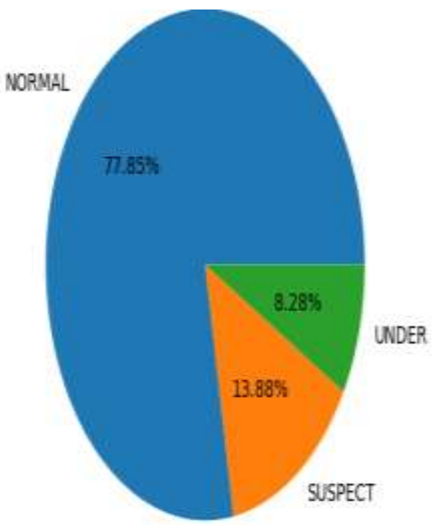
- **Data cleaning** is a critically important step in any machine learning project.
- In this module data cleaning is done to prepare the data for analysis by removing or modifying the data that may be incorrect, incomplete, duplicated or improperly formatted.
- In tabular data, there are many different statistical analysis and data visualization techniques you can use to explore your data in order to identify data cleaning operations you may want to perform
- Check for any missing values in the dataset.
- Visualize the class distribution using countplot and pie chart.
- Apply the preprocessing function to obtain training and testing sets (X_train, X_test, y_train, y_test).


"birth_weight","gestation","parity","age","height","weight","smoke"
120,284,0,27,62,100,0
113,282,0,33,64,135,0
128,279,0,28,64,115,1


baseline value,accelerations,fetal_movement,uterine_contractions,light_decelerations,severe_decelerations,prolonged

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
IJARST
ISSN: 2457-0362
www.ijarst.in





Correlation Analysis:

• A heatmap of the correlation matrix is generated using sns.heatmap(df.corr(), annot=True) to visualize the relationships between different features.

• Standard scaling of the features using StandardScaler.

3.1.3 MODEL TRAINING

- A training model is a dataset that is used to train an ML algorithm. It consists of the sample output data and the corresponding sets of input data that have an influence on the output.
- The training model is used to run the input data through the algorithm to correlate the processed output against the sample output. The result from this correlation is used to modify the model.
- This iterative process is called “model fitting”. The accuracy of the training dataset or the validation dataset is critical for the precision of the model.
- Model training in machine language is the process of feeding an ML algorithm with data to help identify and learn good values for all attributes involved.
- There are several types of machine learning models, of which the most common ones are supervised and unsupervised learning.
- In this module we use supervised classification algorithms to train the model on the cleaned dataset after dimensionality reduction.

Logistic Regression: Logistic regression is a linear classification algorithm that models the probability of a sample belonging to a particular class. The models m2, m3, and m4 are logistic regression models with different settings, such as default parameters, balanced class weights, and One-Versus-One (OVO) strategy.

Random Forest Classifier : Random forests are ensemble learning methods that construct a multitude of decision trees at training time and output the class that is the mode of the classes (classification) of the individual trees. The model m8 uses specific hyperparameters like n_estimators, class_weight, criterion, and min_samples_leaf.

AdaBoost Classifier AdaBoost is an ensemble learning method that combines multiple weak classifiers to create a strong classifier. In this code, an AdaBoost classifier is used with a Decision Tree base estimator and specific hyperparameters like n_estimators and learning_rate.

3.1.4 PREDICTION

• A Flask web application is created to provide a user interface for making predictions using the trained models.

• Flask routes are defined for handling different functionalities, including loading the models and making predictions.

• HTML templates are used for creating user interfaces to take input features and display prediction results.

4.RESULTS AND DISCUSSION

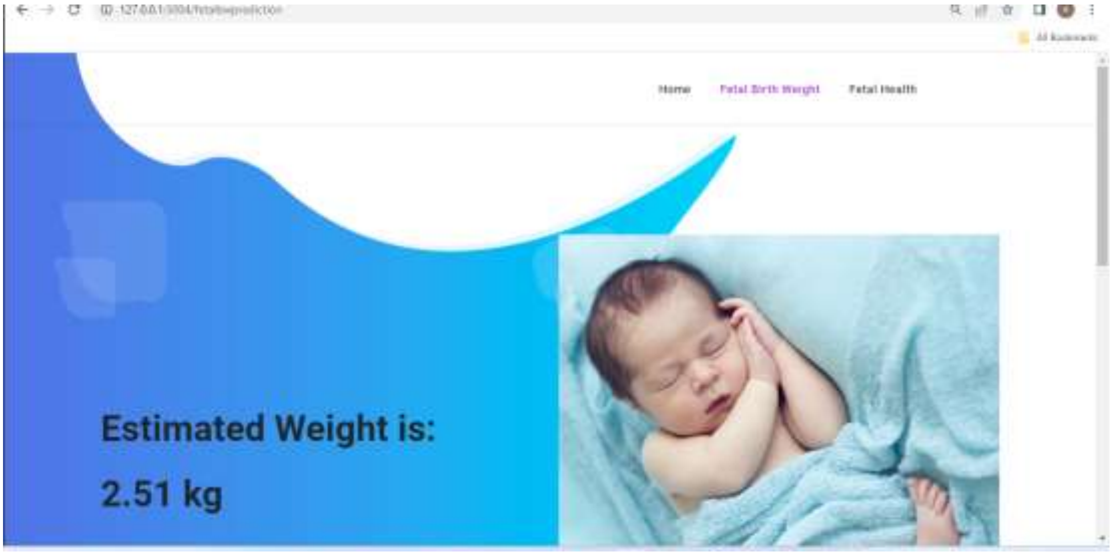


Fig 1:Based on input parameters we got results

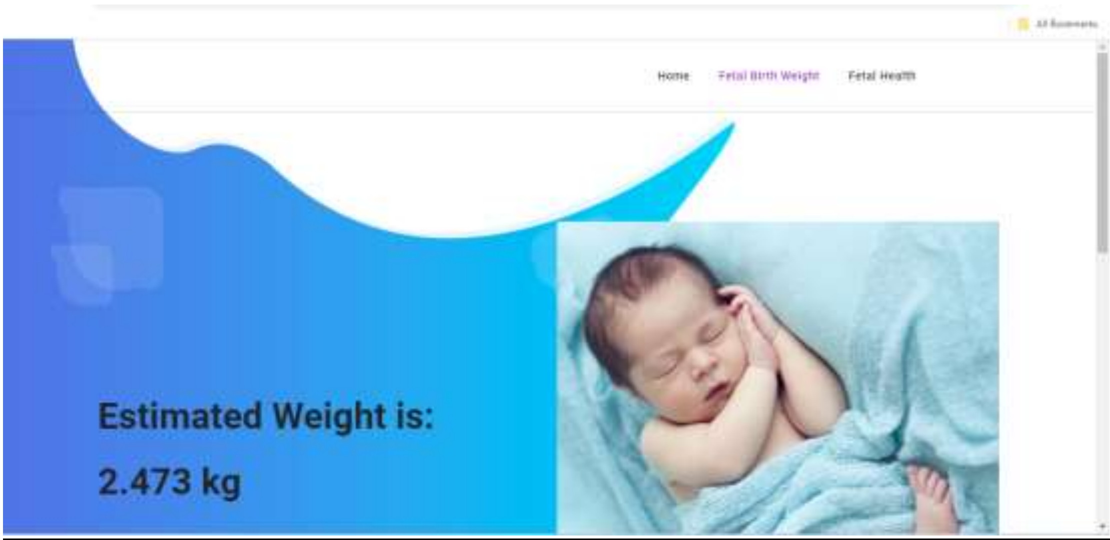


Fig 2:Based on input parameters we got results

5.CONCLUSION:

The "Birth Weight Predictor" online application represents a substantial leap in prenatal care by integrating machine learning capabilities and the user-friendly Flask framework. The application delivers a personalised and accurate estimation of newborn birth weight by integrating a complete dataset comprising numerous mother health variables such as age, weight, height, medical history, and habits. The use of complex machine learning techniques, such as the voting classifier with Random Forest, the Boosting algorithm, and logistic regression, improves the model's predictive skills. This guarantees that important patterns and linkages in maternal data are detected, resulting in accurate birth weight forecasts. The Flask-based online interface improves accessibility and simplicity for healthcare providers, pregnant parents, and researchers alike. Because of the simplicity with which maternal data may be entered via the intuitive interface, birth weight projections are quickly available, enabling early insights into probable outcomes and facilitating informed decision-making.Overall, the "Birth Weight Predictor" software is a useful tool in the field of prenatal care. Its role to improving early intervention, personalised treatment, and minimising problems illustrates the promise of combining technology and healthcare for the benefit of both mothers and newborns. As healthcare technology advances, this research offers as a shining example of how machine learning applications can improve the quality of care delivered to expecting mothers and their infants.

REFERENCES:

1. Yu Z, Han S, Zhu J, Sun X, Ji C, Guo C. Pre-pregnancy body mass index in relation to infant birth weight and offspring overweight/obesity: a systematic review and meta-analysis. PLOS ONE. 2013;8(4):61627.

2. Shepard M, Richards V, Berkowitz R, Warsof S, Hobbins J. An evaluation of two equations for predicting birthweight by ultrasound. Am J Obstetr Gyneco. 1982;142(1):47–544.

3. Hadlock F, Harrist R, Carpenter R, Deter R, Park S. Sonographic estimation of birthweight. The value of femur length in addition to head and abdomen measurements. Radiology. 1984;150(2):535–40.

4. Zhu T, Zhao X, Ai M, Ma R, Lei J, Liu J. Comparison of accuracy of six calculations for predicting fetal body weight. Chin Mater Child Health Care. 2016;(20):4179–81.

5. Möst L, Schmid M, Faschingbauer F, Torsten H. Predicting birth weight with conditionally linear transformation models. Chin Mater Child Health Care. 2016;25(6):2781–810.

6. Hong C, Ji Y. Comparison of common methods for birthweight prediction and clinical value. Chin Women Children Health Res. 2017;5:2781–810.

7. Kuhle S, Maguire B, Zhang H, Hamilton D, Allen A, Joseph K, Victoria M. Comparison of logistic regression with machine learning methods for the prediction of fetal growth abnormalities: a retrospective cohort study. BMC Pregn Childbirth. 2018;18(1):333.

8. Farmer R, Medearis A, Hirata G, Platt L. The use of a neural network for the ultrasonographic estimation of birthweight in the macrosomic fetus. Am J Obstetr Gynecol. 1992;166(5):1467–72.

9. Cheng Y, Hsia C, Chang F, Hou C, ChiU Y, Chung K. Zeolites and synthetic mechanisms. In: 6th World Congress of Biomechanics (WCB 2010), Singapore 2010;1514–1517

10. Mohammadi H, Nemati M, Allahmoradi Z, Forghani H, Sheikhani A. Ultrasound estimation of birthweight in twins by artificial neural network. J Biomed Sci Eng. 2011;4(1):46–50.

11. Feng M, Wan L, Li Z, Qing L, Qi X. Fetal weight estimation via ultrasound using machine learning. IEEE Access. 2019;7:87783–91.

AUTHOR'S PROFILE

Ms. M. Anitha Working as Assistant & Head of Department of MCA, in SRK Institute of technology in Vijayawada. She done with B. Tech, MCA, M. Tech in Computer Science. She has 14 years of Teaching experience in SRK Institute of Technology, Enikepadu, Vijayawada, NTR District. Her area of interest includes Machine Learning with Python and DBMS.

Mr. K. Hareesh completed his Master of Computer Applications. Currently working as an Assistant Professor in the department of MCA at SRK Institute of Technology,

Enikepadu, Vijayawada, NTR District. His area of interest includes Networks, Machine Learning.

Ms. G. Mounika is an MCA Student in the Department of Computer Application at SRK Institute of Technology, Enikepadu, Vijayawada, NTR District. She has Completed Degree in B.Sc. (MPCS) from Vidya Vikas College, Akividu, West Godavari District. Her area of interest are DBMS and Machine Learning with Python.