INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact Factor: 6.078
(Volume 7, Issue 4 - V7I4-1342)
Available online at: https://www.ijariit.com

Fetal birth weight estimation in high-risk pregnancies through Machine Learning Techniques

Meghana S.
meghanas.17cs@saividya.ac.in
Sai Vidya Institute of Technology, Bangalore, Karnataka

Y. Sree Rushitha
sreerushithay.15cs@saividya.ac.in
Sai Vidya Institute of Technology, Bangalore, Karnataka

Susan Syeda
susansyeda.17cs@saividya.ac.in
Sai Vidya Institute of Technology, Bangalore, Karnataka

Priyanka M.
priyankam.17cs@saividya.ac.in
Sai Vidya Institute of Technology, Bangalore, Karnataka

Dr. Shantakumar B. Patil
shantakumar.bpatil@saividya.ac.in
Sai Vidya Institute of Technology, Bangalore, Karnataka

ABSTRACT

Low birth weight of the fetus is considered one of the most critical problems in pregnancy care, which will affect the health of the newborn and in more severe cases will lead to its death. This situation is the reasonfor the high infant mortality rate throughout the world. In terms of health, artificial intelligence technologies, especially those based on machine learning (ML), canearly predict problems related to the health of the fetusthroughout pregnancy (even at birth). Therefore, the project proposes an analysis of several ML techniques that can predict whether the fetus will lose weight at birth in its gestational age. The importance of early diagnosis of problems related to fetal development depends on the possibility of increasing the number ofdays of pregnancy through timely intervention. This intervention will help to improve the weight of the fetusat birth, thus reducing neonatal morbidity and mortality.

Keywords: *Fatal Birth Weight Estimation,Machine Learning, Linear Regression, Random Forest Regressor, XGBRegression*

1. INTRODUCTION

Low fetal birth weight is considered to be one of the most critical issues in pregnancy care. This situation is the reason for the high infant mortality rate around the world. ML technology can predict whether a fetus will beyounger than its gestational age at birth.

The importance of early diagnosis of problems related to fetal development lies in the possibility of increasing the number of days of pregnancy through timely intervention.

Fetal weight is a key factor in predicting short- and long-term health consequences. According to the birth weight (BW), the World Health Organization (WHO) divides newborns into three groups, namely low birthweight (LBW, BW<2500g), normal birth weight (NBW, 2500g≤BW<4000g) and high birth weight. (HBW, BW) ≥ 4000 g) which isalso called macrosomia.

Low birth weight is related to fetal and newborn mortality and developmental delay, and can also lead to long-term childhood diseases such as mental retardation and learning disabilities. Therefore, it is very important to accurately estimate the weight ofthe fetus during pregnancy and correctly identify the low birth weight fetus or the giantbaby. Once the risk is determined, appropriate clinical decisions and preventive measures can be taken to reduce the morbidityand mortality of pregnant women or newborns.

Therefore, continuous monitoring of fetal health through information and communication technology can reduce mortality and morbidity and help improve thequality of life of mothers and babies. Fetal weight is a key factor in predicting short- andlong-term health consequences.

We proposed a system to predict fetal birthweight using machine learning techniques and algorithms like linear regression, random forestregressor and XGB Regressor based on different parameters entered by the user in the front end and classify them based predicted birth weight as low, normal and abnormal birth weight.

2. LITERATURE SURVEY

M. Feng, L. Wan, Z. Li, L.Qing and X. Qi: In this study, they proposed a novel fetal weight estimation model which combined SVM based classification with DBN toimprove the performance of EFW in allfetal weight ranges, they also solved theimbalanced learning problem by utilizing SMOTE based data augmentation. It was demonstrated from the result that the proposedmodel outperformed the regression formulas.Their study revealed that DBN is a promising approach for fetal weight estimation, it also provedthat classify fetuses into different groups and predictbirth

weight using different significant parameters are effective.

M. W. L. Moreira, J. J. P. C.Rodrigues, V. Furtado, C. X. Mavromoustakis, N. Kumar and I. Woungang: During this paper the utilization of AI techniques joined with novel technologies will scale back the highmorbidity and mortality rates worldwide, particularly in developing countries. Hence, this paper compared many cc techniques employing a real info of pregnant ladies United Nations agency suffered some hypertensive disorder throughout physiological state. The results shown that hybrid strategies supported ensemble learning area unit capable of withefficiency predicting the expected weight of the fetus at birth.

J. Yu, Y. Wang and P. Chen:during this paper, the EFSVR is projected to estimate craniate weight estimation for LBW infants. correct weight estimation for LBW fetuses ought to alter many issues. thanks to the poor quality of ultrasound pictures and inter- or intra observer variances, ultrasound activity knowledge for the model construction area unit sometimes inaccurate. to reinforce the strength of the burden estimation rule, the FSVR is projected to alleviate the impact ofinaccurate knowledge to the model coaching. that the generalization performance of the FSVR model seriouslydepends on the best setting of many empirical parameters. They gift a way supported the NSGA-II to pick out the bestFSVR parameter settings.

F. Sereno, P. Marques de Sa, A. Matos and J. Bernardes: The SV approach here delineate achieves calculable craniate weight errors below those obtained by victimization twenty six regression equations .In addition calculablecraniate weight to be utilized in clinical management wants adaptation to native conditions of measurement the crucial biometric parameters. The generalization and combination of ensembles of neural nets so as to make sure that knowledge variability inherent to the dynamics of the growing foetus phenomena.

R. Czabanski, M. Jezewski, J. Wrobel, J. Jezewski and K. Horoba: during this paper, we have a tendency to investigated the flexibility of application ofthe ANBLIR to predict the chance of low- fetal birth weight basing on the classification of quantitative description of craniate cardiotocogrphic signals. along side analysis of various learning strategies of the ANBLIR, we have a tendency to examined the influence of the patients knowledge distribution and conjointly the role of the underrepresented abnormal category on the prediction quality. The experiments show that it's higher to use thetheme one single CTG trace assigned to at least one patient. The obtained results indicate conjointly a rise of the chance of low-fetal birth weight prediction with adecrease of the craniate age.

M. Abdollahianand N.Gunaratne: during this work they projected an easy and economical mathematical model supported real knowledge collected over a pair of years to estimate the delivery weight for low birth weight babies. solely real recorded knowledge were analyzed victimization multi statistical regression model to assess the impact of many predictors. The p-valuereminiscent of individual characteristics area unit accustomed establish the foremost important reduced model for the prediction.The analysis show that the regression supported solely gestation age, baby's height and head circumference will make acase for hour of the variation within the newborn weight for LBW babies.

3. SYSTEM DESIGN

- Fetal weight dataset is taken andloaded.
- The dataset divided as train data andtest data.
- The data is preprocessed in orderto fill the missingvalues in the dataset, standardization and labelencoding etc to increase the accuracy.
- The features which are directlyaffecting the final results were extracted.
- The model is built using machine learning algorithms like logistic regression, support vector machine, decision tree, knn, naïvebayes.
- The model is trained with thepreprocessed data.
- The model is tested and accuracyis calculated for different ML algorithms.

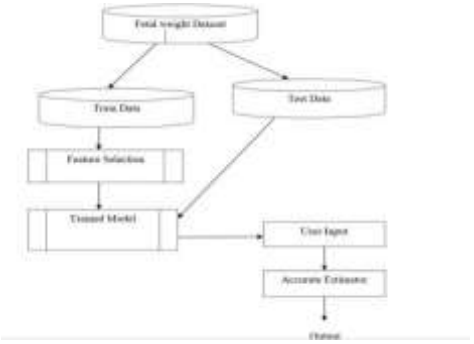

Fig. 1: System Architecture

High leveldesign

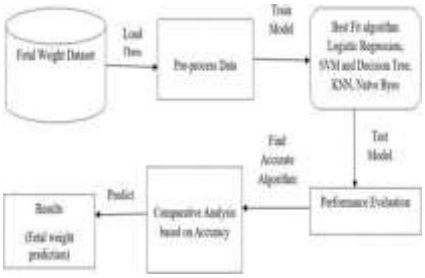

Fig. 2: Flow Diagram

- Fetal weight dataset is taken andloaded.
- The data is preprocessed in order to fill themissing values in the dataset standardization and label encoding etc, toincrease the accuracy.
- The model is built using machine learning algorithms like logistic regression, supportvector machine,decision tree .
- The model is trained with thepreprocessed data.
- The model is tested and accuracy iscalculated for different ML algorithms.
- The algorithm with best accuracy isfinalized and that model will predictthe fetal weight.

Activity Diagram

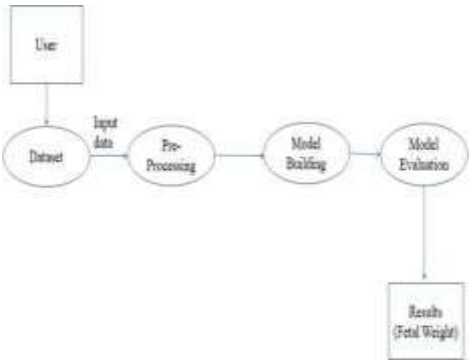

Fig. 3: Activity Diagram

- The user will provide the datasetto the system.
- The dataset is preprocessed in order to increase the accuracy of themodel.
- The model is built using different algorithms.
- The model is evaluated and model withbest accuracy is finalized.
- The finalized model will predictthe results.

Low Level Design

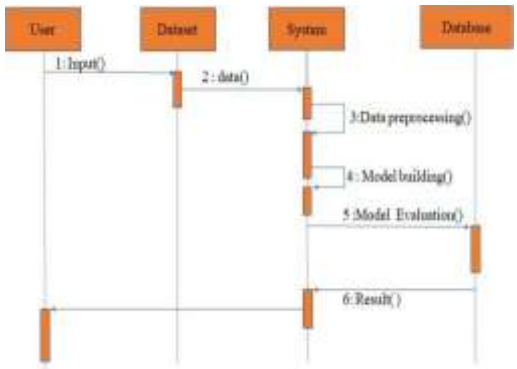

Fig. 4: Sequence Diagram

- The user will give dataset as input to the system.
- The system will store the dataset given by the user in its database.
- The system will do pre processing of the data stored.
- The model is built using various ML algorithms and trained using pre processeddata.
- The model is evaluated and the algorithm with best accuracy is finalized.
- The finalized model will predictthe results.

4. IMPLEMENTATION

Python is a high-level, object- oriented, interpreted programming language with dynamic semantics. Its advanced built-in data structure, combined with dynamic writing and dynamic linking, make it highlyattractive for rapid application development and as a scripting language. Python's simple and easy-to- learn syntax emphasizes readability, thus reducing the cost of maintaining the program. Python supports modules and packages, which encourage program modularity and code reuse. The Python interpreter and the extensive standard library are provided as open source. Python is widely regarded as the preferred language for ML (machine learning) teaching and learning.

Python IDE:

- IDE (or integrated development environment) is a program dedicated to software development.
- IDE integrates a series of tools specially designed for software development. These tools generally include:
- Editors designed to process code (eg with syntax highlighting and autocompletion)
- Build, run, and debug tools
- Some form of source code control
- When you install Python, IDLE is also installed default.
- Its main functions include Python shell window (interactive interpreter), autocompletion, syntax highlighting, smart indentation, and a basic built-in debugger.
- We use Python version 3.6.8.

HTML

HTML (Hypertext Mark up Language) is thecode that is used to structure a web page andits content. For example, content could be structured within a set of paragraphs, a list ofbulleted points, or using images and data tables.

HTML is a markup language that defines the structure of your content. HTML consists of aseries of elements, which you use to enclose, or wrap, different parts of the content to makeit appear a certain way, or act a certain way. The enclosing tags can make a word or image hyperlink to somewhere else, can italicize words, can make the font bigger or smaller, and so on.

Xampp Control Panel

XAMPP is the title used to compile free software. The name is an acronym, and each letter represents one of five key components. software package contains Apache Web server, MySQL (or MariaDB) relational database management system and Perl and PHP programming languages. The original X represents the operating system it uses: Linux, Windows, and Mac OS X.

Apache

open source web server is the mostwidely used web content server in the world. The server application is available as free software from the Apache Software Foundation.

MySQL/MariaDB

In MySQL, XAMPP contains one of the most popular relational database management systems in the world. Combining the Apache Web server and the PHP programming language, MySQL provides data storage for Web services. The current version of XAMPP has replaced MySQL with MariaDB (a community-developed MySQL project branch made by the original developer).

PHP

The PHP server-side programming language allows users to createdynamic websites or applications.PHP can be installed on all platforms and is compatible with many different database systems.

Perl

The Perl scripting language is used for systems management, webdevelopment, and network programming. Like PHP, it also allows users to write dynamic web applications. In addition to these core components, this free-to-use Apache distribution also contains some other useful tools that vary depending on your operating system.

These tools include Mercury mail server , phpMyAdmin database management tool,Webalizer, OpenSSL and Apache Tomcat web analysis software solutions, and FileZilla or ProFTPd FTP server.

Application areas

An XAMPP server can be installed and used with a single executable file quickly and easily, functioning as a local test system for Linux, Windows, and Mac OS X. The software packet contains the same components that are found on common web servers. Developers have the chance to test out their projects locally and to transfer them easily to productive systems. But XAMPP isn't suitable to use as a public server, because **many safety features have been deliberately left out** to simplify and speed up the system for testing.

Flask

Flask is a Python web framework, which is built with a small

kernel and easy-to-use extension concepts.

Flask is considered more than the Django pythonic web framework, becausein common cases, the equivalent Flaskweb application is more explicit. It is alsoeasy for beginners to start using Flask, because there is almost no boilerplate code to start and run a simple application. Flask is a popular Python web framework,which means it is a third-party Python library for developing web applications.

Linear regression

Linear regression is probably one of the most important and widely used regression techniques. It is one of the simplest regression methods. One of its main advantages is that theresults are easy to interpret.

When implementing a linear regression of a dependent variable y on the set of independent variables $\mathbf{x} = (x_1, \ldots, x_r)$, where r is the number of predictors, it assumes that there is a linear relationship between y and $\mathbf{x}$: y = □□

□□□□Ꝑ… + $\beta_r x_r + \varepsilon$. This equation is a regression equation. β_0, β_1, ..., β_r is the regression coefficient and ε is the random error. Linear regression calculates the estimator of the regression coefficient or simply predicts theweight, expressed as b_0, b_1, ...,. They defined the estimated regression function $(\mathbf{x}) = b_0 + b_1 x_1$

+… + $b_r x_r$. This function should capture the dependencies between input and output well enough. For each observation i = 1, …, the estimated or predicted response ($\mathbf{x}_i$) of n shouldbe as close as possible to the corresponding real response y_i. The difference between all the observations $i = 1,\ldots, n$ y_i ($\mathbf{x}_i$) is called a residual. Regression consists of determining thebest predictive weight, that is, the weight corresponding to the smallest residual. To get the best weight, you usually minimize the residual sum of squares (SSR) of allobservations i = 1, …, n: SSR = Σ_i (y_i f ($\mathbf{x}_i$)) ². This method is called ordinary least squares

The actual response changes y_i, i = 1, ..., n, in part due to the dependence of the predictorvariable $\mathbf{x}_i$. However, there is an additional inherent difference in production. The coefficient of determination, expressed as R^2, tells you what changes in y can be explained byusing a specific regression model that depends on $\mathbf{x}$. A larger R^2 indicates a better fit, which means that the model can better explain changesin production under different inputs. The value R^2 = 1 corresponds to SSR = 0, which is a perfect fit, because the predicted and actualresponse values exactly match each other.

Random Forest is a supervised learning algorithm. It can be used for classification and regression. It is also the most flexible and easy-to-use algorithm. The forest is made up of trees.It is said that the more trees, the stronger the forest. The random forest creates decision trees on randomly selected data samples, obtains predictions from each tree, and selects the best solution by voting. It also provides a good feature importance index.

Technically speaking, it is an integrated method(based on the divide and conquer method) of decision trees generated on randomly divided data sets. This collection of decision tree classifiers is also called a forest. A single decision tree is generated using attribute selection indicators, such as information gain, gain ratio, and Gini index for each attribute. Each tree depends on an independent random sample. These predictions are then averaged to produce a single result. Averaging makes the random forest better than a single decision tree, improving its accuracy and reducing overfitting. The random forest regressor prediction is the average of the predictions produced by the trees in the forest.

XGB Regressor

XG Boost is a powerful machine learning algorithm especially where speed and accuracyare concerned. XG Boost (extreme GradientBoosting) is an advanced implementation of gradient boosting algorithm. Extreme GradientBoosting (XG Boost) is an opensource library that provides an efficient and effective implementation of the gradient boosting algorithm. Regression predictive modellingproblems involve predicting a numerical value such as a dollar amount or a height. XG BOOSTcan be used directly for regression predictive modelling.

5. RESULTS

This model has been trained using multiple algorithms. The algorithm with highest accuracy is selected as finalized algorithm.

SL. No	Models	Accuracy
1	Linear Regression	36%
2	Random ForestRegressor	38%
3	XGB Regressor	42%

6. CONCLUSION

Current standards for ultrasound assessment of fatal growth may cause up to 15% of foetuses tobe incorrectly classified as SGA. Growthrestriction is a sign of serious health problems, usually because the foetus does not get enoughnutrition or oxygen in the womb. The understanding of many aspects of fatal growth and the physiology of their limiting pathways is still weak. As a clinical recommendation, several ultrasound technology models have been developed. However, a precise disease method diagnosis has not yet been found. As we all know, the more growth restriction begins, the greater its severity. ML technology is an important tool to help experts identify suchchanges early. We have proposed a system that uses machine learning techniques and algorithms to accurately predict fatal birth weight.

Therefore, our project predicts that an earlyfoetus with a birth weight of less than 2.5 kg isalso a low birth weight, a birth weight greater than 2.5 kg and less than 4.5 kg is a normal birthweight, and an abnormal birth if the weight exceeds 4.5 kg. Here we use machine learning techniques and algorithms such as linear regression, random forest regression and XGB regression to predict fatal birth weight. XGB regression prediction is more accurate than linear regression and random forest regression. Finally, we regarded the XGB regressor as ourfinal algorithm for predicting fatal birth weight.

7. REFERENCES

[1] M. Feng, L. Wan, Z. Li, L. Qing and X. Qi, "Fetal Weight Estimation via Ultrasound Using Machine Learning," in IEEE Access, vol. 7, pp. 87783-87791, 2019, doi: 10.1109/ACCESS.2019.2925803.

[2] M. W. L. Moreira, J. J. P. C. Rodrigues,

[3] 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.

[4] J. Yu, Y. Wang and P. Chen, "Fetal Weight Estimation Using the Evolutionary Fuzzy Support Vector Regression

for Low-Birth-Weight Fetuses," in IEEE Transactions on Information Technology in Biomedicine, vol. 13, no. 1, pp. 57-66, Jan. 2009, doi: 10.1109/TITB.2008.2007080.
[5] R. Czabanski, M. Jezewski, J. Wrobel, J. Jezewski and K. Horoba, "Predicting the Risk of Low-Fetal Birth Weight From Cardiotocographic Signals Using ANBLIR System With Deterministic Annealing and ${\bm \varepsilon}$ - Insensitive Learning," in IEEE Transactions on Information Technology in Biomedicine, vol. 14, no. 4, pp. 1062- 1074, July 2010, doi: 10.1109/TITB.2009.2039644.
[6] F. Sereno, P. Marques de Sa, A. Matos and J. Bernardes, "Support vector regression applied to foetal weight estimation," IJCNN'01. International Joint Conference on Neural Networks. Proceedings (Cat. No.01CH37222), Washington, DC, USA, 2001, pp. 1455- 1458 vol.2, doi: 10.1109/IJCNN.2001.939576.
[7] G. A. Pradipta and P. D. Wulaning Ayu, "Fetal weight prediction based on ultrasound image using fuzzy C means clustering and Itterative Random Hough Transform," 2017 1st International Conference on Informatics and Computational Sciences (ICICoS), Semarang, 2017, pp. 71-76, doi: 10.1109/ICICOS.2017.8276340.
[8] M. Abdollahian and N. Gunaratne, "Low Birth Weight Prediction Based on Maternal and Fetal Characteristics," 2015 12th International Conference on Information Technology - New Generations, Las Vegas, NV, 2015, pp. 646-650, doi: 10.1109/ITNG.2015.108.
[9] F. Sereno, P. Marques de Sa, A. Matos and J. Bernardes, "Support vector regression applied to foetal weight estimation," IJCNN'01. International Joint Conference on Neural Networks. Proceedings (Cat. No.01CH37222), Washington, DC, USA, 2001, pp. 1455- 1458 vol.2, doi: 10.1109/IJCNN.2001.939576.
[10] R. Czabanski, M. Jezewski, J. Wrobel, J. Jezewski and K. Horoba, "Predicting the Risk of Low-Fetal Birth Weight From Cardiotocographic Signals Using ANBLIR System With Deterministic Annealing and ${\bm \varepsilon}$ - Insensitive Learning," in IEEE Transactions on Information Technology in Biomedicine, vol. 14, no. 4, pp. 1062- 1074, July 2010, doi: 10.1109/TITB.2009.2039644. (ISSN: 2454-132X)

BIBLIOGRAPHY

Dr. Shantakumar B. Patil
Sai Vidya Institute of Technology, Bangalore, Karnataka, India

Y. Sree Rushitha
Sai Vidya Institute of Technology, Bangalore, Karnataka, India

Meghana S.
Sai Vidya Institute of Technology, Bangalore, Karnataka, India

Susan Syeda
Sai Vidya Institute of Technology, Bangalore, Karnataka, India

Priyanka M.
Sai Vidya Institute of Technology, Bangalore, Karnataka, India

