*International Journal of Engineering & Technology*, 7 (2.22) (2018) 5-8

International Journal of Engineering & Technology

*Website: www.sciencepubco.com/index.php/IJET*

***Research paper***

# Machine Learning for High Risk Pregnancies Pre-Term Birth Prediction: A Retrospective

**M. Ramla[1]*, S. Sangeetha[2] , S. Nickolas[3]**

*[1]Research Scholar, Department of Computer Applications, National Institute of Technology, Trichy, India.*
*[2]Assistant Professor, Department of Computer Applications, National Institute of Technology, Trichy, India.*
*[3]Associate Professor, Department of Computer Applications, National Institute of Technology, Trichy, India.*
**Corresponding author E-mail:ramzsami@gmail.com*

**Abstract**

Any birth before 28 weeks of gestation is termed as Pre-Term. This has substantial impact in the emotional reactions of mothers. The post-traumatic stress notably in the mother could be of chronic psychological risk. Moreover, it is to be addressed in the global scenario for sustainable development. Predicting stillbirths is still a distant reality. A plethora of works have been carried out and this paper present the summaries and analysis of current research. The primary focus of the paper is to throw light on the challenging issue of Preterm Birth Prediction. Myriad of machine learning techniques are used by various researchers each with its own estimation accuracy and type of ML model.

***Keywords:*** *Pre-Term Birth, Stillbirth, High Risk Pregnancies, Machine Learning, Predictive Analytics.*

## 1. Background

According to WHO, 'A Stillbirth is a baby born with no signs of life at or after 28weeks of gestation.' It can result in the guilt feel of the mother. The scepticism starts when there is no fetal movement and the conformation is made through ultra sound. India has the dubious distinction of having the highest number of stillbirths in the world according to The British Medical Journal, 'The Lancet'. About 0.75 million neonates die every year in India and our country leads the world. The main cause for StillBirth is poorly understood. But the five main causes are fetal growth restriction, childbirth complications, maternal disorders and infections, congenital abnormalities. The attributes that contribute to stillbirths are numerous. Preterm birth is a challenging and complex real-world problem that pushes the boundary of state-of-the-art data mining methodologies.

## 2. Introduction

Predictive Analytics is the branch of advanced analytics used to make predictions about unknown future events. It combines Data Mining, Statistics, Machine Learning, AI to analyses current data and make predictions for the future.

The objective of Predictive Analytics is to help organizations change information into significant bit of knowledge. Assimilate the volume of information and make decision in optimal fashion bringing a proactive outcome and behaviors based on the data.

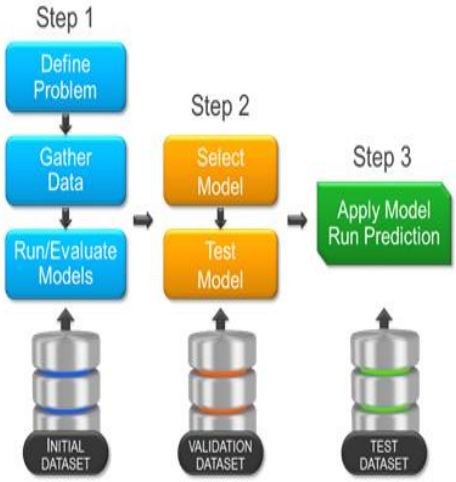


***Fig:1 Steps in Predictive Analytics***

Especially in health care domain, physicians possibly cannot commit to memory all the knowledge they need for every situation. Time and expertise is required to analyze the information and integrate it with the patients own medical profile. Health Care embraces Predictive analytics in a slow pace and making the data available for PA is a challenge.
With the aim to develop a predictive model for early detection of pregnancies at high risk of preterm, a systematic review of Machine Learning techniques used by various researches experimented on different datasets are summarized here.

## 3. Review of Literature

Review of medical literature finds an explosion of articles that apply machine learning to a wide variety of health care problems.

Van Syne et al. (1994) provided a proof of the concept for using machine learning and expert system to predict preterm risks. LERS was used to generate production rules directly from data. [1] A Bayesian classification program was successfully tested on the data sets. Additional machine learning techniques including conceptual clustering and neural networks can be added with the final expert system which will be verified by the experts.

Goodwin (2000) and her colleagues have explored the use of data mining techniques to predict preterm labor. [2] The purpose of their study was to develop tools and techniques to help better understand the causes of premature birth. Five different modelling techniques that used neural networks, logistic regression, CART and software called PVRuleMiner and FactMiner were employed. Receiver Operations Characteristics (ROC) analyses are made as it is particularly used with medical diagnoses. They considered attributes such as Multiple gestation, Uterine Fibroids, Previous Still born, Hypertension, marital status, Race, Poor weight gain etc., New findings detected several small but high-risk subpopulations and several variables had high positive predictive value but were rarely present in the population.

Christina Catley et. al. (2006) employed ANN as a screening tool for preterm birth. ANNs trained using obstetrical data have been found to be a potentially useful clinical decision support tool in the early estimation of PTB [3]. This work does not merely attempt to replace the physicians intuition and judgement , but attempts to augment physicians decision making by providing a screening tool to identify mothers at high risk of deliv-ering prematurely from a heterogeneous population. Attempts to further increase the sensitivity of the model can be made with richer socio demographics data.

Yavar Naddaf et al (2008) suggested that decision support tools are required to help doctors predict preterm births. Different classification techniques such as Naïve Bayes, Decision Trees, SVM, Logistic Regression, and associative classifier are applied on datasets to predict preterm births. Contrast Rule Mining is performed to select attributes that discriminate the most between normal and preterm cases. [4] But the predictive model did not show much improvement as the set of available clinical attributes did bot cover the space required for predicting PTB.

Luiz Fernando et al. (2009) developed a fuzzy model to estimate the probability of neonatal mortality. It was based on the fuzziness of the variables like birth weight, gestational age, APGAR score, and previous report of stillbirth. The model showed a good accuracy and the inference was made using Mamdani's method.

Matharage et al. (2011) worked with an unsupervised clustering technique called Growing Self Organizing Map (GSOM) to analyse the stillbirth data and they presented patterns which can be crucial to medical researchers. [6] GSOM is an unsupervised clustering technique which is used to uncover any hidden patterns among stillbirths. A GSOM was trained to identify groupings in maternal demographic data and current stillbirth pregnancy (CSP) data, 6 clusters based on CSP were found and interesting information about the clusters were revealed.

Vovsha et al. (2014) suggested that previous researchers focus on individual risk factors correlated with preterm birth. Their work considered PTB to be binary classification problem and various logistic regression and SVM methods were applied. This study is a step to harness the health care data to a powerful prediction model. They showed significant improvement on existing results, Moreover, this methodology is used off-the-shelf to handle very unbalanced classes of examples, large number of features and types. The authors conclude by raising the issues of filling in the missing values for nulliparous women, the intractability of the problem, and the volume if data needed for prediction.[7]

Bukowski et al., (2014) showed a population-based case control study of all stillbirths from utmost five geographical areas in the US. They observed that Stillbirth is associated with fetal growth restriction and excessive fetal growth [8]. Enumerated list of characteristics were considered for the study and these can be taken as the vital predictors for supporting the predictive analytic process.

Rudresh .Shirwaikar et al. (2016) focussed in prediction of apnea episodes using machine learning algorithms. They developed a classification model using decision trees C5.0, SVM and ensemble approach which includes random forest. Bagged decision tree with default 25 decision trees and a auto tuned boosted C5.0. Random Forest with default ensemble of 500 trees were used. Reusing the model to predict other neonatal diseases such as jaundice and sepsis can be done further[9].

Kumari Deepika and Seema (2016) have used machine learning algorithms for the diagnosis of diabetes and heart disease. All the patients' data are trained by using different classifiers such as Naïve Bayes, SVM, Decision Tree and Artificial Neural Networks.[10] From experiment, it has been found that SVM gives highest accuracy rate. Comparative study of different classifiers have also been made based on their accuracy rate.

Kayode et al. (2016) developed a stillbirth prediction model by considering the features like maternal comorbidity, residential place, occupation, birth parity, bleeding and presentation of fetus[11]. The model was extended using the fetal growth rate as the major predictor. This was proved to be a promising model in a low resource setting and the same could be externally validated in future.

Rebecca Knowles et al. (2016) leveraged Machine learning and Natural Language Processing to detect high risk pregnancies. They showed that simple text features from unstructured records can outperform the baseline classifications [12]. Only physician's clinical notes were considered for flagging the patient as high risk. Clustering the patients based on the risk factors can also be done.

## 4. Summary of Related Work

| REF No | Objective | Methodology | Data Set / Features | Findings | Gaps/ Future Work |
|---|---|---|---|---|---|
| [1] | Build a prototype expert system using machine learning and statistical analysis on perinatal records | Machine Learning with a program named LERS (Learning from examples using Rough Sets) was used | 20000 samples from Tokos HealthDyne and St.Lukes perinatal centre was collected | High accuracy rate of 88% was achieved in predicting preterm pregnancies | Prediction accuracy was affected because of large number of dichotomous data.<br>Additional techniques like conceptual clustering and neural networks can be included.<br>Expert verification is required.<br>Focus on collecting prospective data samples |
| [2] | Develop tools and techniques to understand the causes of premature birth | Neural Networks, Logistic Regression, CART, PVRule Miner and FactMiner were experimented | Duke University Medical Center TMR Perinatal Data with 63,167 records and nearly 4000 to 5000 variables per record<br>Attributes:<br>Multiple gestation, Incompetent Cervical OS, Uterine Fibroids, Previous Stillborn, Education, Hypertension, Urine Screen, Race, Poor Weight Gain | Several small but high-risk subpopulations and several variables had high positive predictive value but were rarely present in the population | Results on TMR data did not perform well because several variables had high positive predictive values but were rarely present in the population. |
| [3] | Assess if ANN has the potential use in obstetrical outcome estimates | The backpropagation feedforward ANN was trained and tested on cases with 8 input variables describing the patients obstetrical history | Perinatal Pattnership Program of Eastern and Southeastern Ontario (PPPESO) Niday database<br>Attributes:<br>Smoking, Intention to breast feed, Previous PTB, Previous Term Birth, Singleton Birth, Gender, Age, Parity | Offers trained ANN as a webservice to physicians for the clinical prediction for mothers at risk of PTB | A maximum sensitivity of 36.6% was achieved using only a limited 8 variable input available prior to 22 weeks of gestation. Decreased sensitivity and correct classification rate is the current limitation. Attempt to increase the sensitivity can be made base on richer sociodemographic data and obstetric history. |
| [4] | Predicting Preterm Birth based on maternal and fetal data | Applied many classification techniques like Naïve Bayes, Decision Trees, SVM, Logistic Regression, associative classifier on historic maternal and fetal records | DataSet: Northern and Central Alberta Perinatal Outreach Program, 243948 cases including 21193 preterm cases. 244 Attributed | Feature selection using Contrast Set mining was proposed to improve the accuracy of prediction | Prediction performance was very poor even after doing feature selection by contrast mining. The reason may be that demographic information are not considered and they are an important discriminator |
| [5] | Fuzzy Linguistic Predictive model to estimate probability of neonatal mortality | Fuzziness of the variables were considered and the risk of neonatal was given as percentage. | Real data file from Brazilian city was used for validation of the computing model | Mamdani approach of Inference was used | Accuracy was 0.9. The number of fuzzy rules grow exponentially and this can impair the models performance |
| [6] | Analyse stillbirth data and present patterns to medical researchers. | An unsupervised clustering techniques GSOM was used | Experimented on stillbirth database at a tertiary referral hospital in Australia.<br>Attributes: Prolonged rupture of membranes, Hypertension, vaginal bleeding etc., | 6 clusters on demographic data and 9 clusters on CSP data were formed.<br>Discovered that the main risk factor is Consanguinity | Data available for time-based analysis is inadequate. |
| [7] | An application of ML towards the problem of predicting preterm birth | PTB is taken as a binary classification problem with SVM and Logistic Regression | Dataset: Maternal fetal Medical Unit Network (MFMU)<br>3073 cases were studied.<br>Top Features: Term Delivery, Marital status, Race, Mom age, Parity, Income, Previous Preterm deliveries. | Linear SVM provides a robust baseline for the quality of performance.<br>Sensitivity: 40% | How to deal with classes that overlap? Filling in the missing values for nulliparous women does not make sense |
| [8] | Establishing the relationship between the fetal growth and risk of still birth | Fetal growth abnormalities were categorized as SGA(Small Gestational Age) ang LGA (Large Gestational Age) | Stillbirth Collaborative Research Network (SCRN) population-based case-control study | Observed that stillbirth is associated with fetal growth restriction and excessive fetal growth. | Study reveals the vital predictors for predicting the PTB. |
| [9] | Prediction of | Decision Tree C5.0, SVM, | 229 samples of neo- | Random forest algo- | Model can be reusable to other diseases. |

| | | | | | |
|---|---|---|---|---|---|
| | Apnea episodes during first week of child birth | Random Forest was used. | nates from Neonatal Intensive Care Unit (NICU) of Kasturba Hospital. | rithm with accuracy of 0.88 and kappa of 0.72 was found to be most accurate model<br>Sensitivity is low because of class imbalance | Undersampled data leads to decrease in accuracy and specificity.<br>Oversampled data leads to moderate accuracy, sensitivity and specificity |
| [10] | Chronic Disease prediction using Data mining | Naïve Bayes, Decision Tree, SVM, ANN are used for the diagnosis of diabetes and heart disease | UCI Machine Learning Repository | SVM gave high accuracy (95.55%) for heart data set and Naïve Bayes proved good (73.58%) for Diabetes dataset | Utilization of the model for individual or clinical decision support system can be done further. |
| [11] | Prediction model for early detection of pregnancies at high risk of stillbirth | All potential predictors were entered into a multivariate logistic regression. Significant predictors were identified using stepwise backward selection | Federal Medical Center, Bida Nigeria.<br>6573 cohort was studied | External validation of the model can be done to include data related to maternal HIV status as recommended. | Prediction is done on a low resource setting. Extended model can be developed by integrating ultra sound. |
| [12] | High risk pregnancy prediction from Clinical Text<br>Reduce time spent in reading patient records. | Supervised binary classification problem based on the free text associated with each patient. Flag the patients as high risk. Uses machine learning with NLP | John Hopkins Health Care LLC<br>15,028 records<br>Baseline Accuracy performs at 56.8% | Simple text features from unstructured records can outperform baseline classification | Only structured data from clinical physician notes are considered. Clustering of patients based on their risk factors can be done.<br>Combine structured and unstructured data to improve prediction. |

## 5. Discussion

Different authors have worked with multitude of machine learning algorithms like decision trees, Neural Networks, Logistic Regression, CART, SVM, Naïve Bayes, GSOM, Random Forest for prediction of pre-term pregnancies. Few authors have also considered the fuzziness of the predictor variables. It is anticipated that further works will make use of yet -to-be-explored data mining techniques to delve deeper into mining the massive volume of data. A fertile area of research would be to increase the prediction accuracy through deep learning of the data.

## 6. Conclusion

A good deal of research has been carried out on diagnosis of Preterm pregnancy. As health care data is highly sensitive and it requires privacy preserving and data anonymity, effective concerns to bridge the gap in the data is a challenge. Moreover, prediction of preterm pregnancy in a low resource setting has been done and only very few factors are taken into consideration. The selection of machine learning approach depends on the nature of data sets. It is best to combine the approaches for enhanced accuracy. Further research should emphasis on these aspects to build an effective evidence-based medicine.

## References

[1] M.M.Van Dyne, L.K. Woolery, J.Gryzmala Busse, C.Tsatsoulis, "Using Machine Learning and Expert Systems to Predict Preterm delivery in Pregnant Women", DOI: 10.1109/CAIA.1994.323655, IEEE

[2] Linda Goodwin, Sean Maher," Data Mining for Preterm Prediction", 2000 ACM 1-58114-239-5/00/003.

[3] Christina Catley, Monique Frize, C.Robin Walker and Dorina C. Petriu, 'Predicting High-Risk Preterm Birth Using Artificial Neural Networks', IEEE Transactions on Information Technology in Biomedicine, Vol.10, No.3, July 2006

[4] Yavar Naddaf, Mojdeh Jalali Heravi, Amit Satsangi, 'Predicting Preterm Birth based on Maternal and Fetal data', Google Scholar, 2008

[5] Luiz Fernando, Paloma Maria S, Racha Rizol, Luciana B.Abiuzi, ' Establishing the risk of neonatl mortality using a fuzzy predictive model', Cad Saude Publica, 2009

[6] Matharage S. Alahakoon O., Alahakoon D, Kapurubandara S, Nayyar R, Mukherjee M, Jagadish U, Yim S, Alahakoon, ' Analysing Stillbirth Data using Dynamic Self Organizing Maps', 2011 IEEE DOI 10.1109/DEXA.2011.14

[7] Illia Vovsha, Ashwath Rajan, Ansaf Salleb-Aouissi, Anita Raja, Axinia Radeva, Hatim Diab, Ashish Tomar, Ronald Wapner, "Predicting Preterm Birth is Not Elusive: Machine Learning paves way to individual Welness', AAAI Symposium Series, 2014

[8] Radek Bukowski et al., 'Fetal growth and Risk of StillBirth: A population based case-control study', PLOS Medicine, Vol:11 Issue:4

[9] Rudresh D.Shirwaikar, U.Dinesh Acharya, Krishnamoorthi Makkithaya, M.Surulivelrajan, Leslie Edward Simon Lewis, ' Machine Learning Techniques for Neonatal Apnea Prediction', Journal of Artificial Intelligence, DOI: 10,3923/jai.2016.33.38 (2016)

[10] Kumari Deepika, Dr.S.Seema, "Predictive Analytics to Prevent and Control Chronic Diseases", 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATT), 2016

[11] Gbenga A.Kayode, Diederick E.Grobbee, Mary Amoakoh-Coleman, Ibrahim Taiwo Adeleke, Evelyn Ansah, Joris A.H. De Groot, Kerstin Klipstein-Grobusch, "Predicting Stillbirth in a low resource setting", DOI:10.1186/s128884-016-1061-2, BMC Pregnancy and Child Health., 2016

[12] Rebecca Knowles, Mark Dredze, Kathleen Evans, Elyse Lasser, Tom Richards ,JonathanWeinerc, Hadi Kharrazi, "High Risk Pregnancy Prediction from Clinical Text", Johns Hopkins HealthCare LLC, 2016.