CMPE-257 Machine Learning

Individual Research Project

Machine Learning for Preterm Birth Prediction

Harshada Bhide-Apte

Department of Computer Engineering

Fall 2017

Table of Contents

Abstract

Normal pregnancy involves a full-term of 40 weeks gestation. The problems associated with pre-term child birth can be avoided with the system that can better predict pre-term delivery. But accurately predicting preterm birth can be very difficult using current manual methods of prediction, even by the most experienced doctors. Hence, the medical information systems based on advanced machine learning techniques are vital in predicting possible preterm child births and thereby helping to avoid them. Machine learning algorithms and neural networks are used for detecting biological patterns in the existing medical records data of pregnant women. The data preprocessing and validation techniques are used to classify preterm births from full-term births. Large observational datasets are available of medical records of pregnant women who had full-term and pre-term deliveries. Algorithms can be applied to learn from these datasets. Various feature extraction and selection techniques are used to process un-correlated features to evaluate them for predicting preterm birth. The expert systems have proven to be more accurate in predicting preterm birth. Research is still being conducted on various algorithms and classifiers for improved predictive accuracy. This paper intends to study various machine learning techniques used in the preterm birth prediction systems.

1 Introduction

Normal pregnancy or full-term pregnancy has a 40 weeks gestation period. Preterm birth, or premature birth, is the birth of a baby at fewer than 37 weeks gestational age. Babies born of preterm birth are at greater risk for medical problems such as cerebral palsy, delayed development, hearing and sight problems. Preterm birth is the most common cause of death among infants worldwide. The health risks are higher, the earlier a baby is born. Uterine contractions are considered as major symptoms of preterm labor. But the exact causes of preterm birth are still not precisely known. Diabetes, extreme weights, high blood pressure, being pregnant with more than one baby, vaginal infections, smoking and stress, among others are some of the factors that increase the risk of preterm birth. Each year 5% to 18% of all deliveries happening worldwide are premature.

Many machine learning scientists have done substantial work to develop an expert system for preterm delivery risk assessment of pregnant women. Machine learning techniques use knowledge based development methodology and validation techniques to analyze medical databases containing information collected from women during and just after their pregnancy. The dependent variable or decision variable is usually weeks of gestation at delivery. The goal of these expert systems is to improve clinical outcomes for preterm birth risk assessment through the development of a knowledge base and using that knowledge base to induce rules to predict decision variables.

1.1 Preterm Birth

Preterm birth also known as premature birth is usually defined as birth before 37 completed weeks of gestation. But over the period of last few decades, many different studies have considered slightly different number of weeks as preterm birth. Preterm is defined ambiguously in the literature of the prior research performed about using machine learning for predicting preterm birth. The preterm births can be classified into two categories spontaneous and indicated. The births which happen naturally without medical interventions are called Spontaneous births. A small number of births may not come under either of the two categories.

1.2 Nulliparous pregnancy

The first-time mothers are called nullipara or para 0. As these women have no prior record of pregnancy, it is more difficult to predict preterm birth risk in their case. About 40% of pregnancies in the United States fall under this category. In practice, prior history of premature birth is used as the most predictive indicator in clinical settings. Since the major input variable of prior history of preterm or full-term labor is not available for these women, a stronger system which can still predict accurate preterm risk is necessary.

1.3 Benefits of preterm birth prediction

Preterm birth can be delayed or even prevented, if proper medical care is provided to the pregnant women who are at high risk of premature labor. The hormone progesterone, if taken during pregnancy, may prevent preterm birth. Many other medications including corticosteroids and nifedipine may affect outcomes. That’s why, predicting preterm birth in advance can reduce the medical risks to the mother and the infant. There are measures which can be taken to provide extra care for those who need it. The important thing is to make accurate predictions about which pregnant women will have preterm birth and which won’t.

2 Background

Traditionally the predictions about preterm birth are made manually. Manual risk assessment tools are only 17% to 38 % accurate. This poor predictability is mainly because of the lack of underlying conceptual model of preterm birth risk. Many risk scoring and screening instruments are available, but no conceptual or theoretical model of preterm risk has been reported, which may account for the poor reliability and validity of traditional manual screening techniques. Another problem related to premature delivery risk assessment is poorly defined and complex information base. The knowledge base about preterm labor is unorganized and not validated through research. This makes it of little use to patients and providers of prenatal care. The poor accuracy of these prediction models has led to the increasing trend to treat all pregnancies as though they are at "high risk' for premature delivery.

2.1 The Coefficient of Premature Delivery Risk (CPDR)

Papiernik-Berkhauer proposed this empirical method of predicting preterm birth in 1969. [4] In this method, various maternity attributes are grouped into four categories – social status, work conditions, obstetric history and pregnancy characteristics. These attributes are presented in a two-dimensional table. They are allotted priorities ranging from 1 to 5 according to the importance given to them by perinatal experts. The sum of the priorities gives the risk of preterm birth. This system proved to have much lower performance accuracy when used on different population. It was later modified in the 1980 and 1996 to improve performance. But still is considered a poor measure for the preterm birth risk assessment.

2.2 Machine Learning

Over the past few decades four paradigms of machine learning have been considered for use in preterm birth risk assessment systems. [1]

1) Analytic paradigm – It uses a strong underlying theory to learn and build concepts. The lack of strong underlying theory made this paradigm inappropriate to use for premature delivery risk prediction.
2) Genetic paradigm – It creates new solutions to a problem and then tests their fitness. It was also considered not very useful in preterm birth prediction.
3) Neural network – Neural network or the connectionist paradigm recognizes patterns in input data and classifies the input. It learns how to classify based on previous classifications available on similar input patterns. This paradigm is very commonly used in preterm birth risk prediction systems.
4) Inductive paradigm – It works on the sets of data where the classification of an example is known, and the system learns to discriminate between different classifications given the data values associated with different examples. This is used in some of the expert systems developed to predict preterm birth.

2.3 LERS

A machine learning based program named Learning from Examples using Rough Sets (LERS) generated 520 meaningful rules to be used in a preterm birth prediction expert system prototype. [1] The inductive paradigm was used for this system. It was developed in the Computer Science Department of the University of Kansas. This system uses rough set theory for managing uncertainty in knowledge acquisition. Unlike other approaches the uncertainties are not removed from the decision table and ignored by the learning algorithm. LERS computes the lower and upper approximations for the preterm birth and full-term birth concepts. Then it induces the rules from these approximations. LERS assumed equal priorities for all the attributes. Therefore, no bias was added. LERS used 3 different large datasets. Because of their varying sizes and dissimilar variables, each database was processed separately. Exploratory factor analysis was conducted on the dataset followed by multivariate regression analysis. Rules induced from the data were validated using content validity techniques and perinatal experts. LERS tested the machine learning algorithms with data that contained missing values, obvious errors and inconsistencies. All the 520 rules were verified by two certified perinatal nurses and were deemed useful. But it was asserted that each individual rule does not provide sufficient information. It was also able to determine which examples in the test data remained unclassified by the rules, thereby pointing out where more data and rules were needed. [1]

2.4 Neural Network classification based on EHG records

The open dataset used for this research contained medical records of 300 pregnant women (262 had full term delivery and 38 had preterm delivery). The raw data containing EHG signals was obtained using three different channels hence there are three signals per record. It was pre-processed using in multiple steps including data segmentation, feature extraction and relevant feature classification using linear as well as non-linear machine learning techniques. Then several feature extraction techniques were used to extract features from the raw data. Using various measures these features were ranked and the top four unrelated features were selected for further analysis in the classification stage. Advanced artificial neural network classifiers of seven different types were used on these features and several validation techniques were considered to determine their overall accuracy. The results of these validation techniques were compared leading to the conclusion that the Radial Basis Function Neural Network(RBNC) gave the best output. To address the issue of oversampling, Synthetic Minority Over-Sampling Technique (SMOTE) was used to generate total of 262 preterm birth records from the original 38 records. This new dataset is evenly split between term births and preterm births. This study evaluates some of the back-propagation trained feed-forward neural network classifiers and some perceptron classifiers. [3]

These results were obtained by evaluating each classifier using predetermined measures on the original dataset. Each experiment was conducted 30 times on randomly selected test data. After applying SMOTE technique for oversampling preterm records, the results were obtained for the new balanced dataset. The results showed the increase in sensitivity and decrease in specificities for all algorithms. They also showed drastic increase in the accuracy of the classifiers. The mean error rates, produced by these seven classifiers, are lower than the expected baseline error rate from the initial results. The significant improvement in overall results indicated that the additional features provide better separation between term and preterm birth delivery. [3]

2.5 Columbia University Research

The research was performed by comparing two approaches for developing prediction systems – SVM approach with linear and non-linear kernels and logistic regression. The research paid special attention to the first-time mothers as their pregnancy history is not making the risk prediction more difficult. The analysis was performed on the data from 10 different medical centers which constituted a very rich database. Various steps were performed to harness this data and organize it in a structured fashion. The complexity of the data was handled by organizing features into groups. Each dataset was randomly split into training and test data in 80-20 ratio. The research has derived predictive models for different stages of pregnancy. The preterm birth prediction is considered as binary classification problem with full-term birth and preterm birth being the two output classes. Every patient is described as a vector of all the features and a label, (x_i, y_i), $\{+1, -1\} \ni y_i$. Then the researchers applied SVM and logistic regression and compared the error rates.

3 Technical Descriptions

Manual preterm birth prediction has yielded poorer results over the period. But it is feasible to use machine learning to develop knowledge-based expert system rules for preterm birth risk assessment. Many studies have proved them to be more accurate than traditional methods. Machine learning algorithms of varying strengths and restrictions are developed to extract patterns from data for the creation of classification systems, decision trees and production rules.

3.1 Oversampling of dataset

Only 5% to 18% of all deliveries are preterm. Therefore, most prenatal datasets are unbalanced and contain more full-term records than preterm records. This has a significant impact on machine learning algorithms, as classifiers are more prone to detecting the majority class. Hence, the probability of detecting a premature birth becomes low. To solve this issue, minority class of preterm birth is oversampled. Techniques like the Synthetic Minority Over-Sampling Technique or SMOTE are used for oversampling preterm class.

3.2 Inductive Expert Systems

It was used in systems like LERS. It works with a decision table that presents data from the real-world which is used for decision making. In the decision table examples are characterized by attributes and decisions. The values of decision variables are provided by experts in the field. A concept, denoted as [(d, w)], is a set of all examples that have value w for decision d. An attribute-value pair is selected by looking for attributes with the highest priorities. Attribute priorities are allocated by a domain expert. The next criteria for attribute selection is its relevance to the target. The lower approximation of a concept is the set of all examples that certainly adhere to the concept after considering all attributes. The upper approximation of a concept is the set of all examples that possibly adhere to the concept. The rules computed from lower approximations are “certain rules”, and the rules computed from upper approximations are called “possible rules”. Certain rules are completely supported by the training data. Possible rules are supported by some of the data while contradicted by some other data in the same dataset.

Possible rules are quantified by “rough measure” which is the ratio of the number of all examples correctly described by the rule to the number of all examples described by the rule. The higher the rough measure the more reliable the rule. [1]

3.2.1 Sample decision table

Example (Patient)	Features			Decision (Delivery)
	Pregnancy No.	Age of mother	Bleeding	
P1	1	< 20	yes	Preterm
P2	3	30-39	No	Fullterm
P3	2	20-29	No	Preterm
P4	2	20-29	No	Fullterm
P5	1	20-29	Yes	Fullterm

3.2.2 Establishing concepts

From the above decision table two concepts are built as follows -

Concept	Examples set
[(delivery, preterm)]	{P1, P3}
[(delivery, fullterm)]	{P2, P4, P5}

3.2.3 Computing Approximations

The lower and upper approximations are derived from the data as follows –

Approximation	Concept	Examples Set
Lower approximation	[(delivery, preterm)]	{P1}
	[(delivery, fullterm)]	{P2, P5}
Upper approximation	[(delivery, preterm)]	{P1, P3, P4}
	[(delivery, fullterm)]	{P2, P3, P4, P5}

3.3 Neural network based system

The raw data is pre-processed using various techniques like data segmentation, feature extraction and classification. A features vector is generated from the extracted features. The features are then ranked using several measures like statistical significance and linear discriminant analysis.

Then top unrelated features are selected for further analysis. Many advanced artificial neural network classifiers have been used for preterm prediction systems. The major classifiers which proved most useful are [3] –

1) Back-propagation trained feed-forward classifier (BPXNC) – To train this network, the inputs are fed forward to optimize the weights. Then the error is fed backwards to further optimize the weights. Then it uses the input-output values from training data to improve weights to reduce the difference between predicted and observed output values. This training cycle is repeated multiple times to reduce the error and get the best possible output accuracy.
2) Levenberg–Marquardt trained feed-forward classifier (LMNC) – Its working is much like BPXNC but in a memory intense way. It also has a well-defined stopping point for training cycles. The training stops when the performance of 1000 artificially generated tuning examples is reached per class there by leaving no scope for further improvement of performance.
3) Perceptron linear classifier (PERLC) - It uses supervised learning algorithm in which it assumes that true outputs of the training data are available and uses them for training purpose. The inputs are adjusted iteratively to product the accurate output. It is the simplest neural network classifier but has the disadvantage that it does not have hidden layers. That's why it must use bias in output accuracy.
4) Radial basis function classifier (RBNC) – It has only one hidden layer of neurons which contains radial basis units. It is efficient for complicated pattern recognition and is widely used in biomedical systems. Its mapping properties can be adjusted by weights in the output layer.
5) Random neural network classifier (RNNC) – It's a feed forward network with one hidden layer containing sigmoid neurons. This hidden layer has equally distributed weights and biases with zero standard deviation and mean. To achieve this the input layer scales the features to achieve unit variance.
6) Voted perceptron classifier (VPC) – It is an advanced perceptron neural network algorithm. It works with linearly separable datasets having large margins.

3.4 Support Vector Machines (SVM)

Researchers at Columbia university have use the support vector machines with linear and radial basis function kernels for their expert systems. The skewness of the datasets was handled by modifying the cost function as follows,

$$\min_{w,\mathcal{E}} \ \frac{1}{2}\,||w||^2 + C_- \sum_{y_i=-1}^{n} \mathcal{E}_i + C_+ \sum_{y_i=1}^{n} \mathcal{E}_j$$

$$s.t \quad y_k\,[w^T x_k + b\,] \geq 1 - \mathcal{E}_k, \qquad \forall k$$

Where, x = features vectors

y = output labels

w = weight vector

Ɛ =slack variables

C_{-}, C_{+} = regularization parameters

Overall equal weight was given to each class by assigning different penalties for misclassification.

3.5 Logistic regression

Logistic regression is widely used in biomedical domain. To handle the skewed datasets oversampling techniques are used in this method. This method was experimented by Columbia university researchers and the results were compared with their SVM results. Logistic regression model selection methodologies [4] –

1) Forward selection – It is a greedy algorithm. At each step, it selects the covariate that best improves the fit. The stopping point is when adding covariates is not productive anymore. In this method, a predictor is never removed once it has been added.
2) Stepwise selection – It is similar in working to forward selection. But it also allows to remove an already added predictor to improve the fit.
3) Lasso regression – This method assigns penalties to generate sparse solutions for huge datasets by eliminating trivial covariates. It also performs feature selection at a level. This method is proven to yield better results compared to forward selection and stepwise selection models.
4) Elastic net regression – This method is used when there is correlation between different predictor variable groups. This correlation could be directly obvious or indirect relation. This model combines the sparsity induction and ridge regression. In sparsity induction, trivial predictor variables are omitted from the analysis of huge datasets. Ridge regression, on the other hand includes the complete group of covariates if one of the predictor variables from it is added for analysis.

3.6 Stabilized Sparse Logistic Regression (SSLR)

It is based on is based on l_1-penalized logistic regression and is stabilized by a graph of feature correlation. [5] In this method BOOTSTRAP is used to compute the importance of the features. The feature weights are averaged. Feature importance is computed as averaged weight * feature standard deviation. This enforces the feature strength and stability as it is insensitive to scaling. The top k most important features are used for generating the rules. The weights of these features are linearly transformed and rounded to integers. This method builds a linear prediction rule. For binary classification, it uses logistic regression to compute the probability outcome –

$$P(y = 1\,|x) = \frac{1}{1 + e^{-f(x)}}$$

To deal with the large number of features in the medical dataset, a sparse solution is derived by minimizing l_1 – penalized loss which is calculated using the formula –

$$L = -\sum_{d=1}^{n} \log P(\,y^d|x^d) + \lambda \sum_i |w_i|$$

Where, d= data points

λ=0 is the penalty factor

The penalty factor reduces the weights of the trivial features to near 0. But the introduction of sparsity may lead to instability of the model. To make the model more stable, correlated features are given similar weights.

$$L = L_0 + \lambda \sum_i (\propto |w_i| + \frac{1-\propto}{2}(\,w_i - \sum_{j \neq i} S_{ij} w_j)^2)$$

Where,

$L_0 = -\sum_{d=1}^{n} \log P(\,y^d|x^d)$,

$\propto \in [1,0]$,

$S_{ij} > 0$, similarity between features i and j

Similarity matrix S can be computed using various methods like calculating the cosine between data columns. For preterm analysis α is set to 0.5 and λ is set to 5.

The prediction rule is used to calculate the scores for all the patients which are converted into risk probability using univariate logistic regression thus producing a risk curve.

4 Discussions

Any preterm birth risk assessment system should aim to develop easy to interpret prediction rules. The uncertainties in these rules should be quantifiable for them to be useful in clinical practice. These expert systems should have accurate classifiers to distinguish between preterm and full-term birth possibilities. These systems should allow a compact set of risk factors. They should be able to provide a reasonable explanation for the prediction made. The prediction rules should be usable in practice and have interpretability and stability. This enforces that the rules should be self-explanatory and provide stable results over data resampling rendering reproducibility.

4.1 Challenges in creating an expert system for preterm birth prediction

1) Ambiguity of strong underlying concept – There is no one strong conceptual model for preterm birth risk assessment. This poses the first problem in deciding the target for the expert system.
2) Problems in data acquisition – Data is inconsistently recorded and sometimes the data attributes needed were not collected. Many times, patients miss a visit, withdraw or change medical care provider.
3) Uncertainty of dataset – It is caused by data errors, ambiguity of exact meanings of data, and doubtful connections between conditions and decisions.
4) Inconsistency in data - Inconsistency is observed in datasets where two examples having identical attribute values have different decision values.
5) Data organization – Data is often voluminous and poorly organized making it difficult to process it.
6) Unrelated data – The health care providers often collect a great deal of data that has little to do with preterm birth risk.
7) Skewed class distribution – As the number of women delivering prematurely is very less, the class distribution between full-term and preterm births is very uneven. This creates a difficulty for classification algorithms to product accurate output.
8) Leakage Problem – This is the biggest problem in observational datasets which affects feature generation. In this problem, the recorded information implicitly gives away the decision to be predicted. The results of the tests performed late in pregnancy may indicate the possibility of preterm birth.

4.2 Specificity and Sensitivity of algorithms

The algorithm sensitivity is considered as the ability of an algorithm to correctly identify the pregnancy cases with preterm birth (true positive rate). On the other hand, the specificity of an algorithm is the ability of the algorithm to correctly identify the pregnancy cases with full-term birth (true negative rate).

The specificity and sensitivity values were taken from the published research papers which used various machine learning algorithms for preterm birth risk assessment systems. For some algorithms, slightly different values of sensitivity and specificity were observed when tested on different datasets or for each data collection point during pregnancy. I have taken the average of those values to compute one single value per algorithm.

The following formulae are used to calculate specificity and sensitivity.

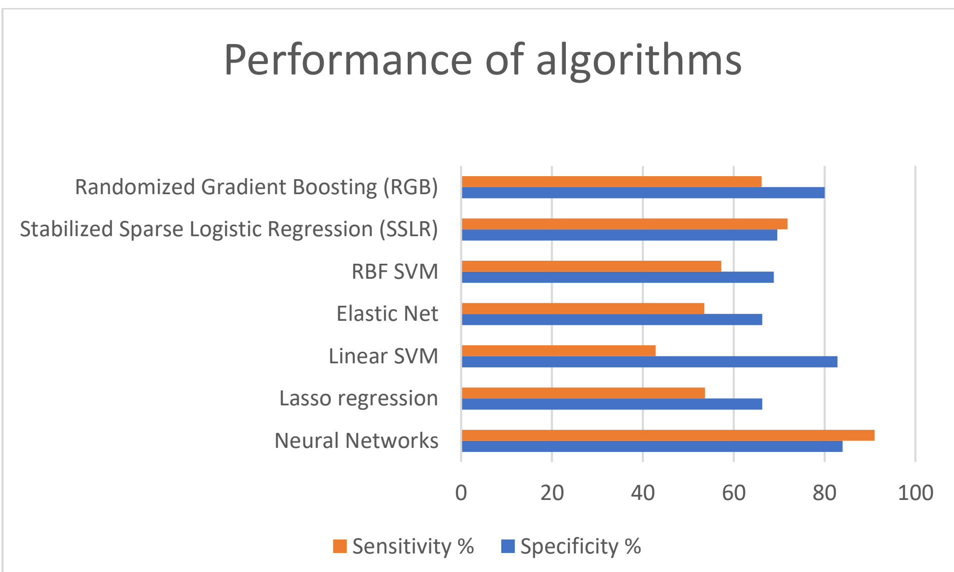
$$Sensitivity = \frac{TP}{TP + FN}$$

$$Sensitivity\ \% = Sensitivity * 100$$

$$Specificity = \frac{TN}{TN + FP}$$

$$Specificity\ \% = Specificity * 100$$

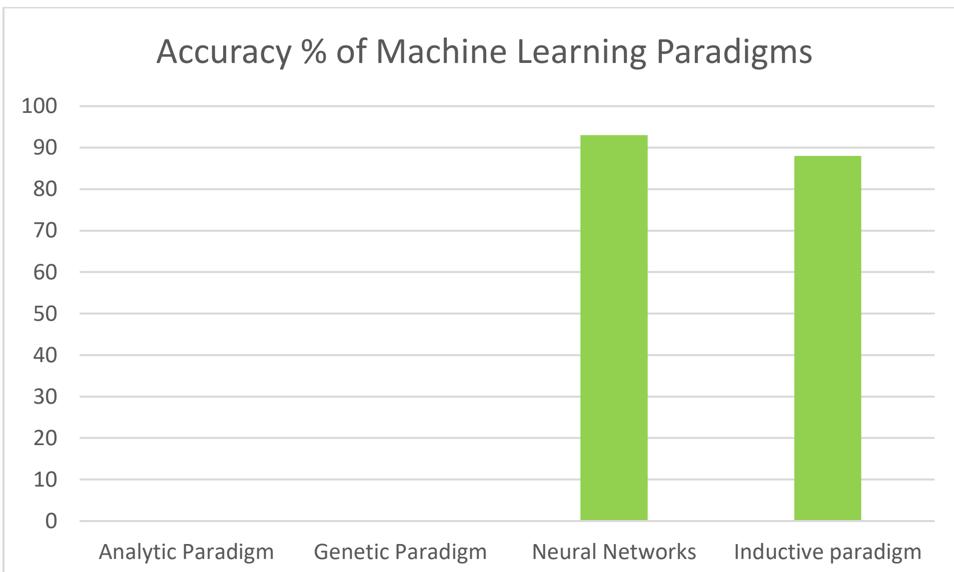
Performance of Machine Learning Algorithms		
Algorithm	**Specificity %**	**Sensitivity %**
Neural Networks	84	91
Lasso regression	66.3	53.7
Linear SVM	82.83	42.83
Elastic Net	66.3	53.5
RBF SVM	68.8	57.26
Stabilized Sparse Logistic Regression (SSLR)	69.59	71.83
Randomized Gradient Boosting (RGB)	80.03	66.17



4.3 Accuracy of various machine learning paradigms

Multiple researchers have attempted to use different machine learning paradigms to tackle the problem of preterm birth prediction. Analytical and genetic paradigms have proved to be inappropriate for premature labor risk assessment systems. Inductive paradigm has yielded very good accuracy of 88% compared to traditional methods. In the recent research results, neural networks seem to be performing much better than any other previously attempted paradigms. The k-nearest neighbor neural network algorithm is the most precise in classifying the prenatal features, to yield the prediction output with a maximum accuracy of 93%.

Accuracy of Machine Learning Paradigms		
Paradigm	Prototype System	Accuracy %
Analytic Paradigm	Not useful	-
Genetic Paradigm	Not useful	-
Neural Networks	Neural Network classification based on EHG records	93
Inductive paradigm	LERS	88



5.0 Summary

The development of prenatal expert systems is important for predicting possible preterm deliveries and thereby avoiding them. This survey of research articles in the combined field of biomedical and machine learning domains proves that machine learning techniques improve the accuracy of preterm birth prediction compared to the traditional methods. The focus of the current research being performed in this regard is focused on improving the sensitivity of the classifier algorithms, because it is vital to predict even the borderline pre-term births in comparison with miss-classifying a full-term pregnancy. The extremely advanced machine learning algorithms for detecting biological patterns in the medical records by data preprocessing and validation techniques are used to classify preterm and term births.

The definition of preterm birth is ambiguous and the databases are less amenable to study. Preterm risk assessment is a complex and disorganized knowledge domain. This makes it difficult to design an expert system for preterm birth prediction. But database quality is increasingly getting better due to the carefully planned and quality controlled data collection methods. Each of the systems prototyped based on machine learning has surpassed the manual prediction accuracy. It seems that preterm birth prediction may not fit a linear model and alternative analysis may be more appropriate in future studies.

References

[1] Linda K, Jerzy G., "Machine Learning for an Expert System to Predict Preterm Birth Risk", in Journal of the American Medical Informatics Association Volume 1, Number 6, Nov -Dec 1994, Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC116227/pdf/0010439.pdf

[2] Linda K, Jerzy G., "Improving Prediction of Preterm Birth Using a New Classification Scheme and Rule Induction", Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2247776/pdf/procascamc00001-0707.pdf

[3] Paul Fergus, Ibrahim Idowu, Abir Hussain, Chelsea Dobbins, "Advanced artificial neural network classification for detecting preterm births using EHG records", in Neurocomputing, 0925-2312, (2015), Available: http://www.sciencedirect.com/science/article/pii/S0925231215017695

[4] I. Vovsha, A. Rajan, A. Salleb-Aouissi, A. Raja, A. Radeva, H. Diab, A. Tomar, R. Wapner, "Predicting Preterm Birth Is Not Elusive: Machine Learning Paves the Way to Individual Wellness", The AAAI Spring Symposium.

[5] T. Tran, W. Luo, D. Phung, J. Morris, K. Rickard, S. Venkatesh, "Preterm Birth Prediction: Deriving Stable and Interpretable Rules from High Dimensional Data", July 2016, Available: https://arxiv.org/pdf/1607.08310.pdf