JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

PRETERERM BIRTH PROGNOSTIC PREDICTION USING CROSS DOMAIN DATA FUSION

[1] Dr S.Chandra Sekaran, [2]B.Neelu, [3]M.Murali

[1] Professor, Department of Computer Science and Engineering, P.S.V. College of Engineering and Technology, Krishnagiri.

[2]Assistant Professor, Department of Computer Science and Engineering, P.S.V. College of Engineering and Technology, Krishnagiri.

[3] PGStudent, Department of Computer Science and Engineering, P.S.V. College of Engineering and Technology, Krishnagiri.

ABSTRACT

In recent times most of the babies are delivered preterm due to various factors. The risk of complication for premature newborns are increased in future (example: brain problem, breathing problem, hearing problem etc.,). Sometimes infant mortality rate increased due to Preterm or premature birth. There is a need to predict preterm birth earlier is necessary. To analyze the data of preterm birth only with medical data predictive analytics is gives the partial insights about preterm birth. In addition to that mother's clinical data, the behavioral, psychological, environmental circumstances data needs to integrated to analyze and design a prognostic model to prevent and predict preterm birth. In this paper we will proposed that cross domain data fusion prognostic model that use the multiple domain data other than medial data is fused to predict the pattern of preterm birth, with the big data classification algorithm techniques gives the best accuracy of predicting preterm birth.

KEYWORDS

Preterm birth, big data, crosses domain data fusion, classification algorithm, data integration, and predictive analytics.

INTRODUCTIN

Preterm birth or Premature as a birth occurring before 37 gestation weeks of pregnancy. normal pregnancy is considered to last about 40 weeks. Preterm birth (PTB) is a serious public health problem that adversely affects both families and the society . It is a leading cause of neonatal mortality and illness across the world and also the second major cause of child deaths under the age of five years . A baby born earlier than 37 gestation weeks is referred to as a preterm birth (PTB) or 259 days after a woman's last menstruation. Around the world, nearly sixteen million babies are born prematurely each year, taking into account even more than around one in ten children. On the other hand, preterm birth rates differ greatly across the world. Premature birth is the leading cause of infant mortality and morbidity , Premature babies, especially those born very early, often have complicated medical problems. Such as Small size newborn, with a disproportionately large head , Fine hair covering much of body. Laboured breathing or respiratory distress and many more long term and short term complications :

short-term complications: Breathing problems, Heart problems, Brain problems, Temperature control problems, Blood problems, Metabolism problems, Immune system problems

long-term complications: Cerebral palsy, Impaired learning, Vision problems, Hearing problems, Dental problems.

PTB Classification

PTB is classified into several categories based on the pregnancy week during birth. The gestational age is between a woman's starting date of her last normal menstrual period (LMP) and her due date . The following are the four types of PTB :

1 Extreme PTB : The range is under 28 weeks of pregnancy. When a baby is born just before the 28th week of pregnancy, it is called extreme PTB.

2.Very PTB : The range is between 28 to 32 weeks of pregnancy. It is noted that between 28 and 32 weeks of gestation baby is born.

3.Moderate PTB : They range from 32 to 34 weeks of pregnancy. It is noted that between 32 and 34 weeks of gestation baby is born.

4.Late PTB: The range between 34 to 37 weeks of pregnancy.

Challenges and Difficulties:

Early detection of pregnancies with a heightened hazard of spontaneous preterm birth (sPTB) could aid premature babies to have fewer stillbirths and side effects later in life. PTB is detected in about half of all women with no identified clinical risk features. PTB rates were not reduced by PTB diagnostic procedures, including an obstetric consultation, mother's characteristics, or a transvaginal ultrasound check-up of a cervix PTB stays a challenging and composite real-world challenge. And also, the nature of pregnancy data presents a challenge because it fluctuates constantly is disruptive, and missing data for critical groupings of factors is common Lack of reflexes for sucking and swallowing, leading to feeding difficulties

In real world, analysis and make a decision regarding an object (e.g: person, machine etc.) doesn't depend on single domain (e.g.: social networking, geographic, real time media etc.) or single source.

To provide better investigation and quality inference, data from difference source need to be combined (fused).**Data Integration** is used to **integrate data** from different source to enhance the objective of information but it won’t fit for large data set. **Data fusion** is data analysis technique that fuses many individual type of data (large data set) with representing same object. Working together of multiple data with respect to an object produce an effect greater than sum of their individual effect. The key challenge of data fusion is to retrieve and fuse data across different domain is difficult. The study is to analyse the preterm birth using cross domain data fusion in big data predictive analytics and classifier algorithms.

BACKGROUND

Earlier research on preterm birth primarily focused on preterm birth risk factors, cervical length, and biochemical assessment. Common risk factors include the age of the pregnancy women, history of preterm labor, many pregnancies, diabetes, asthma, hypertension, thyroid disease, anemia, infection, Obesity, genetic influences, nutritional deficiencies, smoking, alcohol consumption, stress, excessive physical work recreational drugs, cervical length, etc., Unfortunately, their ability to predict is severely limited . As a result, many number of researchers have tried to predict premature birth using a machine learning approach on a collection of known clinical characteristics. These factors points to the inefficiency of previous methods for predicting the risk of labor in pregnant women . Women who seem to have a preterm birth, on the other hand, frequently have no known risk factors . A model that combines the mother's clinical data, the behavioral, psychological, environmental circumstances data history predicts spontaneous preterm labor better than the only medial factor.

In medical fields, human and artificial intelligence (AI) decision-making results in high-performance outcomes. AI is the art of emerging methods to solve problems typically related to human intelligence. In the ground of computer science, Machine learning (ML) is an artificial intelligence technique.ML focuses on using a number of algorithms and data to imitate the human way of learning, thereby increasing accuracy. ML includes different learning types of techniques: unsupervised, supervised, Evolutionary Learning, reinforcement learning, Semi-Supervised and Deep Learning. **Figure- 1** shows the types of the machine learning algorithm. In the initial stage of AI in medicine, they were standalone structures. There is no direct connection between AI and electronic data sets. Clinical data provide new learning health systems that open new opportunities and challenges. Beyond standard models, specific risk prediction has recently been improved using machine learning (ML) technologies. Many machine learning algorithms can represent intricate non-linear interactions between predictor characteristics and results. ML approaches can start understanding the framework from information without being specifically designed. A large amount of data is necessary to develop strong models with high accuracy using the ML technique . ML takes advantage of various variables from electronic health record (EHR) data for PTB prediction . look at studies that use machine learning algorithms to predict preterm birth, which could be useful in perinatal medicine. Fortunately, most countries' health records health records include information on a person's socio demographic, medical history, and obstetric. As a result, Health Records are good data sets for machine learning representations to study from and finally predict the desired outcome. Studies into using machine learning on HR data to find effective predictive frameworks for the timely identification of PTB have increased. This systematic review will look at the literature on using machine learning to identify PTB risk in mothers using HR data. Electronic health records (EHR) information, uterine electromyography (EMG) information, and electro hysterography (EHG) have been used in the majority of investigations to date. In recent years, attempts are made to integrate the characteristics of the mother and her medical history were collected to predict the risk exposures of PTB. Furthermore, many predictive systems based on maternal socio demographic factors have been investigated in machine learning algorithms.

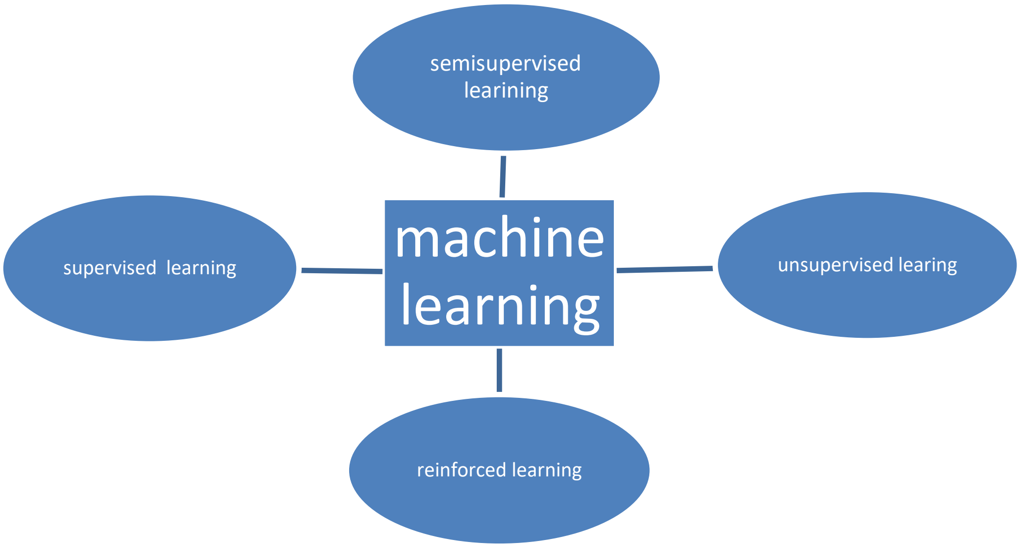

Figure1.different machine learning types

In this study, cross domain data fusion and its applications are discussed comprehensively. A novel framework for prognostic prediction model is proposed for health care domain.To analyse the medical data will give only the partial insight about an object. Design an efficient diagnostic and prognostic prediction model **figure-2**, along with patient medical details fuse data on this,

• Habits
• Living environment
• Friends
• Movie going behavior
• Social media and
• GPS Trajectories etc.

To control preterm birth using clinical data analytics is only gives the partial insight about an object. In addition to that mother's clinical data, behaviors, habits, and other environmental circumstances data also analyzed to design a prognostic model to prevent and predict preterm birth. There is a need to develop a prognostic model to indicate the preterm birth and also design a prediction model. To develop a model, data from different domain are fused. The various medical factors are studied and listed for occurrence of preterm birth,

• Malnutrition
• Overweight
• Thyroid deficiency
• Infections .etc.

Other than this medical factor, various behavioral and psychological factors are playing major role in preterm birth . To decrease the risk and develop the prediction model, the proposed frame works fuses clinical, behavioral and psychological factors. Analyze the behavior and psychology of an object, various data from different domain are derived as an individual data 4sets. From that datasets, required knowledge is extracted and fused using data fusion algorithm.

Elements used in proposed framework:

Objects: In this approach, two types of objects are used for analytics,

• Preterm baby

• Mother

Domain: The four vital domains are choose to extract data set,

[1] Health Care Domain

Clinical Dataset: Preterm baby and mother clinical data i.e. (blood pressure, diabetics, prenatal genetic test, ultrasound etc.)

1. Social Media Domain

Social media dataset: Friends details, likes, posts, and shared content .etc.

2. Movie Domain

Movie Dataset: Movie genre details i.e. (horror, comedy, and drama)

3. Transport Domain

GPS Dataset: Location data, Travel path data Prognostic Prediction

By extracting knowledge from each domain dataset and perform knowledge fusion to find, The common pattern among the objects that leads to preterm pregnancy.

Example: Movie watching habits, most preterm pregnancy woman are watched horror movies than comedy or drama

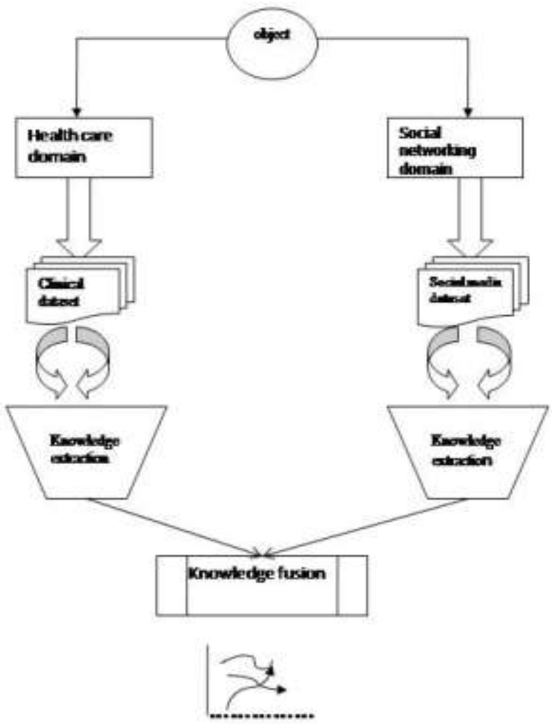

Figure 2. Framework of Prognostic prediction model of preterm

RELATED WORKS

This section focuses mainly on the existing methodologies related to prediction of PTB using machine learning, statistical analysis, and data mining techniques. Some of them are discussed in this section.

The study of Mercer et al was designed to develop a risk-score-based model for predicting PTB. The model can be trained using a multivariate logistic regression technique to explore various risk factors using clinical data available between 23 and 24 weeks' gestation.

A machine learning-based risk prediction conceptual model (RPCM) for PTB was introduced with the help of novel feature selection approach using entropy-notion to predict the birth cases (TB and PTB) from the obstetrical records.

Goodwin et al. employed the machine learning model to generate 520 predictive rules for PTB with the application of data mining techniques. The study in the deep learning models for predicting preterm delivery using existing electronic medical records (EMRs) of mothers available in healthcare centre

Weber et al. performed a cohort study to predict spontaneous preterm. The prediction of PTB was performed using numerous classifiers, namely, K-nearest neighbours, lasso regression, and random forests. This study has taken into the consideration of demographic, race-ethnicity, and maternal characteristics.

Mailath-Pokorny et al. explored the predictive features for preterm delivery that occurs within 2 days after admission and before 224 days of gestation using the multivariate logistic regression model. The predictive features considered are age of the mother, gestational age during admission, maternal history, vaginal bleeding, cervical length, preterm history, and preterm premature rupture of membranes (PPROM) in their study.

Son and Miller presented a prediction model for PTB using cervical length measurement in women with a singleton gestation. To accomplish better predictive performance, they attempted to determine the best cut points of cervical length.

Elaveyini et explored the major risk factors of preterm birth using artificial neural networks. PTB prediction was based on the feed-forward back propagation algorithm. Over the past decades, majority of research studies have been done to enhance the accuracy of prediction of PTB .

Researchers are continually making their best efforts to analyse and explore the principal risk factors for preterm delivery . The present article focuses on the cross domain data fusion to predict the PTB.

Shortcomings in the Existing Clinical Models: In recent years, a significant number of clinical prediction model have been developed to improve the accuracy of learning models. However, to the best of authors' knowledge, most of them suffer from selecting the most accurate features from the medical dataset in linear time. Hence, there is a scope for improving the performance of machine learning classifiers and reducing learning time in cross domain data fusion.

CONCLUSION

In this study, the proposed model (Preterm birth prediction using cross domain data fusion) can be used for prediction of PTB based on excellent features (mother's clinical data, the behavioral, psychological, environmental circumstances data) is integrated to analyze and design a prognostic model to prevent and predict preterm birth. The work focuses on cross domain data fusion approach by applying machine learning classifier algorithms in order to classify all birth cases into term birth and PTB. Comparing the performances of the classifiers on cross domain data. The model supports the decision-making process in maternity care by identifying and alerting the pregnant women at risk of preterm delivery thereby preventing possible complications, reducing the diagnosis cost, and ultimately minimizing the risk of PTB.

The limitation of the present research is that the risk factors for PTB are limited in size and dataset is small, which could be increased to improve the performance of the PTB prediction in the future studies. However, expert knowledge and clinical judgement may still be needed to interpret this risk and take appropriate action in individual cases.

REFERENCES

[1] Kannimuthu Subramanian , Premalatha Kandhasamy "UP-GNIV: an expeditious high utility pattern mining algorithm for item sets with negative utility value", International Journal of Information Technology and Management, Vol. 14, No. 1, pp. 26- 42, 2015.

[2] L. J. E. Meertens, P. van Montfort, H. C. J. Scheepers et al., "Prediction models for the risk of spontaneous preterm birth based on maternal characteristics: a systematic review and independent external validation," *Acta obstetricia et gynecologica Scandinavica*, vol. 97, no. 8, pp. 907–920, 2018.

[3] R. Pari, M. Sandhya, and S. Sankar, "Risk factors based classification for accurate prediction of the Preterm Birth," in *Proceedings of the 2017 International Conference on Inventive Computing and Informatics (ICICI)*, pp. 394–399, IEEE, Coimbatore, India, November 2017.

[4] A. Belle, R. (iagarajan, S. M. R. Soroushmehr, F. Navidi, D. A. Beard, and K. Najarian, "Big data analytics in healthcare," *BioMed Research International*, vol. 2015, Article ID 370194, 16 pages, 2015.

[5] R. Raja, I. Mukherjee, and B. K. Sarkar, "A systematic review of healthcare big data," *Scientific Programming*, vol. 2020, Article ID 5471849, 15 pages, 2020.

[6] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, "Disease prediction by machine learning over big data from healthcare communities," *IEEE Access*, vol. 5, pp. 8869–8879, 2017.

[7] Y. Li, T. Li, and H. Liu, "Recent advances in feature selection and its applications," *Knowledge and Information Systems*, vol. 53, no. 3, pp. 551–577, 2017.

[8] H. Huan Liu and L. Lei Yu, "Toward integrating feature selection algorithms for classification and clustering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 4, pp. 491–502, 2005.

[9] C. Gao, S. Osmundson, D. R. Velez Edwards, G. P. Jackson, B. A. Malin, and Y. Chen, "Deep learning predicts extreme preterm birth from electronic health records," *Journal of*

Biomedical Informatics, vol. 100, p. 103334, 2019.

[10] A. Weber, G. L. Darmstadt, S. Gruber et al., “Application of machine-learning to predict early spontaneous preterm birth among nulliparous non-Hispanic black and white women,”
Annals of Epidemiology, vol. 28, no. 11, pp. 783–789, 2018.

[11] M. Mailath-Pokorny, S. Polterauer, M. Kohl et al., “Individualized assessment of preterm birth risk using two modified prediction models,” *European Journal of Obstetrics &
Gynecology and Reproductive Biology*, vol. 186, pp. 42–48, 2015.

[12] M. Son and E. S. Miller, “Predicting preterm birth: cervical length and fetal fibronectin,” *Seminars in Perinatology*, vol. 41, no. 8, pp. 445–451, 2017.

[13] U. Elaveyini, S. P. Devi, and K. S. Rao, “Neural networks prediction of preterm delivery with first trimester bleeding,” *Archives of Gynecology and Obstetrics*, vol. 283, no. 5,
pp. 971–979, 2011.

[14] T. Khatibi, N. Kheyrikoochaksarayee, and M. M. Sepehri, “Analysis of big data for prediction of provider-initiated preterm birth and spontaneous premature deliveries and
ranking the predictive features,” *Archives of Gynecology and Obstetrics*, vol. 300, no. 6, pp. 1565–1582, 2019.

[15] M. Colstrup, E. R. Mathiesen, P. Damm, D. M. Jensen, and L. Ringholm, “Pregnancy in women with type 1 diabetes: have the goals of St. Vincent declaration been met concerning
foetal and neonatal complications?” *8e Journal of Maternal- Fetal & Neonatal Medicine*, vol. 26, no. 17, pp. 1682–1686,
2013.