

Original Investigation

Using an innovative stacked ensemble algorithm for accurate prediction of the preterm birth

Ramalingam et al. Using an innovative stacked ensemble algorithm for accurate prediction of the preterm birth

Pari Ramalingam, M. Sandhya, Sharmila Sankar

Department of Computer Science and Engineering, B. S. Abdur Rahman Crescent Institute of Science and Technology, Chennai, India

Address for Correspondence: Pari Ramalingam
e.mail: pari_ramalingam@yahoo.com

DOI: 10.4274/jtgg.2018.0105

Received: 1 August, 2018

Accepted: 30 November, 2018

Abstract

Objective: A birth before the normal term of 38 weeks of gestation is called a preterm birth (PTB). It is one of the major reasons for neonatal death. The objective of this article is to predict the PTB well in advance so that it is converted to a term birth.

Material and Methods: This study uses the historical data of expectant mothers and an innovative stacked ensemble (SE) algorithm to predict the PTB. The proposed algorithm stacks the classifiers in multiple tiers. The accuracy of the classification is improved in every tier.

Results: The experimental results from this study show that PTB can be predicted with more than 96% of accuracy using the innovative stacked ensemble learning.

Conclusion: The proposed approach helps the physicians in Gynecology and Obstetrics departments to decide whether the expectant mother needs a treatment or not. Only the patients for whom a PTB is predicted, the treatment is given to delay the birth or in many cases to convert the PTB to a normal birth. This, in turn, can reduce the mortality of babies due to PTB.

Keywords: Preterm birth, neonatal death, risk factors of PTB, stacked ensemble, stacked generalization, meta-learning

I. INTRODUCTION

Births which happen after 37 weeks of gestation and before 39 weeks are termed as normal birth (TB). The babies born before 37 weeks of gestation are considered as premature babies and such births are termed as preterm birth (PTB) [1], [2]. Premature babies typically may have many severe complications like breathing/respiratory problems (apnea), chronic lung disease, jaundice, anemia, infections, bleeding in brain (intraventricular hemorrhage), etc. In the worst cases, the premature babies die in the early days of life itself. Such deaths are termed as neonatal death [3]. The UNICEF report published in 2015 states that PTB is the major cause of the neonatal death [4]–[6]. Due to PTB, some of the maternal also suffer from poor mental health and in some extreme cases they suffer from mental disorder. [7]. The long term consequences of PTB for the babies are cognitive problems (intellectual disability and learning disability), asthma, intestinal problems, vision problems, hearing loss problems, dental problems, poor growth and increased risk of sudden infant death syndrome.

When the expectant mother undergoes for prenatal checkups, the clinical pathological may indicate the possibility of a PTB. In Obstetrics and Gynecology (O&G) world, these indicators are called as risk factors of PTB [8]–[11]. The Obstetrician Gynecologist analyzes these risk factors and diagnoses the birth as either TB or PTB. While diagnosing PTB, the Obstetrician Gynecologist also take into consideration the behavioral and social characteristics of the expectant mother [12]. Hence they are also considered as risk factors of PTB. All the risk factors are not critical in nature and they do not equally contribute to PTB. Hence the risk factors are categorized into primary risk factors and secondary risk factors based on their criticality. The primary and secondary risk factors associated with PTB are listed in table I.

TABLE I. RISK FACTORS ASSOCIATED WITH PTB [13], [14]

Primary Risk Factors	Premature rupture of membranes
	Presence of fetal fibronectin in vaginal discharge
	Cervix shortens earlier than third trimester
	Excessive amount of amniotic fluid
	Conceived with assisted reproductive technologies
	Prior history of PTB
	Prior history of abortion
	Multiple gestation
	Vaginal or urinary tract infections
Short inter-pregnancy interval	
Secondary Risk Factors	Under weight (< 45 kg.)
	Short stature (height < 145 cm)
	Stress/hypertension
	Heavy work
	Family history
	Prior history of pregnancy loss
	Lack of antenatal checkup
	Low economic status
	Smoking
Diabetes	

Getting the evidence for PTB in clinical pathology is a challenging task. More than that, some of the clinical tests are expensive to afford for patients from developing countries. Hence predictive analytics is the way to go. Predicting a PTB as a TB leads to fatal consequences, thus the learning algorithms with high accuracy are the need of the hour. Ensemble learning gives better accuracy than the individual learning algorithms and hence is suitable for predicting PTB [15], [16]. Especially if the base learners are diverse and are moderately performing, the ensembles perform effectively [17]. Using a trainable combiner to learn from the predictions of base learners, generalizes better than the traditional ensembles [18]. Such learning systems are termed as Stacked Ensemble (SE) systems [19], [20]. They use base classifiers to train level-0 models and a generalizer to learn from the predictions of level-0 models. Hence the predictions of the base classifiers form the input space for the generalizer. These predictions are termed as meta-features and the generalizer is said to perform meta-learning [21]–[23].

This study uses an innovative SE algorithm for accurate prediction of PTB. It differs from the traditional SE in producing the meta-features. Rather than using the predictions of level-0 models as meta-features, it combines them using multiple combination schemes to produce the meta-features. The meta-features along with the critical features are used to train the generalizer. The combination schemes produce the joint distributions of the level-0 predictions. The predictions from level-0 models are the abstraction of the mapping between the input space and the actual labels. Hence these joint distributions not only map the level-0 predictions to the actual label, but also map indirectly the input space with the actual label. This in turn produces meta-features

which abstracts in a better way, the relationship between the input space and the actual labels. Due to this the proposed algorithm performs better than the traditional SE algorithm. The performance of the algorithm is measured using its accuracy and recall. The following are the contributions of this study: (i) introduce an innovative SE algorithm to improve the prediction accuracy (ii) Use this algorithm to predict the PTB accurately and (iii) motivate the research community to use this algorithm for the classification problems. Organization of the remaining sections: Section II describes the work carried out in predicting the PTB. Section III depicts the proposed algorithm in detail. Experimental results along with the inferences are reported in section IV. Section V is the conclusion and the scope for future work.

II. RELATED WORK

Using machine learning or statistical analysis for predicting the PTB based on historical data is gaining momentum in Obstetrics and Gynecology. Bittar et al. [24] used statistical analysis to predict the spontaneous PTB for high risk group of expectant mothers who had prior PTB. They used the cervical length and the level of protein-1 (pHGF β -1) in cervical secretions as the input for their statistical analysis. In the first step, they used pHGF β -1 and cervical length to perform the logistic regression analysis to predict PTB. For examining the contributions of these two factors on PTB, they performed multiple logistic regression analysis. A dataset with 105 expectant singleton mothers were used for their analysis. They found that pHGF β -1 level measured during 30th week of gestation helps in predicting the PTB accurately. In combination with this measuring the cervical length between 22nd and 24th weeks and using it for the statistical analysis improves the prediction rate of PTB. They achieved a prediction rate of 92% before the 34th week and a prediction rate of 80% before the 37th week. Though their study resulted in high prediction rate, the number of PTB instances in the dataset was only 12. Hence generalizing their result involves a risk element.

A similar kind of study was carried out by Care et al. [25] to predict PTB in women with prior PTB and normal cervical length (greater than 25mm) during 22 – 24 weeks. They used a dataset with 196 instances out of which 134 patients had a normal cervical length and 62 patients had shorter cervical length. Out of the 134 patients with normal cervical length, 28 patients had a PTB and 12 of these had a prior PTB. Out of the 62 patients with shorter cervical length, 25 patients had a PTB. All these patients are White British population demographic. They used SPSS for carrying out the analysis. The analysis revealed that the normal cervix or long cervix did not provide any assurance on recurrence of PTB. They concluded that that the normal cervical length and the demographic information of the patients are not good features to predict the PTB. They also suggested using other factors like amniotic fluid, vaginal discharge, genetic and social & environment factors to predict the PTB.

Predicting PTB outside of the clinical pathology is a better approach for early prediction of PTB. Christina Catley et al. [26] used back propagation feed forward Artificial Neural Network (ANN) for predicting the PTB. They used PPPESO databases for conducting the experiments. The dataset was skewed towards the TB and hence they removed the TB instances to balance the dataset. While dividing the dataset into a training set and a test set, they ensured that the distribution of TB and PTB were not skewed. They used logarithmic sensitivity index for measuring the performance of the prediction. They used MatLab with Neural Network Toolkit for conducting the experiment. They used weight-elimination cost function to improve the classification performance. They used one hidden layer with three hidden nodes. When the skewed dataset was used, the sensitivity reached the peak at 20.4%. The sensitivity of prediction reached the peak at 33.4% with a more balanced dataset. They observed that the number of fetus in the womb in the current pregnancy and previous pregnancies, number of children, smoking after 20 weeks of gestation are the factors with highest connection weights in the ANN. Hence these were major contributors for PTB.

Analyzing the EHG signals for predicting the PTB is another popular approach. This approach uses the signals generated by the contractions and expansions of the uterus. Peng Ren et al. [27] used the EHG signal to classify the birth as TB or PTB. They used the EHG signals of 300 patients available in PhysioBank. Though the signal had wide range of frequency spectrum, they filtered the frequency range of 0.3 Hz to 3 Hz. The signal was decomposed into intrinsic mode functions using empirical mode decomposition. The first ten functions were selected for prediction. They used Gabriel Rilling EMD toolbox for this purpose. The dataset had 262 TB patients and 38 PTB patients. The dataset was balanced using SMOTE. They used principal component analysis to select the components so that it could improve the Area Under the Curve (AUC) value. They used multiple classifiers to classify the PTB. On an average they achieved a maximum AUC value of 86.2%. They analyzed the impact of using the features from 3 channels of EHG against using only the features from channel 3. When the features from only channel 3 were used, the AUC reached a maximum of 89%. Among the classifiers, AdaBoost reached the highest accuracy of 98.6%. The limitation of this approach is the availability of EHG signals of the patients in developing countries.

There are three types of features which can be extracted from the uterine EHG signals. They are linear analysis; non linear analysis and Discrete Cosine Transform analysis. Safaa M. Naeem et al. [28] extracted all these three types of features and analyzed them using ANN. They used three different forms of ANNs like Cascade-forward back propagation network, feed-forward network and Kohonen network. They trained the networks with linear features, non-linear features, DCT features and the combination of these features. They used PCA to select 10 components and used them for prediction. They compared the results obtained for all the types of features with three different forms of ANNs. They concluded that the accuracy reaches a peak value of 90% when feed-forward network with the linear and DCT features of uterine EHG signals were used. After analyzing the consequences of false positives, the accuracy of 90% is not enough for PTB problem.

Application of SE has found mention in many studies. Shuang-Quan Wang et al. [29] applied SE on the dataset constructed by Chou and Elrod [30] to predict the membrane protein types based on pseudo-amino acid composition. Support Vector Machine

(SVM) and Instance-Based Learning (IBL) were used as level-0 classifiers. The combination of these two classifiers provided more information about the input space and its relationship with the class label. To achieve the faster training, Sequential Minimal Optimization (SMO) algorithm was used to train SVM. These classifiers were cross validated using k-fold cross validation. The cross validated predictions from level-0 SVMs were used as input level-1 generalization. Decision Tree (DT) was used as level-1 generalizer. Re-substitution test, jackknife test and the independent dataset test were used to examine the quality of prediction through bias and variance estimation. Among the three tests, high success rate of 85.4% was achieved in jackknife test. SE has achieved remarkably good performance and thus helped in providing the direction for functionally characterizing the gene products using gene sequences.

III. MATERIAL AND METHODS

When the traditional SE is used, the level-1 generalizer inherits some level of bias and variance from level-0 models. Hence the problem of overfitting or underfitting is not eliminated to the maximum possible extent. To address this issue, this study proposes an innovative SE algorithm by stacking classifiers in multiple tiers [31]. The proposed algorithm stacks the classifiers in three tiers namely (i) base tier, (ii) ensemble tier and (iii) generalization tier [32]. The base tier focuses on training a set of suitable learners to achieve moderate accuracy. The second tier uses a set of combinations schemes to combine the predictions from the base learners. The outputs from the combination schemes form the input space for the next tier. The third tier does the meta-learning using the newly formed input space. The performance of this algorithm is optimized using a suitable number of base learners and a suitable number of combination schemes. The choice of meta-learner in the third tier also plays a vital role in improving the accuracy of this algorithm. The base tier ensures reducing the bias, the ensemble tier and the generalization tier ensures reducing the variance. As a result, the bias is perfectly balanced with the variance. Due to this the proposed algorithm improves the classification accuracy. Hence this algorithm is suitable for classifying the PTB based on the historical data of expectant mothers.

In general, for any classifier, cross validation helps in reducing the bias [33], [34]. In the proposed algorithm also 10-fold cross validation is used to train the base learners. The dataset is partitioned into 10 disjoint sets. Each of these 10 sets is used one after the other as a test set. Each fold is used for nine times as a training set. As a result of this, the base learners produce the cross validated predictions. Figure 1 depicts the cross validation of the base learners. As this process is repeated for each of the base learners, the output of the base tier is multiple sets of cross validated predictions. This serves as the input for the next tier.

As depicted in figure 2, the set of cross validated predictions from each of the base learners are used as input in the second tier. The goal of the second tier is to combine the predictions from the base tier and to map them with one of the class labels. Hence it creates a joint distribution of base learners' predictions. The output from each of the combination schemes provides a meta-feature. The quality of meta-features depends on the choice of the combination schemes used in this tier. Better the combination schemes, better the meta-features capture the inherent relationship between the input space and the actual labels. Hence the meta-features plays big role on the accuracy of this innovative stacked ensemble algorithm. The combination schemes to be used in this tier are decided depending upon the problem on hand. The popular combination schemes like Averaging, Majority Voting, etc. works well for most of the problems.

As depicted in figure 3, the meta-features and the top three critical features selected from the original input space form the input space for the meta-learner. The top three features are selected by analyzing the correlation of each of the features with the class labels. The features which have high correlation with the class label are selected.

After analyzing the historical data of patients, we decided to use the following list of base learners, the combination schemes and the meta-learner in this study. This is shown in table II.

TABLE II. CLASSIFIERS USED IN DIFFERENT TIERS

Base Learners
Support Vector Machines (SVM)
Decision Trees (DT)
Logistic Regression (LR)
K-Nearest Neighbors (kNN)
Gaussian Naïve Bayes (GNB)
Stochastic Gradient Descent (SGD)
Passive Aggressive Classifier (PAC)
Perceptron

Combination Schemes	Aggregation (Averaging and rounding off) Majority Voting Weighted majority Voting Confidence Based Voting
Generalizer	Decision Trees (DT)

The experiment was implemented using Python and scikit-learn library [35]. A dataset consisting of the historical data of 2600 patients was used to carry out this study. The dataset was a masked dataset without any reference to the personal details of the patients. Hence the need to get the informed consent and the ethical committee approval did not arise. The details about the data set are given in table III. The data was thoroughly reviewed to check if the dataset had a good mix of all the possible cases: (i) Maternal with risk factors and had a PTB (ii) Maternal without risk factors and had a PTB (iii) Maternal with risk factors but had a TB and (iv) Maternal without risk factors and had a TB. This mix of all the possible cases was also ensured in the training and testing data.

TABLE III. DESCRIPTION OF THE DATASET

# Instances	2600
# Attributes	26
# Classes	2
# Major Instances	1936
# Minor Instances	664
Imbalance Ratio (IR)	2.92

The distribution analysis of the class labels in the dataset reveals that the dataset is an asymmetric one and is skewed towards TB. PTB is the minority class and TB is the majority class. Hence SMOTE (Synthetic Minority Over-sampling Technique) algorithm is used to balance the dataset. Balancing the dataset increases the number of minor instances to match with the number of major instances. This in turn increases the total number of instances. The count of TB and PTB in the dataset before and after SMOTE is shown in figure 4.

As the missing data in the dataset plays a big role in pulling down the accuracy of the models, the dataset was thoroughly analyzed for missing data or null values. If the missing data or null value was found in any of the features and the criticality of feature in which it is found was analyzed. For all the critical features, the mean values were used to replace the missing data. For all the non-critical features, the default values were used. A scatter plot of the historical data was plotted to reveal the outliers. As most of the features are binary in nature, top 5 critical features are selected and concatenated. The concatenated feature is taken along the x-axis and the class label is taken along the y-axis. The central mass of the plot was identified and the points which are farther away from this central mass were analyzed to identify the outliers. The identified outliers were removed from the dataset. Normalizing the dataset also helps in improving the performance of the learning algorithm. Hence different normalization methods were analyzed to select a suitable one for the problem on hand. In this study, the dataset was normalized by carrying out mean cancellation. To assess the impact of primary and secondary factors on the classification accuracy, multiple experiments were conducted with different subset of features in the dataset. The list of experiments conducted is depicted in table IV. These experiments help to understand the contributions of primary and secondary risk factors to PTB. Each of these experiments was repeated for 10 times to ensure the consistency of the results. The average value of the accuracy, the precision, the recall and the F-1 score across the trials are reported in this study. Real Operating Characteristic (ROC) curves were also drawn for each of these experiments and are also reported in this study.

TABLE IV. LIST OF EXPERIMENTS

S. No.	Experiment	Description
--------	------------	-------------

1	Top 5 Secondary	Top 5 features from the secondary risk factors
2	Only Secondary	All the features from the secondary risk factors
3	Top 5 Primary	Top 5 features from the primary risk factors
4	Only Primary	All the features from the primary risk factors
5	All Factors	All the risk factors

IV. RESULTS AND DISCUSSION

A comparative analysis of different performance metrics for SE and the proposed algorithm was carried out. In addition to this, the analysis of how the features subsets improve or degrade the performance metrics was also carried out. This analysis helps in understanding the factors which make major contribution to PTB. The results reveal that the performance metrics reached the maximum when all the features in the dataset were used for training the algorithms. Irrespective of the number of risk factors used for training, the performance of the proposed algorithm is better than the performance of SE. The results of the experiments in which all the risk factors are used for training is summarized in table V.

TABLE V. SUMMARY OF THE RESULTS

Performance Metrics	SE	Proposed Algorithm	Performance Improvement
Accuracy	93.80%	96.90%	3.10%
Precision	94.83%	98.56%	3.73%
Sensitivity/Recall	93.16%	94.44%	1.28%
F1 Score	92.95%	95.01%	2.06%
AUC	92.67%	94.92%	2.25%

The analysis of accuracy for SE and the proposed algorithm is depicted in figure 5. Among the five experiments conducted, the accuracy was at minimum for the experiment conducted only with top five secondary risk factors. There was a big jump in accuracy when other secondary risk factors were also used for training the algorithms. The improvement in accuracy of the proposed algorithm reached the maximum of 12% when only the secondary risk factors were used for training. This implies that, when only trivial features are available, the proposed algorithm still can perform much better than SE. This is mainly due to the reason that the proposed algorithm does not suffer much from overfitting or underfitting. When all the factors are used for training, there is an improvement of more than 3% in accuracy over SE. When only the primary risk factors were used for training, the accuracy of the proposed algorithm is just 1.3% below the accuracy of the proposed algorithm when all the risk factors were used. Hence the contribution of the secondary risk factors in PTB is not significant. Even the maximum accuracy of 93.8% achieved by SE when all the factors are used for training, is 1.8% less than the accuracy achieved by the proposed algorithm with only primary factors. Hence for high dimension datasets also, the proposed algorithm can use minimal features and achieve better accuracy than SE.

The analysis of precision for SE and the proposed algorithm is depicted in figure 6. The observed values of precision are also in similar lines of accuracy. The precision of the proposed algorithm reached the peak value of 98.56% when all the risk factors were used for training. High value of precision for the proposed algorithm implies that the number of false positives were less. High precision implies that most of the TB cases were predicted as NBs. This avoids the unnecessary treatment given to the expectant mothers who otherwise would have undergone the treatment. The improvement in precision reached a maximum of 10% when only secondary risk factors were used for training. Even with trivial factors, the proposed algorithm performs better than SE.

The analysis of sensitivity for SE and the proposed algorithm is depicted in figure 7. High value of sensitivity for the proposed algorithm implies that the number of false negatives were less. High sensitivity implies that most of the PTB cases were predicted as PTBs. This indicates that the patients who need immediate medication do not suffer from the predictions by the proposed algorithm. The improvement in sensitivity reached the maximum of 8.5% when only primary risk factors were used for training. When the top five secondary risk factors were used for training, there is no improvement in sensitivity.

The analysis of F1 score for SE and the proposed algorithm is depicted in figure 8. F1 score is the harmonic mean of precision and sensitivity. As the proposed algorithm achieves improvement in both precision and sensitivity, its F1 score is also better than that of

SE for all the five experiments. F1 score reaches the minimum when only top 5 secondary risk factors were used for training the algorithms. The difference in the F1 score of SE and the proposed algorithm is as high as 13% when only the secondary risk factors were used for training. F1 score reaches the maximum when all the risk factors are used for training.

ROC curves were drawn to analyze the Area under the Curve (AUC). The ROC for SE and the proposed algorithm for the five experiments are depicted in figure 9. The set of graphs in the first row corresponds to SE and the set of graphs in the second row corresponds to the proposed algorithm. In the below graphs, the middle way mark of 50% is represented as dotted lines. As expected, AUC reached the minimum when only top five secondary risk factors were used for training the algorithms. AUC increases with the number of critical factors used for training. More the number of critical factors used for training, more is the AUC. It reached the peak for both SE and the proposed algorithm when all the risk factors were used for training. The minimum values of AUC for the top 5 secondary risk factors imply that the true positive rate is not reaching the peak even if the false positive rate reaches the minimum. This means that the false negatives are high in the prediction. From PTB perspective, this is alarming. High value of false negative implies that a patient who needs the immediate attention and treatment may not receive the treatment.

V. CONCLUSION

The application of the innovative SE algorithm for predicting PTB has achieved better performance than SE for all the experiments conducted in this study. For all the performance metrics considered in this study, the innovative SE algorithm is way ahead when compared with the traditional SE algorithm. The primary risk factors play a major role in predicting the PTB. When the secondary factors are also used along with the primary risk factors, the performance metrics improved marginally (little more than 1%). Hence using only the primary risk factors with the proposed algorithm is the efficient method for PTB prediction. The time complexity of the proposed algorithm with different set of factors can be considered for the future work. As the proposed algorithm is scalable in terms of the number of base learners and the number of combination schemes, the accuracy can further be improved by using a large number of base learners and the combination schemes. Finding the optimal number of base learners and the combination schemes is also an interesting area to explore further. In order to increase the clinical usage of this algorithm, we are considering the possibility of designing a mobile app with a wrapper around the algorithm. The mobile app allows the obstetrician gynecologist to enter the results of the clinical tests of the expectant mother using an interface and get the corresponding prediction. This mobile app hides the complexities of the statistical methods from the end user and thus more and more obstetrician gynecologists get benefited from this algorithm. We are also exploring if this algorithm can be enhanced and extended to analyze other maternal complications.

REFERENCES

- [1] G. C. Di Renzo, L. C. Roura, and null null, "Guidelines for the management of spontaneous preterm labor," *J. Perinat. Med.*, vol. 34, no. 5, pp. 359–366, Jan. 2006.
- [2] World Health Organization, "Preterm Birth," *November 2015*. 2015.
- [3] S. Rahman, "Neonatal Mortality: Incidence, Correlates and Improvement Strategies," in *Perinatal Mortality*, O. Ezechi, Ed. Rijeka: InTech, 2012.
- [4] R. W. Loftin, M. Habli, C. C. Snyder, C. M. Cormier, D. F. Lewis, and E. A. Defranco, "Late preterm birth," *Rev. Obstet. Gynecol.*, vol. 3, no. 1, pp. 10–9, 2010.
- [5] UNICEF, "Levels and trends in child mortality," 2015.
- [6] M. F MacDorman, T. J. Matthews, A. Hindori-Mohangoo, and J. Zeitlin, "International Comparisons of Infant Mortality and Related Factors: United States and Europe, 2010.," *Nat Vital Stat Rep*, vol. 63, pp. 1–7, 2014.
- [7] A. R. Misund, P. Nerdrum, and T. H. Diseth, "Mental health in women experiencing preterm birth," *BMC Pregnancy Childbirth*, vol. 14, no. 1, p. 263, 2014.
- [8] R. J. Perry, V. T. Samuel, K. F. Petersen, G. I. Shulman, N. Haven, and N. Haven, "HHS Public Access," vol. 510, no. 7503, pp. 84–91, 2015.
- [9] G. Chistyakova, I. Gazieva, I. Remizova, L. Ustyantseva, V. Lyapunov, and S. Bychkova, "Risk factors vary early preterm birth and perinatal complications after assisted reproductive technology," *Gynecol. Endocrinol.*, vol. 32, no. sup2, pp. 56–61, 2016.
- [10] "Assessment of risk factors and predictors for spontaneous pre-term birth in a South Indian antenatal cohort," *Clin. Epidemiol. Glob. Heal.*, Jul. 2017.
- [11] F. Fuchs, B. Monet, T. Ducruet, N. Chaillet, and F. Audibert, "Effect of maternal age on the risk of preterm birth: A large cohort study," *PLoS One*, vol. 13, no. 1, pp. 1–10, 2018.
- [12] N. Morisaki *et al.*, "Risk factors for spontaneous and provider-initiated preterm delivery in high and low Human Development Index countries: a secondary analysis of the World Health Organization Multicountry Survey on Maternal and Newborn Health.," *BJOG*, vol. 121 Suppl, pp. 101–109, 2014.
- [13] G. C. Di Renzo, E. Pacella, L. Di Fabrizio, and I. Giardina, "Preterm Birth: Risk Factors, Identification and Management," in *Management and Therapy of Late Pregnancy Complications*, Cham: Springer International Publishing, 2017, pp. 81–94.
- [14] M. E. Ahumada-Barrios and G. F. Alvarado, "Risk Factors for premature birth in a hospital," *Rev. Lat. Am. Enfermagem*,

vol. 24, no. 0, 2016.

- [15] S. Wang and X. Yao, "Relationships between diversity of classification ensembles and single-class performance measures," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 1, pp. 206–219, 2013.
- [16] R. Polikar, "Ensemble based systems in decision making," *Circuits Syst. Mag. IEEE*, vol. 6, no. 3, pp. 21–45, 2006.
- [17] G. Brown, J. Wyatt, R. Harris, and X. Yao, "Diversity creation methods: A survey and categorisation," *Inf. Fusion*, vol. 6, no. 1, pp. 5–20, 2005.
- [18] E. Classifiers, "ERROR CORRELATION AND ERROR REDUCTION IN ENSEMBLE CLASSIFIERS * Kagan Tumer and Joydeep Ghosh," pp. 1–24, 1996.
- [19] D. H. Wolpert, "Stacked Generalization," *Neural Networks*, vol. 5, pp. 241–259, 1992.
- [20] D. W. Opitz and R. Maclin, "Popular Ensemble Methods: An Empirical Study," *J. Artif. Intell. Res.*, vol. 11, pp. 169–198, 1999.
- [21] L. Breiman, "Stacked regressions," *Mach. Learn.*, vol. 24, no. 1, pp. 49–64, 1996.
- [22] R. Vilalta and Y. Drissi, "A perspective view and survey of meta-learning," *Artif. Intell. Rev.*, vol. 18, no. 2, pp. 77–95, 2002.
- [23] L. Todorovski and S. Džeroski, "Combining multiple models with meta decision trees," *Zhurnal Eksp. i Teor. Fiz.*, 2000.
- [24] R. E. Bittar, E. B. Da Fonseca, M. H. B. De Carvalho, S. Martinelli, and M. Zugaib, "Predicting preterm delivery in asymptomatic patients with prior preterm delivery by measurement of cervical length and phosphorylated insulin-like growth factor-binding protein-1," *Ultrasound Obstet. Gynecol.*, vol. 29, no. 5, pp. 562–567, 2007.
- [25] A. G. Care, A. N. Sharp, S. Lane, D. Roberts, L. Watkins, and Z. Alfirovic, "Predicting preterm birth in women with previous preterm birth and cervical length ≥ 25 mm," *Ultrasound Obstet. Gynecol.*, vol. 43, no. 6, pp. 681–686, 2014.
- [26] C. Catley, M. Frize, C. R. Walker, and D. C. Petriu, "Predicting high-risk preterm birth using artificial neural networks.," *IEEE Trans. Inf. Technol. Biomed.*, vol. 10, no. 3, pp. 540–9, 2006.
- [27] P. Ren, S. Yao, J. Li, P. A. Valdes-Sosa, and K. M. Kendrick, "Improved prediction of preterm delivery using empirical mode decomposition analysis of uterine electromyography signals," *PLoS One*, vol. 10, no. 7, pp. 1–16, 2015.
- [28] S. M. Naem, A. F. Seddik, and M. A. Eldosoky, "New technique based on uterine electromyography nonlinearity for preterm delivery detection," *J. Eng. Technol. Res.*, vol. 6, no. 7, p. 107–114, November 2014, 2014.
- [29] S. Q. Wang, J. Yang, and K. C. Chou, "Using stacked generalization to predict membrane protein types based on pseudo-amino acid composition," *J. Theor. Biol.*, vol. 242, no. 4, pp. 941–946, 2006.
- [30] K. C. Chou and D. W. Elrod, "Prediction of membrane protein types and subcellular locations," *Proteins Struct. Funct. Genet.*, vol. 34, no. 1, pp. 137–153, 1999.
- [31] R. Pari, M. Sandhya, and S. Sankar, "A Multi-Tier Stacked Ensemble Algorithm to Reduce the Regret of Incremental Learning for Streaming Data," *IEEE Access*, vol. 6, pp. 48726–48739, 2018.
- [32] R. Pari, M. Sandhya, and S. Sankar, "A Multi-Tier Stacked Ensemble Algorithm for Improving Classification Accuracy," *Comput. Sci. Eng.*, p. 1, 2018.
- [33] R. Kohavi, "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection," *Appear. Int. Jt. Conf. Artificial Intell.*, vol. 5, pp. 1–7, 1995.
- [34] M. Stone, "Cross-validators choice and assessment of statistical predictions," *J. R. Stat. Soc. Ser. B*, pp. 111–147, 1974.
- [35] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.