SN Computer Science (2024) 5:632
https://doi.org/10.1007/s42979-024-02901-6

ORIGINAL RESEARCH

Methylation Data of Parents in the Prediction of a Preterm Birth: A Machine Learning Approach

Pratheeba Jeyananthan[1] · **G. L. D. S. Piyasamara**[1] · **D. C. Sachintha**[1]

Received: 26 May 2023 / Accepted: 12 April 2024

Abstract

Preterm birth is a serious issue which can affect the whole family, especially the mother both physically and mentally. Further, babies also need to face a lot of short-term and long-term complications, sometimes throughout their life. Annually we have approximately 15 million premature babies worldwide, which is the leading cause of death among children. However, early prediction of a preterm birth can help the clinician to give proper treatments to the mother in order to avoid this complication. Previous biological studies showed that there are epigenetic differences between a preterm baby and a full-term baby, and associations between prenatal risk factors and epigenetic changes. Anyhow it is comparably very hard to get epigenetic data from an infant before labor. Hence, this study analyses the methylation data of father and mother individually in the prediction of the possibility of premature birth using machine learning algorithms such as random forest, support vector machine and K-nearest neighbor. As we have high number of features in this data, mutual information is used to select the relevant features of this study. Different number of features are selected using mutual information and their performances are evaluated using all the three machine learning algorithms to reveal the best number of features. Results indicate that top 15 methylation features taken from the father along with random forest classifier outperform other models with a perfect accuracy (AUC = 1 ± 0.00) in the prediction of premature possibilities.

Keywords DNA methylation · Premature birth · Machine learning models · Feature selection · Father-mother–child

Introduction

A birth is called as preterm when the delivery occurs before 37 weeks of gestational age. The current rate of global preterm birth is approximately 11%. It accounts for 18% of all deaths among children under 5 years, and 35% of all deaths among newborns aged less than 28 days [1]. Preterm birth of a baby is not only affects the mental health of a mother but also associated with socio-demographic, obstetric, and neonatal risk factors [2]. As there are various strategies to prevent preterm birth, prediction of preterm births before the delivery will help the doctors to take necessary actions to avoid it [3].

Association between the molecular characterization of parents and preterm children was already studied in biological researches [4]. In some studies, epigenetic changes were studied among mother-infant pairs [5]. There are even more biological studies that showed the connection between epigenetic modification and prematurity [6]. Candidate genes and methylation changes related to preterm labor and very early preterm labor were identified in these type of previous studies [4]. Some other studies identified epigenetic DNA methylation regions correlated with prematurity by studying methylation data from father-mother-infant triads [7]. Further studies supported the hypothesis that epigenetic modification induced by pregnancy-related risk can influence the risk of preterm birth [8].

These biological studies show the connection between epigenetic data and the premature delivery. These days machine learning is widely used in medicine in various ways including automated diagnosis system, drug discovery and especially, omic data along with machine learing is moving

This article is part of the topical collection "Digital Healthcare and Wellbeing" guest edited by Achilleas Achilleos, George A. Papadopoulos, Edwige Pissaloux and Ramiro Velazquez.

✉ Pratheeba Jeyananthan
pratheeba@eng.jfn.ac.lk

[1] Department of Computer Engineering, Faculty of Engineering, University of Jaffna, Kilinochchi, Sri Lanka

Published online: 10 June 2024

towards the personalized medicine [9–13]. In this particular area also, in the literature, few studies used machine learning algorithms in the prediction of preterm birth. However, clinical data obtained from the mother such as personal details, demographic data, pregnancy history and maternal health data were mostly used in those studies [14–18]. Few other techniques including signal processing [19], pathway analysis [20] and uterine records [21] also used in the prediction of preterm birth.

Even though there are many omics-related biological studies in the literature, those studies or data were rarely used in the machine learning-related predictions. However, those biological studies showed a close relationship between premature deliveries and omics scale data. As the literature showed that genetic factors influence the delivery time of a baby [22, 23], here in this study we use the methylation data of the parents of a premature new-born to predict the possibilities of premature labor. An accurate prediction of preterm delivery will help a lot for the medical field and public for a better understanding of the problem and to start an ontime treatment for this crucial issue.

As machine learning models have the capability of predicting a target accurately and quickly, this study compares performance differences between various classification algorithms such as random forest, support vector machines and K-nearest neighbor in the prediction of the possibilities of preterm birth using the methylation data of father and mother individually to see which data has a high correlation with the preterm labor of a baby. Mutual information is used to select the relevant features of the study. Performace difference between distinct number of features also checked to select the best set of features with the highest accuracy. Accuracies between different models are compared and the result shows that top 15 methylation features of the father selected using mutual information along with random forest classifier can predict the premature delivery almost perfectly.

Related Work

These days machine learning is widely used in medicine for many purposes including diagnosis of the disease, biomarker identification, personalized medicine and drug discovery. It is also used in the studies related to women reproductive health and pregnancy, especially to predict the pregnancy complications. One such important complication is preterm birth, which needs more special attention.

There are studies in the literature predicting the preterm birth or possibilities of the preterm birth. Those studies mainly used clinical data, EHG or EHR data [24–30] from the mother. Those studies showed a performance ranges between 65 and 99%. However, the study gave 99% accuracy was obtained from EHG recording data [28], where this data needs more computational power and time to train the model compared to the data used in our study.

On top of this, omics data also used in the literature to predict the preterm birth. miRNA data [31] and metabolome & microbiome [32] data were used in these studies and unexpectedly their perfoamces are aroud 70–80%. These studies show that the previous studies rarely used data from father, where they focused mainly on mother. However, in [33], they suggested that there might be a relationship between the gestational age and paternal DNA methylation. However, this data is not utilized in machine learning related studies.

Here, in this study we are filling the gap in the literature by considering father's epigenetic data in the prediction of preterm birth using machine learning algorithms. Even though genetic factor is influencing the preterm birth, epigenome data is not used in this prediction so far. Hence, this study utilises epigenome data from father and mother in the prediction of preterm birth, compares the performance between them and reveal the role of paternal epigenetic data and preterm labor. Further, this study provides a simple machine learning model with less time complexity to solve this issue.

Materials and Methods

Material

This study uses a dataset from Gene Expression Omnibus (GEO), named preterm birth buccal cell epigenetic biomarkers to facilitate preventative medicine. This is a publicly available data with the accession number, GSE194227 [34]. More detail of the data and the study setup can be obtained from GEO data repository using the accession number.

Methods

Data is downloaded from the above repository and it is preprocessed. During the data preprocessing, null values and duplicate data are removed. Initially, before removing all these data samples, shape of data is [91, 3088298]. After removing all these values, our data shape is [40, 2932684]. We have 40 samples from father and 40 from mother. As the data range is big, in the next step they are normalized between 0 and 1. After that, mutual information is used to select the relevant features of the study. Selected features are used with different machine algorithms to predict the possibility of preterm birth, performance between different models are compared and the one with the best accuracy is selected (Fig. 1).

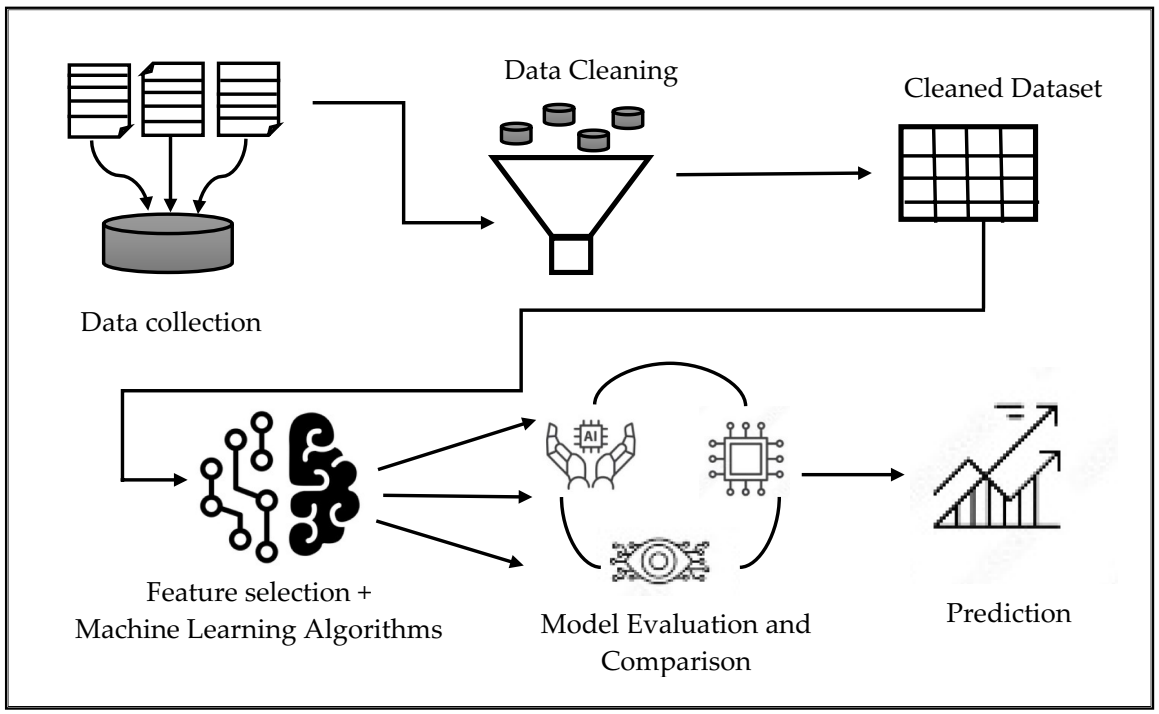

Fig. 1 Flow of methodology. This figure describes the whole process done throughout this work

Mutual Information

Basically, mutual information measures the mutual dependence between two random variables. Mutual information can be defined either on the basis of probability or in terms of entropies.

It can be defined either as,

$$I(X;Y) = \sum_{y\acute{\epsilon}Y} \sum_{x\acute{\epsilon}X} p(x,y) log\left(\frac{p(x,y)}{p(x).p(y)}\right) \quad (1)$$

where, $P(x, y)$ is the joint probability. $P(x)$ *and* $P(y)$ are the marginal probabilities.

If we are using continuous variables then the summation could be replaced by integration.

Or as

$$I(X;Y) = H(X) + H(Y) - H(X|Y)$$

$$= H(X,Y) - H(X|Y) - H(Y|X) \quad (2)$$

Where, $H(x)$ *and* $H(y)$ are the marginal entropies. $H(X|Y)$ *and* $H(Y|X)$ are conditional entropies. $H(X, Y)$ is the joint entropy.

Mutual information can be used in various applications and researches for feature selection. It is one of the most recent feature selection techniques used in researches.

Support Vector Machine (SVM)

Support vector machine is a widely used supervised machine learning algorithm for both classification and regression tasks. Consider a binary classification with a given training set of the form $T = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$ where $x_i \acute{\epsilon} R^n$ and $y_i \acute{\epsilon} \{+1, -1\}$. After training a model with this data, now the goal is to find a function to predict the target class of a new data point using its feature values x. A soft margin SVM formulates this problem as follows:

$$\min_{w,b,\xi} \frac{1}{2}||w||^2 + C\sum_{i=1}^{n} \xi_i \quad (3)$$

$$\text{s.t. } y_i\left((w.x_i) + b\right) \geq 1 - \xi_i, i = 1, 2, \ldots, n$$

$$\xi_i \geq 0, i = 1, 2, \ldots, n$$

Random Forest (RF)

Random forest algorithm is an ensemble algorithm, which can be used for both classification and regression problems. Here, in this study, it is used for the classification task. It uses a number of decision tree classifiers on various subsamples and finally it will average the accuracies (Fig. 2)

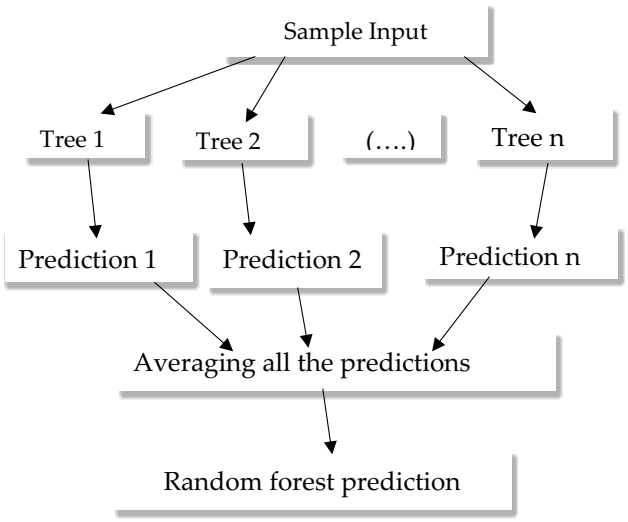

Fig. 2 Random forest Architecture

in order to improve the prediction accuracy and control the overfitting of the model.

K-Nearest Neighbor (KNN) Classifier

KNN is a simple but powerful classifier used in supervised learning. In this technique, a new instance will be assigned to its target class based on the number of closest neighbors. Number of neighbors want to be considered (the K value) can be defined or selected by the user. Out of those k values, this algorithm will find which class has the highest number of closest points and this new instance will be assigned to that class. To find the closest points, some distance measures such as Euclidean distance or Mahalanobis distance can be used.

Accuracy Measures

This study uses accuracy to measure the performance of the model that is defined by the following equation.

$$Accuracy = \frac{True\ Positive + True\ Negative}{TP + FP + TN + FN} \quad (4)$$

Here, True Positive (TP): The values correctly predicted as positive. True Negative (TN): The values correctly predicted as negative. False Positive (FP): The values are falsely predicted as positive. False Negative (FN): The values are falsely predicted as negative.

Results

This study starts with data normalization. As our initial data is not normalized, and the data range difference is too big (Fig. 3A), we normalized the data between 0 and 1 (Fig. 3B). This normalized data is used in the further analysis.

As this dataset is in a very high dimension where 2,932,684 methylation sites were measured on 40 samples, we used mutual information to select the features of the study. Five, ten, fifteen, and finally twenty methylation features are individually selected from mother and father. Those data are used with the classification models in order to predict the labor category, whether it is preterm labor or full-term labor.

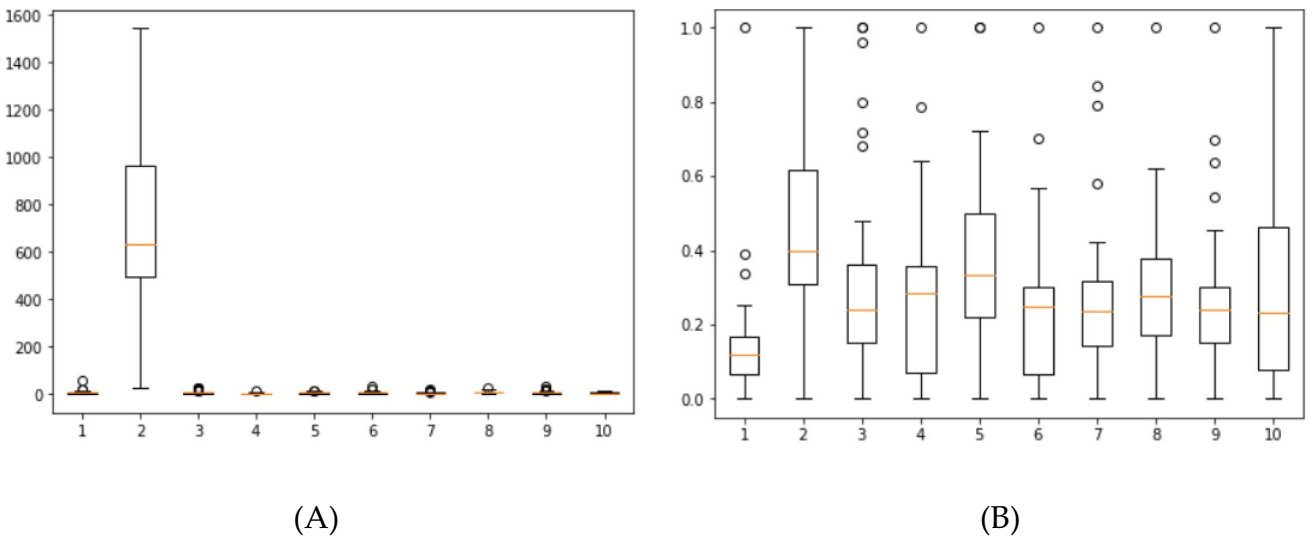

Fig. 3 Data visualisation. Distribution of the data **A** before normatlsation and **B** after normalization. As the range of the data is too big, the values are normalized between 0 and 1

Table 1 Selected Methylation features from Father and Mother in the prediction of preterm birth

Father			Mother		
Chr	Start	Stop	Chr	Start	Stop
1	145,794,001	145,795,000	1	56,465,001	56,466,000
2	109,646,001	109,647,000	1	227,244,001	227,245,000
3	147,583,001	147,584,000	5	82,100,001	82,101,000
4	70,059,001	70,060,000	6	2,753,001	2,754,000
5	137,490,001	137,491,000	6	68,425,001	68,426,000
6	23,570,001	23,571,000	8	22,902,001	22,903,000
7	123,063,001	123,064,000	8	104,742,001	104,743,000
8	21,244,001	21,245,000	10	28,627,001	28,628,000
9	113,436,001	113,437,000	11	93,018,001	93,019,000
10	87,487,001	87,488,000	12	67,658,001	67,659,000
11	24,530,001	24,531,000	13	20,739,001	20,740,000
12	42,721,001	42,722,000	13	24,172,001	24,173,000
13	21,344,001	21,345,000	14	92,500,001	92,501,000
14	26,846,001	26,847,000	14	103,249,001	103,250,000
15	10,831,001	10,832,000	15	49,947,001	49,948,000
16	48,631,001	48,632,000	15	72,745,001	72,746,000
17	81,294,001	81,295,000	21	9,085,001	9,086,000
18	6,471,001	6,472,000	22	34,102,001	34,103,000
19	53,752,001	53,753,000	X	58,920,001	58,921,000
20	25,360,001	25,361,000	X	113,208,001	113,209,000

DNA methylation data from father and mother are used individually in the preterm birth. As the dataset has huge number of features, mutual information is used in the feature selection. Top twenty features selected using mutual information and used in this classification are presents

Selected twenty features from mother and father (Table 1) are presented in Fig. 4 (A) and (B) along with their mutual information values.

Those selected features are subjected to the classification models to predict the possibility of preterm birth, and the accuracy is validated using fivefold cross validation. Initially, top 5 features are used with all three classifiers and the number is then increased to ten, fifteen, and finally twenty. This process is done to select the number of features with the highest accuracy. The same process is applied to the data from both mother and father.

Table 2 shows that Father's methylation data can do an accurate prediction of the possibility of preterm labor. Even though random forest provides an accuracy value of 1 (± 0.00) with each of 5, 10, 15 and 20 features, it is worthy to note that top 15 features give the same accuracy value not only with random forest but also with all the classifiers. This accuracy value is confirmed using precision, recall, F1-score, and AUC value of 1.

However, the mother's methylation data also gives a comparable accuracy value in this prediction, especially with the random forest classifier. Here top 10 features give the best accuracy in all three cases with very low variation.

Even though we have a perfect accuracy while using the methylation data of the father, we tried to use the combined data of both the parents in the same prediction. All the analyses are done in the same way as explained above. Here also we get a perfect accuracy (1 ± 0.00). However, most of the selected features are from the father, and a few top features from the mother also selected in this prediction.

Discussion

Premature deliveries need to be controlled in order to have a healthy community and it should be predicted in the early gestational weeks, not at the later stages. This will help a lot

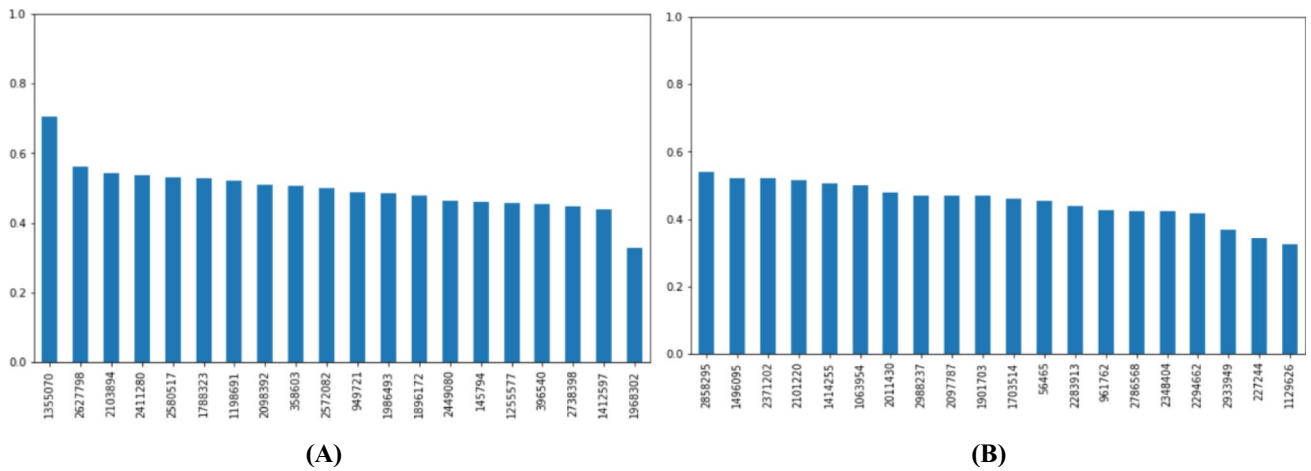

Fig. 4 Mutual information selected top twenty features from **A** Father and **B** Mother. As the dimension of our initial data is too high, mutual information is used to select the relevant features of the study. Selected features (X-axis) are presented against the mutual information value (Y-axis)

Table 2 Methylation data of father and mother is individually used in the prediction of the premature birth of a child

Classification methods	Father's data	Mother's data
Top 5 features		
Random forest	1.00 (±0.00)	0.88 (±0.22)
SVM	0.97 (±0.10)	0.90 (±0.19)
KNN	0.90 (±0.19)	0.93 (±0.20)
Top 10 features		
Random forest	1.00 (±0.00)	0.97 (±0.10)
SVM	1.00 (±0.00)	0.90 (±0.19)
KNN	0.97 (±0.10)	0.93 (±0.12)
Top 15 features		
Random forest	1.00 (±0.00)	0.97 (±0.10)
SVM	1.00 (±0.00)	0.90 (±0.24)
KNN	1.00 (±0.00)	0.82(±0.12)
Top 20 features		
Random forest	1.00 (±0.00)	0.97 (±0.10)
SVM	1.00 (±0.00)	0.97 (±0.10)
KNN	0.97 (±0.10)	0.85 (±0.10)

Mutual information is used to select the top 5, 10, 15, and 20 methylation sites, and those selected data are used in the prediction using three different classification algorithms such as RF, SVM, and KNN. Results after fivefold cross-validation are summarized

for the doctors to take necessary actions ontime to prevent preterm birth. There are various machine learning prediction models which predict preterm birth using several other data and their accuracies range between 60 to 99% [24–29, 31, 32]. Those studies are summarized in Table 3. They show that their accuracies are not too good, especially for this scenario, where every wrong decision may lead to a disaster up to the level of the loss of life of an infant. In one of those studies, they achieved 99% accuracy using a dataset which needs more computational power and time in the training of the model compared to our data.

In this prediction, each and every false negative prediction will also lead to many psychological effects, particularly on the mother. Here we need to note that stress of the mother is identified as one cause of this preterm birth [35–37]. So we need to be very careful with every wrong predictions. Hence, our accuracy in this study is validated using many performance metrics such as precision, accuracy and F1-score.

However, there are biological studies already showed the relationship between omics scale data and premature deliveries [38–41]. Various studies in the literature showed that preterm birth is highly influenced by genetic factors, especially from the maternal side [42–44]. Studies showed that if a mother already has a premature delivery, she is more likely to has the same in the coming deliveries as well [45]. Also, few studies proposed that 25–40% of premature deliveries maybe by hereditary, and maternal genomes has more effect on preterm birth compared to the fetal genome [46].

Studying the literature shows that most of the studies mainly considered either the maternal side or the fetal side in the genome-wide studies. Even in their studies, [45] reported that if a mother herself was born preterm, she is more likely to deliver preterm babies and this cannot be applied to the father. As a consequence of these findings, very few studies considered father's data along with mother's data to see the contribution of father's genetic data in the premature labor [7].

As the follow up of the literature, which says that preterm birth is ancestral, here, in this study we analyse the epigenetic data from mother and father separately in the prediction of preterm birth. Our study shows that epigenetic data from the father alone could be used in the prediction of preterm deliveries. This data gives a perfect accuracy, whereas the mother's data alone gives an accuracy comparably less than the father's data. Even though the difference between both the accuracies is very low, this is a surprising result, because mother is always more connected to the maternity-related issues.

Hence we study the experiment detail of the data we used in this prediction in order to justify our results. However, we could not find any special arrangements in their experimental setup or in the recruitment of their samples. At least they did not mention any limitations in their document regarding

Table 3 Accuracies obtained in the literature

Study	Data used	Used machine learning method	Accuracy of the study
[24]	Clinical Data	ANN, Logistic regression, Random forest, Decision trees	60–80% with ANN
[25]	Symptoms	Decision tree, Logistic regression, SVM	90.9% with SVM
[26]	EHG signal	ANN	95%
[27]	EHR data	RNN, Regularized logistic regression, SVM, Gradient boosting	82.7% with RNN
[28]	EHG recording		99%
[29]	Time-lapse images	Deep learning model	96.8%
[31]	miRNA	SVM	71%
[32]	Metabolome	Light GBM, Logistic regression, SVM, Elastic net	81% with LightGBM

In the prediction of preterm birth, various data and machine learning method are used and they gained different accuracies. Those studies are summarized here with the full description for comparing our study with the literature

the sample selection. Even they collected this data from buccal cells, which is very easy to obtain compared to other complex body parts. The only limitation for a machine learning model using this dataset is the limited size of samples in this study. However, one study already used this data set [7] and anaysed both data from the mother and father. In their study, they identified the contribution of the father's methylation data along with mother's data in the preterm birth. They further revealed the contribution of paternal germline to preterm birth.

This study supports our findings, and our study gives a future scope to the researchers in the direction of paternal effects in preterm birth. As we have very few paternal-related studies including [33] in this area, this study suggests that we should consider paternal side in these birth complications. If the mother does not have any other external complications, paternal side effect may have more impact on the healthy delivery of a baby compared to maternal side.

Conclusion

This study compares the performance differences between the methylation data taken from father and mother in the prediction of preterm birth using machine learning algorithms. As this dataset has more features compared to the number of samples, top 5, 10, 15, and 20 features are selected using mutual information. Each set of features is used in the prediction of preterm birth possibility using three different classifiers: random forest, SVM, and KNN classifier. Performance comparison between models after fivefold cross-validation shows that top 15 epigenome features obtain from father can predict premature labor with almost perfect accuracy. This finding is confirmed using the other measures such as precision, recall, F1-score, and AUC values. This study shows that compared to the DNA methylation features from mother of a preterm baby, the epigenetic markers from father carry more information related to the premature labor. However, DNA methylation data of the mother also equally informative in relation to the preterm birth of a baby.

Future Work

In this data, we identified that DNA methylation from the father can predict preterm birth accurately using machine learning techniques. In the future, we will be using other omic scale data also in the same prediction, in order to confirm the paternal effect in the preterm birth.

Funding The author(s) received no financial support for the research, authorship, and/or publication of this article.

Data availability Datasets used in this study are publicly available in GEO data repository. All the models use pre-built functions in python libraries, none of them are implemented here from the scratch.

Declarations

Conflict of interest The author declares there are no competing interests.

References

1. Walani S. Global burden of preterm birth. Int J Gynecol Obstet. 2020;150(1):31–3.
2. Gurung A, Wrammert J, Sunny AK, et al. Incidence, risk factors and consequences of preterm birth—Findings from a multi-centric observational study for 14 months in Nepal. Arch Public Health. 2020. https://doi.org/10.1186/s13690-020-00446-7.
3. Newnham J, Dickinson J, Hart R, Pennell C, Arrese C, Keelan J. Strategies to prevent preterm birth. Front Immunol. 2014. https://doi.org/10.3389/fimmu.2014.00584.
4. Knijnenburg WTA, Vockley JG, Chambwe N, Gibbs DL, Humphries C, Huddleston KC, Klein E, Kothiyal P, Tasseff R, Dhankani V, Bodian DL. Genomic and molecular characterization of preterm birth. Proc Natl Acad Sci. 2019;116(12):5819–27.
5. Vidal AC, Neelon SEB, Liu Y, Tuli AM, Fuemmeler BF, Hoyo C, Murtha AP, Huang Z, Schildkraut J, Overcash F, Kurtzberg J, Jirtle RL, Iversen ES, Murphy SK. Maternal stress, preterm birth, and DNA methylation at imprint regulatory sequences in humans. Genet Epigenet. 2014;6:118067.
6. Parets SE, Conneely KN, Kilaru V, Fortunato SJ, Syed TA, Saade G, Smith AK, Menon R. Fetal DNA methylation associates with early spontaneous preterm birth and gestational age. PLoS ONE. 2013;8(6):e67489.
7. Winchester P, Nilsson E, Beck D, Skinner M. Preterm birth buccal cell epigenetic biomarkers to facilitate preventative medicine. Sci Rep. 2022. https://doi.org/10.1038/s41598-022-07262-9.
8. Menon R, Conneely KN, Smith AK. DNA methylation: an epigenetic risk factor in preterm birth. Reprod Sci. 2012;19(1):6–13.
9. Yulu Z, Zheng G, Yanbo Z, Jianjing S, Leilei Y, Ping F, Yizhi L, Xingang L, Hao W, Ling R, Wei Z, Haifeng H, Xuerui T, Wei W. Rapid triage for ischemic stroke: a machine learning-driven approach in the context of predictive, preventive and personalised medicine. EPMA J. 2022;13(2):285–98.
10. Siddiq M. Revolutionizing drug discovery; transformative role of machine learning. Bull J Multidisiplin Ilmu. 2022;1(2):192–70.
11. Yang Y, Xu L, Sun L, Zhang P, Farid SS. Machine learning application in personalised lung cancer recurrence and survivability prediction. Computat Struct Biotechnol J. 2022;20:1811–20.
12. René B, Pascal C, Matts K, Francois M, Kenta Y, Jérémie G, Chunze L, Ulrich B, Jin YJ. Support to early clinical decisions in drug development and personalised medicine with checkpoint inhibitors using dynamic biomarker-overall survival models. Br J Cancer. 2023;129(9):1383–8.
13. UmaMaheswaran S, Munagala N, Mishra D, Othman B, Sinthu S, Tripathi V. The role of implementing Machine Learning approaches in enhancing the effectiveness of HealthCare service. In: 2022 2nd International Corence on Advance Computing and Innovative Technologies in Engineering (ICACITE). Piscataway: IEEE; 2022.
14. Belaghi RA, Beyene J, McDonald S. Prediction of preterm birth in nulliparous women using logistic regression and machine learning. PLoS One. 2021;16(6):e0252025.

15. Koivu A, Sairanen M. Predicting risk of stillbirth and preterm pregnancies with machine learning. Health Inf Sci Syst. 2020. https://doi.org/10.1007/s13755-020-00105-9.
16. Raja R, Mukherjee I, Sarkar BK. A machine learning-based prediction model for preterm birth in rural India. J Healthc Eng. 2021;2021:1–11.
17. Rittenhouse K, Vwalika B, Keil A, Winston J, Stoner M, Price J, Kapasa M, Mubambe M, Banda V, Muunga W, Stringer J. Improving preterm newborn identification in low-resource settings with machine learning. PLoS One. 2019;14(2):e0198919.
18. Sharifi-Heris Z, Laitala J, Airola A, Rahmani A, Bender M. Machine learning approach for preterm birth prediction using health records: systematic review. JMIR Med Inform. 2022;10(4):e33875.
19. Khan M, Aziz S, Ibraheem S, Butt A, Shahid H. Characterization of term and preterm deliveries using electrohysterograms signatures. In: Khan M, editor. IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON). Piscataway: IEEE; 2019. p. 0899–905.
20. Aung M, Yu Y, Ferguson KK. Prediction and associations of preterm birth and its subtypes with eicosanoid enzymatic pathways and inflammatory markers. Sci Rep. 2019. https://doi.org/10.1038/s41598-019-53448-z.
21. Jager F, Libenšek S, Geršak K. Characterization and automatic classification of preterm and term uterine records. PLOS ONE. 2018;13(8):e0202125.
22. Plunkett J, Feitosa M, Trusgnich M, Wangler M, Palomar L, Kistka Z, DeFranco E, Shen T, Stormo A, Puttonen H, Hallman M, Haataja R, Luukkonen A, Fellman V, Peltonen L, Palotie A, Daw E, An P, Teramo K, Borecki I, Muglia L. Mother's genome or maternally-inherited genes acting in the fetus influence gestational age in familial preterm birth. Hum Hered. 2009;68(3):209–19.
23. DeFranco E, Teramo K, Muglia L. Genetic influences on preterm birth. Semin Reprod Med. 2007;25(1):40–51.
24. Reza AB, Joseph B, Sarah DM. Prediction of preterm birth in nulliparous women using logistic regression and machine learning. PLOS ONE. 2021;16(6):e0252025.
25. Giovanni I, Rakesh R, Indrajit M, Kanti SB. A machine learning-based prediction model for preterm birth in rural India. J Healthc Eng. 2021;2021:1–11.
26. Paul F, Pauline C, Abir H, Dhiya A-J, Chelsea D, Iram SS. Prediction of preterm deliveries from EHG signals using machine learning. PLOS ONE. 2013;8(10):e77154.
27. Gao C, Osmundson S, Edwards DRV, Jackson GP, Malin BA, Chen Y. Deep learning predicts extreme preterm birth from electronic health records. J Biomed Inform. 2019;100:103334.
28. Despotović D, Zec A, Mladenović K, Radin N, Turukalo T. A machine learning approach for an early prediction of preterm delivery. In: Despotović D, editor. 2018 IEEE 16th International Symposium on Intelligent Systems and Informatics (SISY). Piscataway: IEEE; 2018.
29. Bo H, Shunyuan Z, Bingxin M, Yongle Y, Shengping Z, Lei J. Using deep learning to predict the outcome of live birth from more than 10,000 embryo data. BMC Pregnancy Childbirth. 2022. https://doi.org/10.1186/s12884-021-04373-5.
30. Sun Q, Zou X, Yan Y, Zhang H, Wang S, Gao Y, Liu H, Liu S, Lu J, Yang Y, Ma X. Machine learning-based prediction model of preterm birth using electronic health record. J Healthc Eng. 2022;2022:1–12.
31. Burris HH, Gerson KD, Woodward A, Redhunt AM, Ledyard R, Brennan K, Baccarelli AA, Hecht JL, Collier A-RY, Hacker MR. Cervical microRNA expression and spontaneous preterm birth. Am J Obstet Gynecol MFM. 2023;5(1):100783.
32. William FK, Federico B, Martin CL, Jingqiu L, Yoli M, Harry HL, Almut H, Ines T, Christoph AT, Maayan L, Tal K. Preterm birth is associated with xenobiotics and predicted by the vaginal metabolome. Nat Microbiol. 2023;8(2):246–59.
33. Luo R, Mukherjee N, Chen S, Jiang Y, Arshad S, Holloway J, Hedman A, Gruzieva O, Andolf E, Pershagen G, Almqvist C, Karmaus W. Paternal DNA methylation may be associated with gestational age at birth. Epigenet Insights. 2020;13:251686572093070.
34. Omnibus G. E. "GEO," 09 Mar 2022. [Online]. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE194227. (Accessed 2023).
35. Yu J, Wei Z, Wells JC, Fewtrell M. Effects of relaxation therapy on maternal psychological status and infant growth following late preterm and early-term delivery: a randomized controlled trial. Am J Clin Nutr. 2023;117(2):340–9.
36. Akanksha D, Vatsla D, Perumal V, Rajesh S, Aparna S, Ramesh A, Neena M, Juhi B. Prevalence of mental health problems in mothers of preterm infants admitted to NICU: a cross-sectional study. Int J Gynecol Obstet. 2023;160(3):1012–9.
37. Ejder TS, Merve L, Mehtap N, Ejder AS, Şerafettin TK. The relationship of preterm, term, and post-term births to maternal stress and human milk cortisol levels. Breastfeed Med. 2023;18(6):462–8.
38. Reiss JD, Peterson LS, Nesamoney SN, Chang AL, Pasca AM, Marić I, Shaw GM, Gaudilliere B, Wong RJ, Sylvester KG, Bonifacio SL, Aghaeepour N, Gibbs RS, Stevenson DK. Perinatal infection, inflammation, preterm birth, and brain injury: a review with proposals for future investigations. Exp Neurol. 2022;351:113988.
39. Juhi KG, Angharad C, Laura G, Zarko A, Bertram M-M, Ana A. Genome and transcriptome profiling of spontaneous preterm birth phenotypes. Sci Rep. 2022. https://doi.org/10.1038/s41598-022-04881-0.
40. Paquette AG, MacDonald J, Bammler T, Day DB, Loftus CT, Buth E, Mason WA, Bush NR, Lewinn KZ, Marsit C, Litch JA, Gravett M, Enquobahrie DA, Sathyanarayana S. Placental transcriptomic signatures of spontaneous preterm birth. Am J Obstet Gynecol. 2023;228(1):73.e1-73.e18.
41. Camunas-Soler J, Gee EP, Reddy M, Mi JD, Thao M, Brundage T, Siddiqui F, Hezelgrave NL, Shennan AH, Namsaraev E, Haverty C, Jain M, Elovitz MA, Rasmussen M, Tribe RM. Predictive RNA profiles for early and very early spontaneous preterm birth. Am J Obstet Gynecol. 2022;227(1):72.e1-72.e16.
42. Couture C, Brien M-E, Boufaied I, Duval C, Soglio DD, Ann LE, Cox EB, Girard S. Proinflammatory changes in the maternal circulation, maternal–fetal interface, and placental transcriptome in preterm birth. Am J Obstet Gynecol. 2023;228(3):332.e1-332.e17.
43. Paul W, Eric N, Daniel B, Michael KS. Preterm birth buccal cell epigenetic biomarkers to facilitate preventative medicine. Sci Rep. 2022. https://doi.org/10.1038/s41598-022-07262-.
44. Viral GJ, Nagendra M, Ge Z, Louis JM. Genetics, epigenetics, and transcriptomics of preterm birth. Am J Reprod Immunol. 2022. https://doi.org/10.1111/aji.13600.
45. Varner M, Esplin M. Current understanding of genetic factors in preterm birth. BJOG. 2005;112:28–31.
46. Stevenson D, Wong R, Shaw GM. The contributions of genetics to premature birth. Pediatr Res. 2019;85:416–7.