International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 / https://doi.org/10.14445/22315381/IJETT-V70I5P207

Volume 70 Issue 5, 46-59, May 2022

Review Article

A Systematic Review using Machine Learning Algorithms for Predicting Preterm Birth

Dr. R. Surendiran[1], R. Aarthi[2], M. Thangamani[3], S. Sugavanam[4], R. Sarumathy[5]

[1]School of Information Science, Annai College of Arts and Science, Kumbakonam.
ORCID: 000-0003-1596-7874,
[2]Department of Information Technology, Kongu Engineering College, Erode.
ORCID :0000-0001-8939-3434,
[3]Department of Information Technology, Kongu Engineering College, Erode.
ORCID: 000-0001-8864-3315,
[4,5]Department of CSE, Arjun College of Technology, Tamilnadu, India.

surendiranmca@gmail.com

Received: 29 January 2022 Revised: 11 April 2022 Accepted: 30 April 2022 Published: 19 May 2022

Abstract - *Preterm births (PTB) affect nearly 15 million kids worldwide. At present, medical fields aim to reduce the possessions of prematurity rather than avoid it. The cervix is currently measured during a transvaginal ultrasound, used to diagnose the condition. Because of the complexities of this process, preterm births cannot be accurately predicted. Machine learning is becoming more popular for prediction and diagnosis in health care. This study looks at how artificial intelligence can predict preterm labor and birth. According to this study, various machine learning approaches can aid in the diagnosis of preterm births. In terms of predicting preterm birth, machine learning can be well suited for various data types. Electro hysterogram signals, electronic health records, and transvaginal ultrasounds are examples. This review's goal remains to summarize machine learning procedures intended for predicting premature birth.*

Keywords - *Prediction, Preterm birth, Artificial Intelligence, Machine learning.*

1. Introduction

World Health Organization (WHO) defines Preterm birth or Premature as a birth occurring before 37 gestation weeks of pregnancy. A normal pregnancy is considered to last about 40 weeks [1]. Normally during the last week of pregnancy baby in the womb can gain weight, and organs such as the brain and lungs get fully developed. This may lead to long-lasting health issues such as physical disabilities or learning disabilities if this may not do. Before 37 weeks of gestation, newborns need to stay in the neonatal intensive care unit (NICU) for an extended period. This continuous stay in the hospital creates stress among the family members and increases the health care charges, mainly during the first year after birth [2]. For example, in 2005, in the United States, 26.2 billion dollars accounted the annual financial cost related to preterm birth. In 2014, preterm birth was estimated to cost over 500 million dollars in Canada [3-4]. Finding mostly the root of premature birth is a difficult task, and there is no predetermined reason for their occurrence [5].

Earlier research 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., [6]. Women who seem to have a preterm birth, on the other hand, frequently have no known risk factors [7]. A model that combines cervical length and obstetric history predicts spontaneous preterm labor better than both factors [8]. These factors point to the inefficiency of previous methods for predicting the risk of labor in pregnant women for the first time.

Furthermore, many predictive systems based on maternal sociodemographic factors have been investigated. Unfortunately, their ability to predict is severely limited [9]. As a result, numerous researchers have tried to predict premature birth using a machine learning approach on a collection of known clinical characteristics [70].

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 [11-12]. Figure- 1 shows the types of the machine learning algorithm.

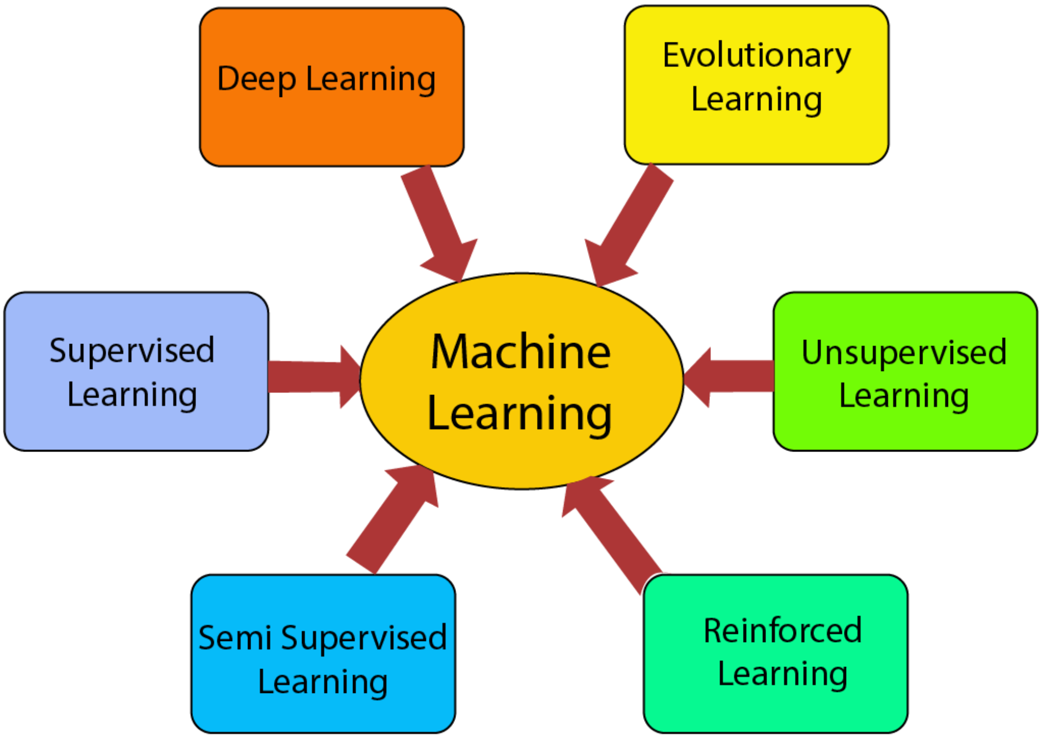

Fig. 1 Different Machine Learning Types

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 [13]. ML takes advantage of various variables from electronic health record (EHR) data for PTB prediction [14].

In this work, Sufriyana et al. [15] look at studies that use machine learning algorithms to predict preterm birth, which could be useful in perinatal medicine. Fortunately, most countries' health records (HRs) include information on a person's sociodemographic, medical history, and obstetric. As a result, HRs 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, successful attempts to use transvaginal (TVS) ultrasound imaging data have emerged. Recently, a few research findings used the deep learning method to predict PTB with the help of ultrasound and MRI images [16] and high-dimensional EHR data.

2. Background

2.1. Preterm birth (PTB)

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 fifteen 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 globe. Premature birth is the leading cause of infant mortality and illness [17]. Fig 2 describes the factors that affect preterm birth.

2.2. PTB Types

PTB is classified as either spontaneous or iatrogenic.

2.2.1. Spontaneous

It could be caused by spontaneous on-set labor or a premature tissue rupture before labor. Determining the source of this kind of preterm delivery is extremely problematic in approximately 50 % of cases

2.2.2. Iatrogenic

It's a type of elective and induction labor that occurs well before the 37th week of pregnancy. For various reasons, such as foetal or maternal health or other medical reasons.

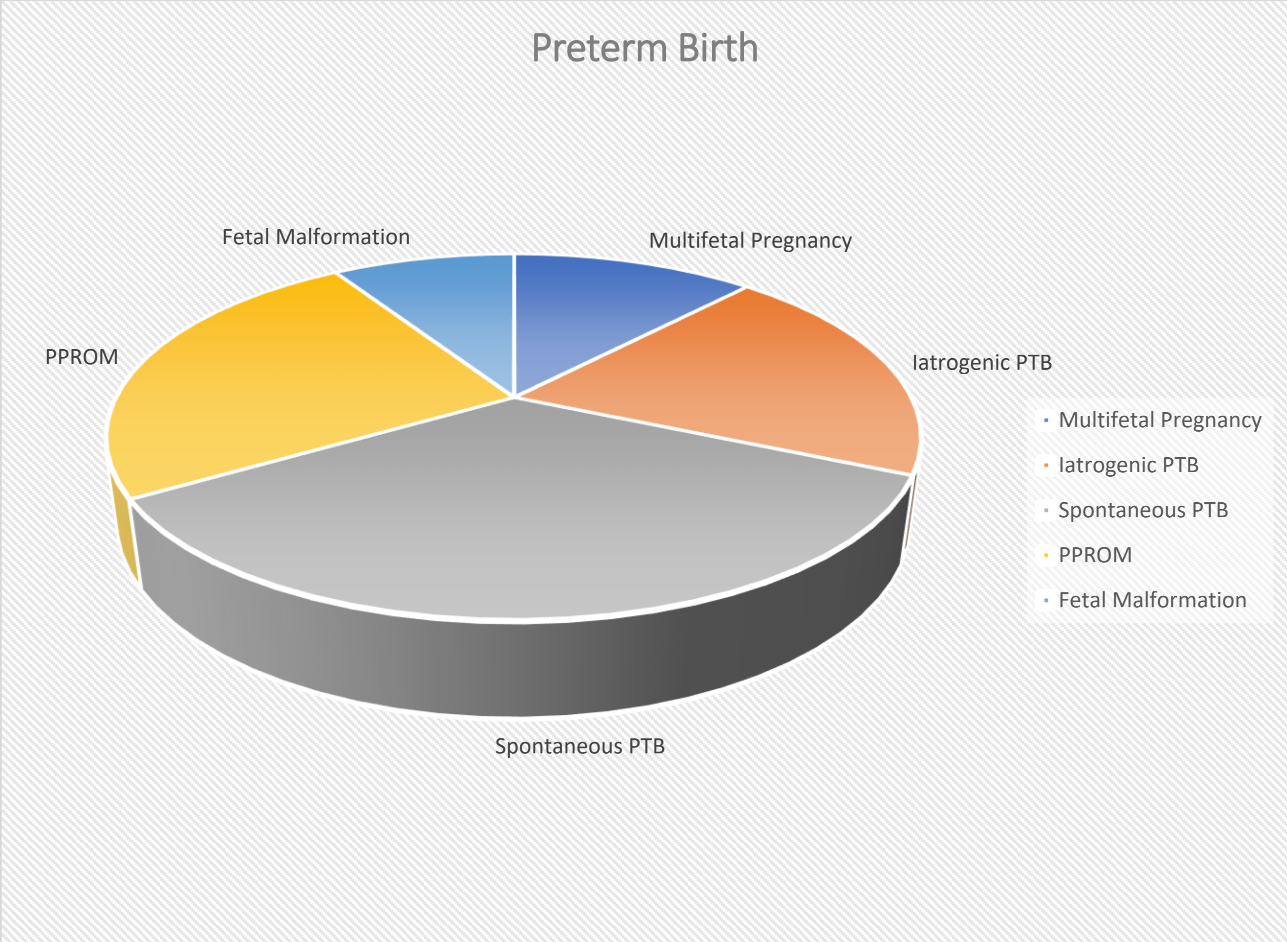

Fig. 2 Epidemiology of Preterm Birth

2.3. 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 [18]. The following are the four types of PTB [7]:

2.3.1. 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.

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.

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.

Late PTB

The range between 34 to 37 weeks of pregnancy. When a baby is born between 34 and 37 weeks of pregnancy.

2.4. 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. sPTB is detected in about half of all women with no identified clinical risk features. PTB rates were not reduced by sPTB 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.

2.5. Machine Learning

Machine learning models in supervised learning learn to predict predefined characteristics or results (also known as labels, goals, target variables, or outcome) related to a model defined by using a collection of attributes (also known as input features, descriptive variables). A machine-learning model accomplishes this by assuming a specific input/output relationship and then training it with a set of models designed together with input and output [19]. Machine learning has several advantages over conventional methods, including scalability and flexibility, which allows it to be used for various tasks like risk prediction, diagnosis, and classification [20-21].

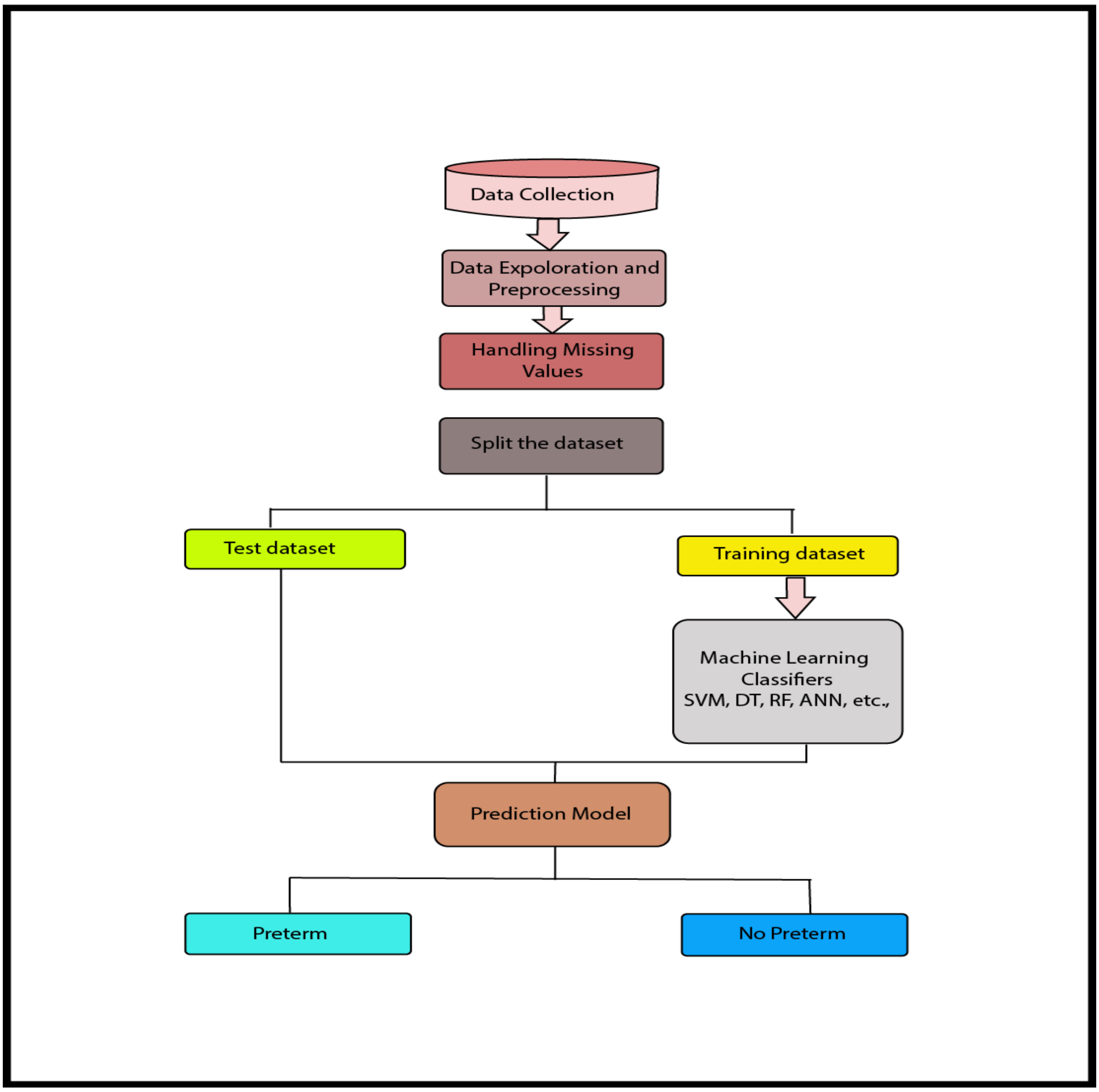

Fig. 3 Outline of Preterm Prediction Systems using Machine Learning algorithms

The articles contain important information about various machine learning classification models that researchers can use [22-24]. The constraints of the selected input/output relationship are typically adjusted iteratively during this training procedure. After being trained, the model is useful to apply for hidden samples and forecasts about the looked-for output. The types of input and output that different machine-learning models can perform and, therefore, how well they can interact with the varying effects of the provided input and output relationship may differ.

Machine learning models include linear regression variants [71] and decision tree-based systems [26], ensemble approaches like random forests [27] or gradient-boosted trees [28], support vector machines [29], nearest-neighbors approaches [30], and Bayesian techniques [31], based upon that function class that was used to create the input/output relationship. Artificial Neural Networks and Deep learning models [32] can be used to analyze tabular data, but other methods frequently perform better than convolutional neural networks in this task [34-35]. The above fig 3 gives details about the flow of the prediction procedure.

3. Related works

3.1. Electro hysterography (EHG)

Electro hysterography (EHG) is a non-invasive procedure used to determine the electrical factor that causes uterine contractions. EHG is a method of using contact electrodes to record electric pulses in the maternal abdomen.

3.2. Handling Imbalanced Data

The Term Preterm EHG Database (TPEHG) is the most widely investigated EHG delivery database. The unequal supply of term and preterm EHG data was addressed using data pre-processing techniques with the help of adaptive synthetic sampling (ADASYN) methodology and synthetic minority over-sampling technique (SMOTE). The accuracy, specificity, sensitivity, and the area under the curve (AUC) of the receiver could all be used to evaluate the results of the created models.

3.3. Metrics for Evaluation

The following formulas are used to calculate accuracy (ACC) (1), the sensitivity (Se) (2), specificity (Sp) (3), and the classifiers:

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad \ldots\ldots\ldots (1)$$

$$Se = \frac{TP}{TP+FN} \times 100 \quad \ldots\ldots\ldots\ldots\ldots\ldots (2)$$

$$Sp = \frac{TN}{TN+FP} \times 100 \quad \ldots\ldots\ldots\ldots\ldots\ldots (3)$$

The area under the curve (AUC) (4) is calculated as follows after selecting the optimal results for each classifier:

$$AUC = \int Se(T)(1 - Sp)'(T)dT \ldots\ldots (4)$$

After retrieving 203 temporal, spatial, and non-linear characteristics from 326 multi-channel EHG recordings, SMOTE has been used to control the database. Using an ensemble classifier, the scientists informed a mean score of F1 as 92.04% later feature selection via a genetic algorithm. A Linear Discriminant Analysis (LDA) algorithm using a regular F1 value of 90.1 percent was used to demonstrate the effectiveness of entropy metrics for the categorization of term preterm EHG recordings [35]. The accuracy of about 31 temporals and spatial characteristics found from 3 EHG channels before the 26th week through pregnancy were assessed. Once 15 features had been selected and ADASYN was used, a random forest classification model with 93 percent accuracy, 89 percent sensitivity, and 97 percent specificity was created [36].

The SampEn, log detector, waveform duration, and variance of only one EHG channel were all retrieved. Using a fusion of such a Case, specifically random neural network and feed-forward neural net classifiers and radial basis function, the authors achieved an AUC of 0.94, a specificity of 84 %, and a sensitivity of 91 %, and using SMOTE for data balancing [37]. By obtaining the root mean square (RMS), median frequency, and peak frequency, SMOTE was used to solve the imbalance problem. A solitary EHG channel is bandpass filtering by 0.34 to 1 Hz for SampEn. The polynomial classifier produced the best results, with a median sensitivity of 97 %, a specificity of 90 %, and an AUC of 0.95 [38].

Previous attempts used 30-minute recordings; however, Despotovic et al. [39] found that splitting the thirty-minute audio tapes into two 15-minute recordings yielded better results. They have unique features based on the signal's nonstationary and empirical method decomposition. An adaptive synthetic sampling (ADASYN) method is used to mitigate the imbalance in unbalanced memorization. On 322 datasets, among which 38 seem to be premature birth, ten-fold cross-validation yields 99 percent accuracy, 98 percent sensitivity, and 99 percent AUC to use a random forest classification algorithm and artificial samples. The obtained outcomes appear to be excellent. From EHG recordings, EMD was used to retrieve 11 IMFs, which were then divided into 6 stages by WPD. Using ADASYN to classify important features, SVM attained an accuracy of about 96.25 %, specificity of about 97.33 %, and sensitivity of about 95.08 % [40].

Nine characteristics were found in the third through 5th IMFs. Next, with the help of ADASYN via data rebalancing and SVM designed for classification, researchers found the accuracy was 98% [41]. Researchers used autoregressive modeling from a single EHG channel to find a novel measure named centroid frequency estimate. The SVM classifier achieved an accuracy of about 99.72%, a specificity of 99.96%, and a sensitivity of 99.48% after using ADASYN [42]. To create the feature vector, we employed the Shannon entropy for the 1st ten deconstructed IMFs. Some classifiers were employed to differentiate between term and preterm deliveries after SMOTE was used. The best result, according to the authors, was obtained with the help of the Adaboost classifier, which had a 0.986 AUC [43]

Researchers have been using autoregressive modeling to retrieve a new centroid frequency estimation proposal using only one EHG channel. Next, by using ADASYN, the SVM classifier had 99.72% accuracy, 99.96% specificity, and 99.48% sensitivity [44]. The dynamic changes that the uterus undergoes throughout the pregnancy are recorded, and the misleading interpolation is corrected. Entropy features describe the uterus's rate of evolution toward delivery. Authors, with the help of the Gaussian Naive Bayes (GNB) classifier and the principal components analysis (PCA), achieve an accuracy of about 0.75, a sensitivity of about 0.84, specificity of about 0.66, and AUC having 0.84 values [45].

The authors present a low processing complexity algorithm for detecting preterm labor using EHG signals. It has been demonstrated that EHG evaluation might be a useful tool for predicting and preventing preterm labor. By using SVM achieves results with 99.56%= accuracy, 98.95% = sensitivity, 99.30%=specificity [46]. The authors only use one EHG channel and don't use synthetic data creation or feature optimization techniques. K-nearest neighbors (KNN), support vector machine (SVM), and decision tree (DT) classifiers were used to determine whether the revealed EHG signal belonged to the premature birth case. SVM is the best among them, with an accuracy of 99.7%, specificity of about 99.7%, and sensitivity equal to 99.5% [47].

Boluda et al. [48] examined how well efficient and reliable algorithms based on classification performed using electro hysterographic recordings (EHG), which include RF, KNN, and ELM, for forecasting forthcoming labor in females through threatened preterm labor (TPL). KNN efficiency depended on the available dataset, but it had a high classification performance. Authors developed a model for predicting iatrogenic premature labor in prenatal females having scarred genital tract. The predictive model we developed can support physicians in assessing the risks of iatrogenic premature births and assist in creating consultations; thereby, improved medical treatment could be provided to enhance the status of patients and unborn children [49].

The objective of Asmi et al. [50] was to improve a robust and widely applicable classification model for predicting preterm labor using ANN. The classification models have fed single-channel and multi-channel EHG, which use novel fetal contractile performance indexes for EHG synchronization. The authors describe the various steps for detecting and predicting preterm birth, including pre-processing, extraction and classification, classification techniques, and attribute selection methods. More EHG signals, according to the authors, can enhance the precision of predicting premature birth. Table 1 compares classifiers in the EHG Dataset, including different classifiers comparisons based on evaluation metrics. Fig 3 shows the comparison chart of classifiers based on evaluation metrics.

Table 1. Comparison of classifiers in EHG Dataset

Author	Data Balancing	Classifier	Parameters
Fergus et al., 2013	SMOTE	Polynomial Classifier	AUC=0.99 Sensitivity=97% Specificity=90%
Ren et al., 2015	SMOTE	Adaboost classifier	AUC=0.986
Fergus et al., 2016	SMOTE	Feed forward, random classifiers, radial basis function neural network	Specificity=84% Sensitivity=91% AUC=0.94
Sadi et al.,2017	No	linear SVM classifier	Accuracy= 95.7% Specificity= 93% Sensitivity= 98.40% AUC= 0.95
Acharya et al., 2017	ADASYN	SVM	Accuracy=96.25% Specificity=97.33% Sensitivity=95.08%
Despotovic et al., 2018	ADASYN	Combining a random forest classifier with artificial samples	Sensitivity=98% Specificity=99% AUC=0.99
Shahbakhti et al., 2019	NO	SVM	accuracy= 99.56%, sensitivity= 98.95% and specificity= 99.30% accuracy= 99.56%, sensitivity= 98.95% and specificity= 99.30% Accuracy=99.56% Specificity =99.30% Sensitivity=98.95%
Degbedzui et al, 2020)	ADASYN	SVM	Accuracy=99.72% Specificity=99.48% Sensitivity=99.96%
Peng et al., 2020	ADASYN	Random forest Classifier	Accuracy= 93 % Sensitivity=89% Specificity=97%
Khan et al., 2020	ADASYN	SVM	Accuracy=98%
Lou et al., 2022	SMOTE	PCA, GNB	Accuracy=0.75, Specificity=0.66, Sensitivity=0.84, AUC=0.84
Mohammadi et al.,2022	NO	SVM	Accuracy =99.7% Specificity=99.7%. Sensitivity=99.5%

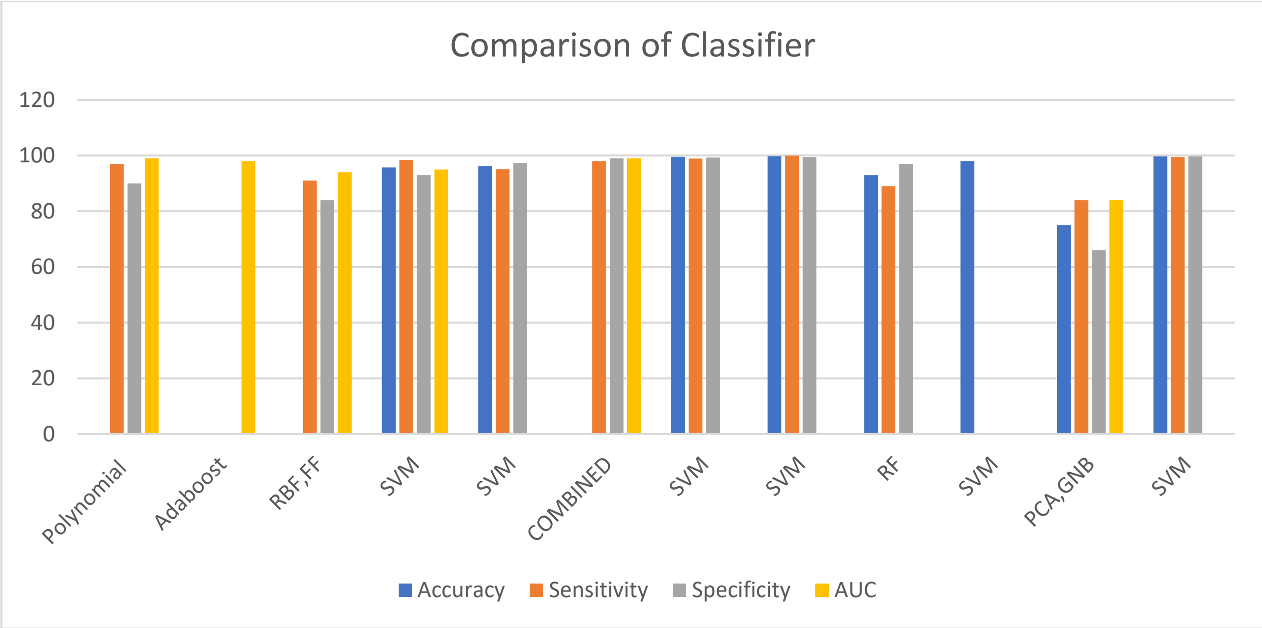

Fig. 4 Comparison of classifiers

3.4. Electronic Health Records

Electronic health records encompass information about the prenatal period progress, the patient's therapeutic history, and other personal information gathered during a medical interview. One of the most difficult problems in using HR-based data is incomplete data. Even though missing data has been identified as a problem in HR studies, only about half of them recognized the existence of incomplete data. They used a variety of methodologies to deal with it. Despite its significance, little investigation has been done in this area, and it is ambiguous just how biased findings impact the predictive model.

The author used data first from the "Preterm Prediction Study," the clinical testing data - sets gathered between 1992 and 1994, suggesting that existing PTB treatments were unavailable. Thus the set of data reflects natural PTB incidence. This article emphasizes predicting preterm labor in nulliparous mothers and understanding its multiple etiological factors. Vovsha's method employs SVM with linear as well as non-linear kernels, but also logistic regression. SVM of radial basis function (RBF) kernel's total score of 0.57 sensitivity and 0.69 specificities is the best performing classifier for all populations [51]. Generates a condensed set of risk factors with quantifiable uncertainties. Handle a large number of irrelevant features. The majority class is under-sampled to balance the data. Randomized Gradient Boosting (RGB) combines Stochastic Gradient Boosting and Random Forests that Tran uses in his work. The findings are presented and documented clearly and concisely. The three biggest risk factors have been numerous fetuses, cervix ineffectiveness, and previous preterm births. The maximum AUC when using RGB is between 0.80 and 0.81 [52].

Esty et al. [53] looked at two datasets that contained birth data. One of the goals of this study is to create a model that can perform better than fibronectin indicator predictions, which are expensive and invasive to screen for. The authors mention that both datasets have many missing variables, but they don't say how many. And from the other hand, features with a little more than 50% missing data are removed. The majority class is downsampled to balance the dataset. The authors employ a C5.0 Decision Tree for classification). Weber's research concentrates on preterm birth in nulliparous women. Logistic regression, lasso regression, generalized additive models, random forest, k-nearest neighbors, elastic net, and ridge regression with mixing parameters are the machine learning classifiers used. The performance of all of the algorithms is comparable. The AUC improves to 0.67 when combined racial-ethnic groups are predicted, comparable to others who use biomarkers [54].

Moreira et al. [55] research create a comprehensive intelligent system for mobile DSSs that can help women who may be at threat of pregnancy-related complications receive better care. As a result, this research may significantly improve maternal and fetal health by predicting preterm birth risk early. SVM is a machine learning (ML) technique for recognizing patterns in a pregnancy database. SVM ML-based technique produced promising results with 0.821 accuracies, 0.839 of true positive rate, 0.268 false-positive rates, and 0.785 the Receiver Operating Characteristic (ROC).

The authors provided a solution depending on logistic regression and the SVM algorithm, both of which are classification algorithms. The number of times of pregnancy, age, diabetes mellitus (GDM and DM), Obesity, and high blood pressure in pregnant women are all important predictors of spontaneous preterm labor. It demonstrates that the medical conditions identified can be used to predict sPTB. According to the findings, all these GDM and DM seem to be the main risk factors for premature births. Accuracy is 0.76, recall is 0.84, specificity is 0.73, and precision is 0.84 [56]. This study compares six machine learning (ML) algorithms for predicting PTB: artificial neural networks (ANN), decision tree logistic regression, SVM, and random forest. Previous preterm birth, age, diabetes, BMI, alcoholism, decision trees, smoking, hypertension, vitro fertilization, and cervical length are all factors taken into account. The model uses ANN to achieve 91.14 percent classification accuracy and 91.80 percent multinominal logistic regression [57].

Extreme preterm birth (EPB) infants brought into the world before the 28th week of pregnancy—were attempted to be predicted. Patient demographic characteristics, diagnosis and treatment procedures, medications prescribed, and laboratory results are all included in the data. Gao et al. represent each medical concept using a bag of words (BOW) in an addition word embedding. In BOW, the term frequency-inverse document frequency (TF-IDF) stabilizes the significance of each healthcare perception. A skip-gram stays used in word embedding to match the most connected words to a specified word. The models developed employ SVM and LR GB. LR seems to attain excellent BOW and word embedding performance, with an AUC of 0.780 [58].

An Artificial Neural Network (ANN) algorithm and algorithm like naive Bayes are used in the model proposed in research on preterm birth prediction. The results show that the help of the Artificial Neural Network (ANN) to predict preterm birth yields a 90.67 percent accuracy rate and a ROC value of 0.954. The Naive Bayes algorithm has a ROC value of 0.929 and an accuracy of 84.53 percent. As a result, the Artificial Neural Network (ANN) algorithm predicts preterm birth with a 6.14 percent accuracy and a ROC value of 0.025 [59]. We estimated spontaneous PTB at 28- and 32-weeks using variables, particularly during the first and second trimesters, with the help of multivariable logistic regression methods. AUC during 28 weeks duration in nulliparous and multiparous women and the first trimester was 68.5 percent and 73.4 percent, respectively. The AUCs for 2nd models in nulliparous and multiparous women were 72.4 percent and 78.2 percent, respectively [60].

Based on HER data, a prediction of premature birth in women who have had cervical cerclage was proposed, including features such as age, previous preterm birth, cervical length during cerclage, and uterine abnormality. KNN, neural networks, and random forest are among the classifiers compared by the authors. They have a precision of 98 percent and a specificity of 94 percent [61]. The authors suggested that new risk models be tested to find new ones used in healthcare situations. They use artificial neural networks, logistic regression, and gradient boosting decision trees, all cutting-edge machine learning algorithms. The AUCs for early stillbirth were 0.76, late stillbirth was 0.63, and preterm birth was 0.64. [13]

The characteristics of the mother and her medical history were collected. To predict the risk exposures of PTB, the data will be entered into the Foetal Medicine Foundation (FMF) online tool. Univariate and multivariate logistic regression assessments have been performed to evaluate the impact of parental character traits on PTB occurrence [62]. The authors are interested in determining preterm labor depending on clinical symptoms through longitudinal EHR. The problems based on prediction have been formulated as a classification process with noisy labels. A Recurrent Neural Network, including an attention Mechanism, serves as a base classifier. We believe that a data subgroup both with noisy and clean labels is available. To train a deep learning model with clean and noisy labels, we suggest an Alternating Loss Correction (ALC) technique [63]. This research aims to find alterations in the maternal plasma of pregnancies, delivered premature babies later spontaneously appear with symptoms without a previous record of premature births or known risk factors [64]. Thus, this study aims to add other maternal chrono disruption factors and see if they can achieve better preterm birth prevention. Indeed, the decision tree achieved as a prediction model that shows light coming in through the window or the liveliness status of the bedroom during the night is an important factor in predicting preterm delivery [65]. Following Table 2 compares classifiers using the HER dataset, and Fig 5 gives the chart comparison of the dataset.

Table 2. Comparison of classifiers in EHR Dataset

Authors	Classifier	PTB Gestation age	Results
Vovsha et al. (2014)	SVM	<32 weeks	Sensitivity- 0.57 Specificity= 0.69
Tran et al. (2016)	Logistic regression, Random forest stochastic gradient boosting, randomised gradient boosting,	<35 and <37 weeks	AUC=0.81
Esty et al (2018)	Decision trees, neural networks	<34 and <37 week	AUC = 0.81, Specificity = 0.72 Recall = 0.91,
Weber et al (2018)	K-NN, random forest, elastic net, Generalised Additive Models (GAM), ridge regression, lasso regression,	20-32 weeks	AUC = 0.67
Prema and Pushpa Latha (2019)	logistic regression, SVM		Recall = 0.84, Precision=0.84, Specificity = 0.73, Accuracy = 0.76
Lee and Ahn, (2019)	SVM, random forest, logistic regression, Naive Bayes, neural networks, decision trees	>20 and <37 week	Accuracy = 0.92
Gao et al. (2019)	SVM, gradient boosting (GB), linear regression (LR), and Long-term, short-term memory	<28 week	AUC=0.827, PVP=0.033 sensitivity=0.965, specificity=0.698,
Diah *et al*. (2020)	the Naive Bayes algorithm and Artificial Neural Network (ANN) algorithm		Accuracy=6.14% ROC=0.025
Koivu and Sairanen (2020)	ANN, LR	<37 week	AUC of 0.64
Belaghi et al., (2021)	multivariable logistic regression		AUC (1st trimester) 68.5% and 73.4 %, (2nd ^{t} trimester) 72.4 % and 78.2 %
Rawashdeh et al., (2021)	K-NN, random forest, neural networks, Naive Bayes classifier, decision trees,	<26 week	Accuracy = 0.95, Specificity = 0.94 AUC = 0.98, Recall = 1.0,

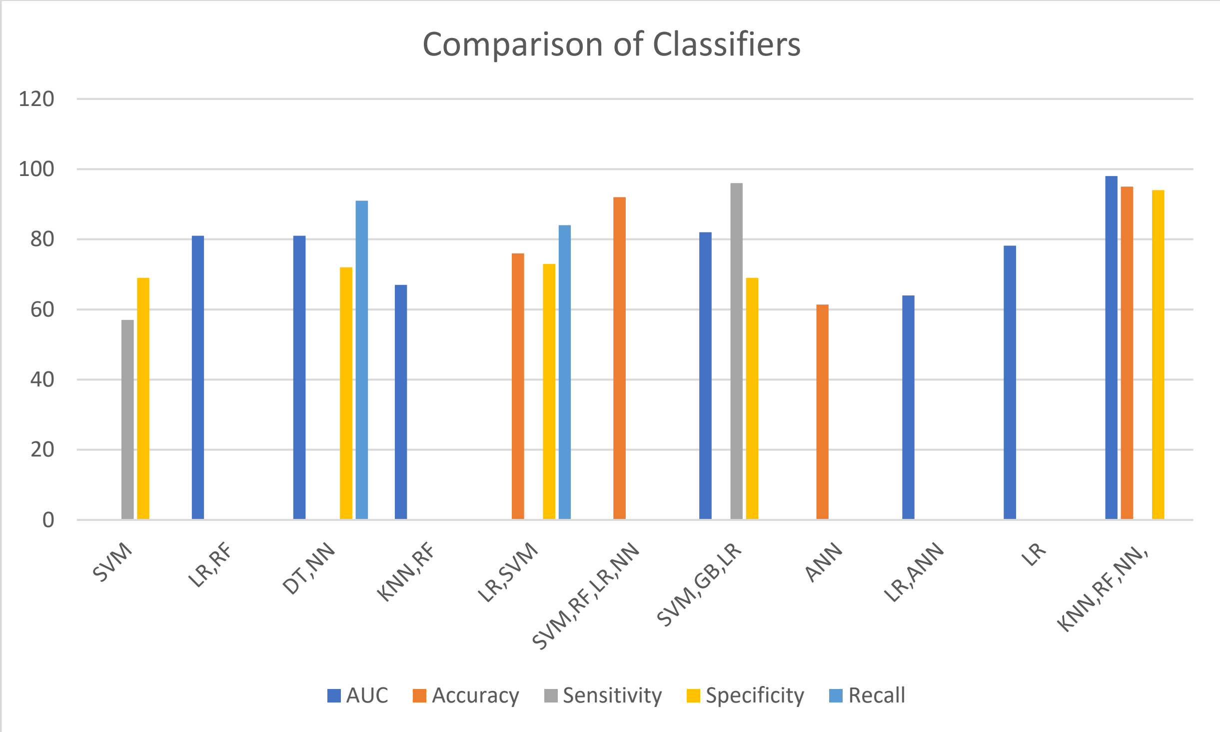

Fig. 5 Comparison of classifiers

3.5. Transvaginal Ultrasound

Ultrasound waves are used to visualize and diagnose a variety of pathologies during a non-invasive vaginal ultrasound imaging exam. It entails the doctor inserting a probe into the vaginal canal to envision better and analyze the reproductive organ. The authors proposed segmenting the cervix using a convolutional neural network (UNet). Retrieve two physical and biological ultrasound markers from the resulting images: anterior cervical angle (ACA) and cervical length (CL). He employs classic machine learning classifiers such as SVM and Naive Bayes to classify the results. The Naive Bayes algorithm yields the best results, with an AUC of 78.13 percent, an accuracy of 7.5%, recall of 74%, and precision of 85% [16]. The cervix is segmented and classified simultaneously on transvaginal ultrasound images. The Y-Net network, a convolutional neural network, is used in the algorithm. Our method outperformed advanced methods in predicting preterm birth from transvaginal ultrasound images. Shows the outcome Specificity = 0.97, recall = 0.68 [66].

After cerclage in the first 26 weeks of pregnancy, a decision support system predicts the duration of spontaneous delivery in a high-risk group. Assist physicians in defining the management timeline and reducing neonatal complications. The dataset's highly imbalanced class distribution problem was solved using SMOTE. After that, four classification models were used to create the prediction model: Decision Tree, K-Nearest Neighbours (K-NN), Neural Network (NN), and Random Forest. The Random Forest algorithms produced excellent outcomes in terms of G-mean and sensitivity, with 0.96 G-mean and 1.00 sensitivity values [67]. Following Table 3 compares classifiers using the TVS dataset, and Fig 6 gives the chart comparison of the dataset.

Table 3. Comparison of Classifiers in TVS Dataset

Authors	Techniques	Gestation week	Results
Wodarczyk et al. (2019)	deep neural network, Support Vector		Recall = 0.74, Precision = 0.85, AUC = 0.78, Accuracy = 0.78,
Wodarczyk et al., (2020)	U-Net, Fully Convolutional Network, and Deeplabv3	Between 23 – 42 weeks	Specificity= 0.97, Recall = 0.68,
Rawashdeh et al., (2020)	Decision Tree, Random Forest, KNN, NN, linear regression, Random Forest, K-star Gaussian process and LWL	<26 weeks	AUC = 0.98, Recall=1.0, Specificity = 0.94, Accuracy = 0.95,

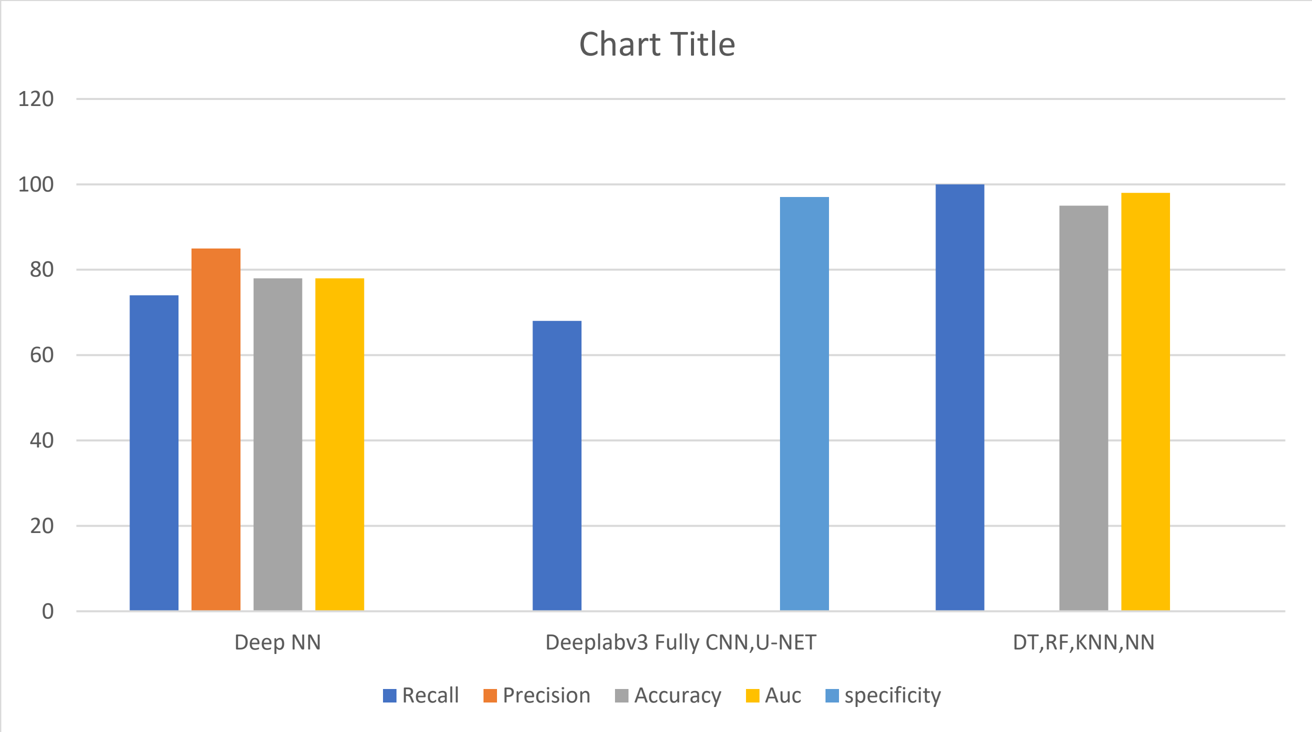

Fig. 6 Comparison of classifiers

Another study used layer-wise importance transmission to predict preterm birth using 3D convolutional NN. Backpropagation algorithm layer-wise relevance propagation allocates similarity scores per each input voxel in this case. The results of 157 magnetic resonance scans of newborn babies between the ages of 23 and 42 weeks were used in this study. Therefore 94 % is accuracy, 100 % is a true positive rate, and 86% is a true negative rate [68]. Together within the 2.09 window, 95 percent IC (2.090–2.097) predict the week of delivery. Health professionals can be alerted by addressing PTB risk factors. Elastic Net, Extreme Gradient Boosting, Linear Regression, Decision TreeRidge Regression, Quantile Ordinal Regression - LASSO have compared algorithms. The best method for determining gestation week and PTB can use EXtreme Gradient Boosting [69]. Below Fig 7 shows the preterm birth rate between 2017 to 2020.

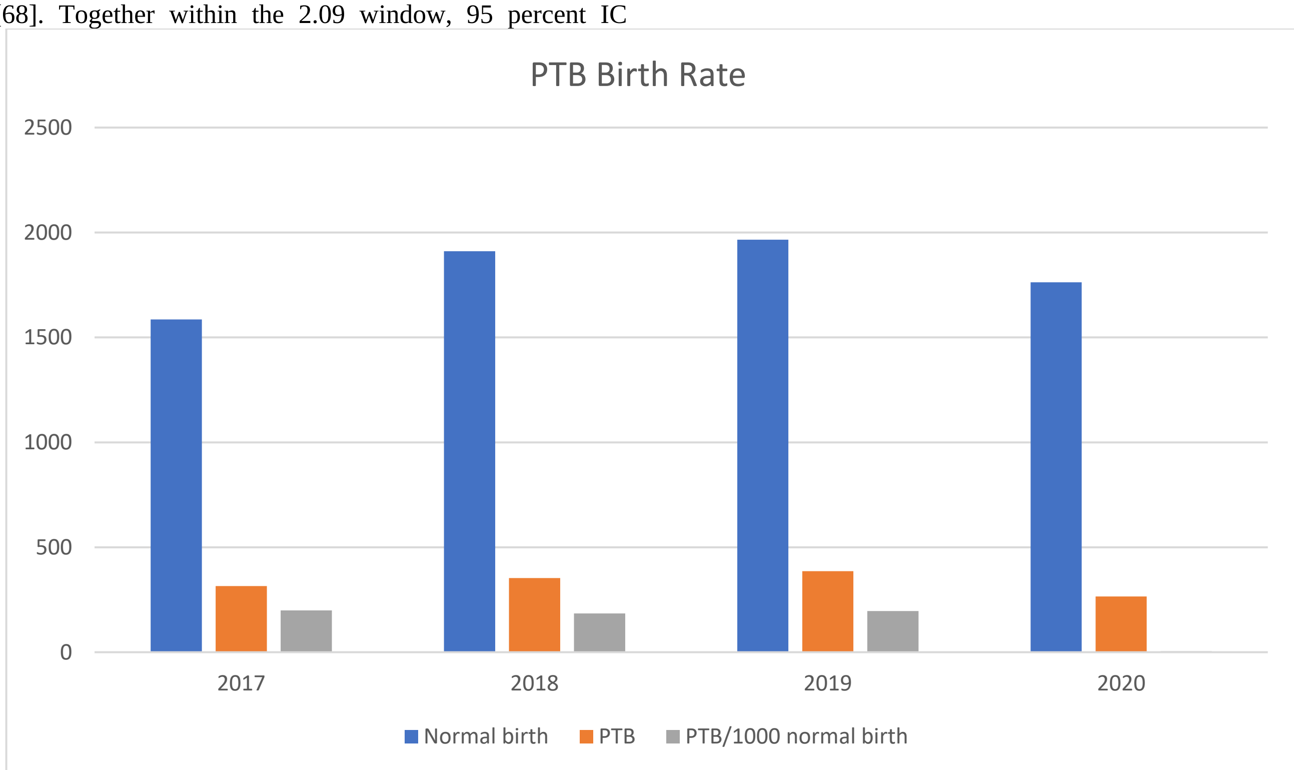

Fig 7. PTB Birth Rate

4. Conclusion

A review of the current implementation of machine learning methods in perinatal medicine is done in this paper. Finally, interpretable ML applications are related to real data and results. The output is objective and helps identify the most key parameters for doctors. Continue to support ML exploration in this area to solve problems that could be used to reduce perinatal complications in a variety of health care settings. The usefulness of the ML model can also be improved by enlarging the data size and optimizing the data type. Centralized procedures for handling incomplete data, imbalance control, and specific instance groups also were necessary to accomplish more accurate and reliable results. According to the reviewed studies, machine learning methods can help optimize premature birth detection capability and provide information that can help recognize women with PTB. Finally, develop a dependable, objective method for evaluating labor and intervening earlier.

References

[1] World Health Organization. Preterm birth. (2015), URL http://www.who.int/ media centre/factsheets/fs363/en/.

[2] Cost effects of Preterm birth: a Comparison of healthcare Costs associated with early preterm, late preterm, and full-term birth in the first 3years after birth J Jacob, MLehne, A Mischker, N Klinger. – The European Journal (2017)-Springer.

[3] Behrman, R. E., & Butler, A. S. Preterm birth: Causes, Consequences, and Prevention. (2007).

[4] Hintz, S. R., Newman, J. E., &Vohra, B. R. Changing Definitions of Long-term follow-up: should “long term” be even longer? In Seminars in perinatology . WB Saunders. 40 (6) (2016) 398-409.

[5] World Health Organisation. (1977). WHO: Recommended Definitions, Terminology, and Format for Statistical Tables related to the Perinatal Period and use of a new Certificate for cause of Perinatal deaths? Modifications Recommended by FIGO as amended Acta Obstet Gynecol Scand, 56(3), (1976) 247-53.

[6] Radford, S. K., Costa, F. D. S., Júnior, E. A., & Sheehan, P. M. Clinical Application of Quantitative Fetal Fibronectin for Predicting Preterm birth in Symptomatic Women. Gynecologic and Obstetric Investigation, 83(3) (2018) 285-289.

[7] Blencowe, H., Cousens, S., Chou, D., Oestergaard, M., Say, L., Moller, A. B., & Lawn,J. Born too soon: the Global Epidemiology of 15 Million Preterm births. Reproductive Health, 10(1) (2013) 1-14.

[8] Berghella, V., Palacio, M., Ness, A., Alfirevic, Z., Nicolaides, K. H., &Saccone, G. Cervical Length Screening for Prevention of Preterm birth in a Singleton Pregnancy with Threatened Preterm Labor: Systematic Review and Meta-Analysis of Randomized Controlled Trials using Individual Patient-level data. Ultrasound in Obstetrics & Gynecology, 49(3) (2017) 322-329.

[9] Mercer BM, Goldenberg RL, Das A, Moawad AH, Iams JD, Meis PJ, et al. The preterm prediction study: a clinical risk assessment system. Am J Obstet Gynecol 1996 Jun;174(6):1885-1895?

[10] Surendiran,R., and Duraisamy, K., An Approach in Semantic Web Information Retrieval. SSRG International Journal of Electronics and Communication Engineering,1(1), ISSN:2348-8549, (2014) 17-21. https://doi.org/10.14445/23488549/IJECE-V1I1P105

[11] Shahid N, Rappon T, Berta W. Applications of Artificial Neural Networks in Healthcare Organizational Decision-Making: a scoping review. PLoS One (2019).

[12] Fatima M, Pasha M. Survey of Machine Learning Algorithms for Disease Diagnostic. J Intell Learn SystAppl 09 (2017) 1–16.

[13] Koivu A, Sairanen M. Predicting Risk of Stillbirth and Preterm Pregnancies with Machine Learning. Health Inf Sci Syst 8(1) (2020) 14.

[14] Safi Z, Venugopal N, Ali H, Makhlouf M, Boughorbel S. Analysis of Risk Factors Progression of Preterm Delivery using Electronic Healthrecords. (2020).

[15] Sufriyana H, Husnayain A, Chen YL, Kuo CY, Singh O, Yeh TY, et al. Comparison of Multivariable logistic Regression and Other Machine Learning Algorithms for Predictive Prediction Studies in Pregnancy care: Systematic Review and Meta-analysis. JMIR Med Inform; e16503. 8(11) (2020) 17.

[16] Włodarczyk, T.; Płotka, S.; Trzci ´nski, T.; Rokita, P.; Sochacki-Wójcicka, N.; Lipa, M.; Wójcicki, J. Estimation of Preterm Birth Markers with U-Net Segmentation Network. In Smart Ultrasound Imaging and Perinatal, Preterm and Paediatric Image Analysis; Wang, Q., Gomez, A., Hutter, J., McLeod, K., Zimmer, V., Zettinig, O., Licandro, R., Robinson, E., Christiaens, D., Turk, E.A., et al., Eds.; Springer: Cham, Switzerland, (2019) 95103.

[17] S. Saigal and L. W. Doyle, “An Overview of Mortality and Sequelae of Preterm birth From infancy to adulthood,” 8e Lancet, 371 (9608) (2008) 261–269,

[18] 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) IEEE, Coimbatore, India. (2017) 394–399.

[19] Wainberg, M., Merico, D., Delong, A. et al. Deep Learning in Biomedicine. Nat Biotechnol 36(2018) 829–838. https://doi.org/10.1038/nbt.4233

[20] Ngiam K. Y., Khor I. W. Big Data and Machine Learning Algorithms for Healthcare Delivery. The Lancet Oncology. 20(5) (2019)e262–e273. doi: 10.1016/s1470- 2045(19)30149 4.

[21] Rajkumar A., Dean J., Kohane I. Machine learning in Medicine. New England Journal of Medicine. 380(14) (2019)1347–1358. doi: 10.1056/nejmra1814259

[22] Witten I. H., Frank E. Data mining. ACM Sigmod Record. 31(1) (2002) 76–77. doi: 10.1145/507338.507355.

[23] Dhillon A., Singh A. Machine Learning in Healthcare Data Analysis: a Survey. Journal of Biology and Today’s World. 8(6) (2019)1–10.

[24] Ahmad M. A., Eckert C., Teredesai A. Interpretable Machine Learning in Healthcare. Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics; Washington, DC, USA. (2018)559–560.

[25] Surendiran,R., Thangamani,M., Narmatha,C., Iswarya,M., Effective Autism Spectrum Disorder Prediction to Improve the Clinical Traits using Machine Learning Techniques, International Journal of Engineering Trends and Technology (IJETT),., ISSN: 2231-5381, 70(4) (2022) 343-359. https://doi.org/10.14445/22315381/IJETT-V70I4P230

[26] Kotsiantis, S.B. Decision trees: a recent overview. Artif. Intell. Rev. 39 (2013) 261–283

[27] Breiman, L. Random forests. Mach. Learn. 45 (2001) 5–32

[28] Chen, T., and Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining – KDD SIGKDD. 16 (2016) 785–794.
[29] Hsu, C.-W. and Lin, C.-J. A Comparison of Methods for Multiclass Support Vector Machines. IEEE Trans. Neural Netw. 13 (2002) 415–425
[30] Altman, N.S. An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. Am. Stat. 46 (1992) 175–185.
[31] Friedman, N. et al. Bayesian network classifiers. Mach. Learn. 29 (1997) 131–163.
[32] Miotto, R. et al. Deep Learning for Healthcare: Review, Opportunities, and Challenges. Brief. Bioinform. 19 (2018) 1236–1246
[33] Eraslan, G. et al. Deep Learning: new Computational Modeling Techniques for Genomics. Nat. Rev. Genet. 20 (2019) 389–403
[34] Cheerla, A. and Gevaert, O. Deep learning with Multimodal Representation for Pan-Cancer Prognosis Prediction. Bioinformatics 35 (2019) 446–454
[35] Nieto-del-Amor, F.; Prats-Boluda, G.; Martinez-De-Juan, J.L.; Diaz-Martinez, A.; Monfort-Ortiz, R.; Diago-Almela, V.J.; Ye-Lin, Y. Optimized Feature Subset Selection Using Genetic Algorithm for Preterm Labor Prediction based on Electrogastrography. Sensors 21 (2021) 3350.
[36] Peng, J.; Hao, D.; Yang, L.; Dua, M.; Song, X.; Jiang, H.; Zhang, Y.; Zheng, D. Evaluation of Electrogastrogram Measured from Different Gestational Weeks for Recognizing Preterm Delivery: A Preliminary Study Using Random Forest. Biocybern. Biomed. Eng. 40 (2020) 352–362.
[37] Fergus, P.; Idowu, I.; Hussain, A.; Dobbins, C. Advanced Artificial Neural Network Classification for Detecting Preterm Births using EHG records. Neurocomputing 188 (2016) 42–49.
[38] Fergus, P.; Cheung, P.; Hussain, A.; Al-Jumeily, D.; Dobbins, C.; Iram, S. Prediction of Preterm Deliveries from EHG Signals using Machine Learning. PLoS ONE 8 (2013)e77154.
[39] Despotovic, D.; Zec, A.; Mladenovic, K.; Radin, N.; Turukalo, T.L. A Machine Learning Approach for an Early Prediction of Preterm Delivery. In Proceedings of the 2018 IEEE 16th International Symposium on Intelligent Systems and Informatics (SISY), Subotica, Serbia, 13 (15) (2018) 265–270.
[40] Acharya, U.R.; Sudarshan, VK; Rong, S.Q.; Tan, Z.; Min, L.C.; Koh, J.E.; Nayak, S.; Bhandary, S.V. Automated Detection of Premature Delivery using Empirical Model and Wavelet Packet Decomposition Techniques with Uterine Electromyogram Signals. Comput. Biol. Med. 85 (2017) 33–42.
[41] Khan, M.U.; Sajid, Z.; Sohail, M.; Aziz, S.; Ibraheem, S.; Naavi, S.Z.H. Electrohysterogram based Term, and Preterm Delivery Classification System. In Proceedings of the 2020 First International Conference of Smart Systems and Emerging Technologies (SMARTTECH), Riyadh, Saudi Arabia, 3(5) (2020) 83–88.
[42] Degbedzui, D.K.; Yüksel, M.E. Accurate Diagnosis of the Term–Preterm births by Spectral Analysis of Electrohysterography signals. Comput. Biol. Med. 119 (2020) 103677.
[43] Ren, P.; Yao, S.; Li, J.; Valdes-Sosa, P.A.; Kendrick, K.M. Improved Prediction of Preterm Delivery using Empirical Mode Decomposition Analysis of Uterine Electromyography Signals.PLoS ONE, 10 (2015) e0132116.
[44] Sadi-Ahmed, N.; Kacha, B.; Taleb, H.; Kedir-Talha, M. Relevant Features Selection for Automatic Prediction of Preterm Deliveries from Pregnancy Electro Hysterograhic (EHG) records. J. Med. Syst. 41 (2017) 1–13.
[45] Hangxiao Lou, Haifeng Liu, Zhenqin Chen, Zi'ang Zhen, Bo Dong, Jinshan Xu, Bioprocess Inspired Characterization of Pregnancy Evolution using Entropy and its Application in Preterm birth Detection, Biomedical Signal Processing, and Control, Volume 75(2022)103587, ISSN 1746-8094.
[46] M. Shahbakhti, M. Beiramvand, M. R. Bavi, and S. Mohammadi Far, A New Efficient Algorithm for Prediction of Preterm Labor, (2019) 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), (2019) 4669-4672, doi: 10.1109/EMBC.2019.8857837.
[47] Mohammadi Far S, Beiramvand M, Shahbakhti M, Augustyniak P. Prediction of Preterm Delivery from Unbalanced EHG Database. Sensors. 22(4) (2022) 1507. https://doi.org/10.3390/s22041507
[48] Prats-Boluda G, Pastor-Tronch J, Garcia-Casado J, Monfort-Ortíz R, Perales Marín A, Diago V, Roca Prats A, Ye-Lin Y. Optimization of Imminent Labor Prediction Systems in Women with Threatened Preterm Labor Based on Electrohysterography. Sensors. 21(7) (2021) 2496. https://doi.org/10.3390/s21072496.
[49] Zhang, L., Li, H., Li, J. et al. Prediction of Iatrogenic Preterm birth in Patients with Scarred Uterus: a Retrospective Cohort Study in Northeast China. BMC Pregnancy Childbirth 20 (2020) 490 https://doi.org/10.1186/s12884-020-03165-7
[50] Shaniba Asmi P., Kamalraj Subramaniam, Nisheena V. Iqbal, A review of significant Research on Predicting Preterm Birth using Uterine Electromyogram Signal, Future Generation Computer Systems, ISSN 0167-739X, 98(2019)135-143 https://doi.org/10.1016/j.future.2018.10.033.
[51] Vovsha, I.; Rajan, A.; Salleb-Aouissi, A.; Raja, A.; Radeva, A.; Diab, H.; Tomar, A.; Wapner, R. Predicting Preterm Birth is not Elusive: Machine Learning Paves the way to Individual Wellness. In Proceedings of the 2014 AAAI Spring Symposium Series, Palo Alto, CA, USA, 24 (26)(2014) 82–89.
[52] Tran, T.; Luo, W.; Phung, D.; Morris, J.; Rickard, K.; Venkatesh, S. Preterm birth Prediction: Stable Selection of Interpretable Rules from High Dimensional data. In Proceedings of the 1st Machine Learning for Healthcare Conference, Los Angeles, CA, USA, 19(20) (2016)164–177.
[53] Esty, A.; Frize, M.; Gilchrist, J.; Bariciak, E. Applying Data Preprocessing Methods to Predict Premature Birth. In Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 18 (21)(2018) 6096–6099.
[54] Weber A, Darmstadt GL, Gruber S, Foeller ME, Carmichael SL, Stevenson DK, et al. Application of Machine-Learning to Predict early Spontaneous Preterm birth Among Nulliparous non-Hispanic Black and White Women. Ann Epidemiol 28(11) (2018) 783-9.e1. [doi: 10.1016/j.annepidem.2018.08.008] [Medline: 30236415]
[55] M. W. L. Moreira, J. J. P. C. Rodrigues, G. A. B. Marcondes, A. J. V. Neto, N. Kumar AndI. De La Torre Diez, A Preterm Birth Risk Prediction System for Mobile Health Applications Based on the Support Vector Machine Algorithm, 2018 IEEE International Conference on Communications (ICC), (2018)1-5, doi: 10.1109/ICC.2018.8422616.

[56] Prema, NS; Pushpalatha, M.P. Machine Learning Approach for Preterm Birth Prediction Based on Maternal Chronic Conditions. In Lecture Notes in Electrical Engineering; Sridhar, V., Padma, M., Rao, K., Eds; Springer: Singapore, (2019) 581–588.

[57] Lee, K.S.; Ahn, KH Artificial Neural Network Analysis of Spontaneous Preterm Labor and Birth and Its Major Determinants. J. Korean Med. Sci. 34 (2019).

[58] Gao C, Osmundson S, Velez Edwards DR, Jackson GP, Malin BA, Chen Y. Deep Learning Predicts Extreme Preterm Birth from Electronic Health Records. J Biomed Inform. 100:103334. doi: 10.1016/j.jbi.2019.103334. Epub 2019 Oct 31. PMID: 31678588; PMCID: PMC6899197. (2019)

[59] Puspitasari, Diah, Ramanda, Kresna, Supriyatna, Adi, Wahyudi, Mochamad, Delima, Sikumbang, Erma, Hadisukmana, sulaeman, Comparison of Data Mining Algorithms UsingArtificial Neural Networks (ANN) and Naive Bayes for Preterm Birth Prediction, Journal of Physics: Conference Series, Volume 1641, International Conference on Advanced Information Scientific Development (ICAISD) (2020)

[60] Arabi Belaghi, R.Beyene, J.&McDonald, S.D.Clinical Risk Models for Preterm birth less than 28 weeks and less than 32 weeks of Gestation using a Large Retrospective Cohort. J Perinatol 41 (2021) 2173-2181.

[61] Hasan Rawashdeh, ShathaAwawdeh, Fatima Shannag, EsraaHenawi, Hossam Faris, Nadim Obeid, Jon Hyett, Intelligent System Based on Data Mining Techniques for Prediction of Preterm Birth for Women with Cervical Cerclage, Computational Biology, and Chemistry, 85(2020)107233 ISSN 1476-9271, https://doi.org/10.1016/j.compbiolchem.2020.107233.

[62] Enio Luis Damaso, Daniel Lober Rolnik, Ricardo de Carvalho Cavalli, Silvana Maria Quintana, Geraldo Duarte, Fabricio da Silva Costa, Alessandra Marcolin, Prediction of Preterm Birth by Maternal Characteristics and Medical History in the Brazilian Population, Journal of Pregnancy, Article ID 4395217, 6 pages, (2019). https://doi.org/10.1155/2019/4395217

[63] Boughorbel, S., Jarray, F., Venugopal, N., &Elhadi, H. Alternating loss correction for preterm-birth prediction from ehr data with noisy labels. arXiv preprint arXiv:1811.09782. (2018).

[64] Barbara Lizewska, Joanna Teul, Pawel Kuc, Adam Lemancewicz, Karol Charkiewicz, Joanna Goscik, Marian Kacerovsky, Ramkumar Menon, Wojciech Miltyk, Piotr Laudanski, "Maternal Plasma Metabolomic Profiles in Spontaneous Preterm Birth: Preliminary Results, Mediators of Inflammation, (2018), Article ID 9362820, 13 pages, (2018) https://doi.org/10.1155/2018/9362820

[65] Díaz, E., Fernández-Plaza, C., Abad, I. et al. Machine Learning as a Tool to Study the Influence of Chronodisruption in Preterm Births. J Ambient Intell Human Comput 13(2022) 381–392 https://doi.org/10.1007/s12652-021-02906-6

[66] Włodarczyk, T.; Płotka, S.; Rokita, P.; Sochacki-Wójcicka, N.; Wójcicki, J.; Lipa, M.; Trzci ´nski, T. Spontaneous Preterm Birth Prediction Using Convolutional Neural Networks. In Medical Ultrasound, and Preterm, Perinatal, and Paediatric Image Analysis; Hu, Y., Licandro, R., Noble, J.A., Hutter, J., Aylward, S., Melbourne, A., Turk, E.A., Barrena, JT, Eds.; Springer: Cham, Switzerland, (2020)274–283.

[67] Rawashdeh, H.; Awawdeh, S.; Shannag, F.; Henawi, E.; Faris, H.; Obeid, N.; Hyett, J. Intelligent System Based on Data Mining Techniques for Prediction of Preterm Birth for Women with Cervical Cerclage. Comput. Biol. Chem. 85 (2020) 107233.

[68] Grigorescu, I.; Cordero-Grande, L.; Edwards, A.D.; Hajnal, J.; Modat, M.; Deprez, M. Interpretable convolutional neural Networks for Preterm Birth Classification. arXiv, arXiv:1910.00071. (2019)

[69] T.A.H. Rocha, EBAF de Thomaz, DG de Almeida, et al. Data-driven Risk Stratification for Preterm Birth in Brazil: A Population-Based Study to Develop of a Machine Learning Risk Assessment Approach, The Lancet Regional Health - Americas 3 (2021) 100053.

[70] Baer RJ, McLemore MR, Adler N, Oltman SP, Chambers BD, Kuppermann M, Pantell MS, Rogers EE, Ryckman KK, Sirota M, Rand L, Jelliffe-Pawlowski LL. Pre-Pregnancy or First-Trimester Risk Scoring to Identify Women at High Risk of Preterm Birth. Eur J Obstet Gynecol Reprod Biol. 231(2018) 23540.

[71] Zou, H. and Hastie, T. Regularization and variable selection via the elastic net. JR Stat. Soc. Ser. B Stat Methodol. 67 (2005)301–320