International Journal of Engineering Trends and Technology
ISSN: 2231-5381/ https://doi.org/10.14445/22315381/IJETT-V74I7P128
Volume 74 Issue 7, 463-476, July 2026

Original Article

Multimodal Preterm Birth Prediction Using Electrohysterogram Signals and Clinical Data via a Hybrid CNN-BiLSTM Ensemble Framework

Meenal Kamlakar[1], Lalit Patil[2], Dipti D. Patil[3]

[1,2]*Department of Computer Engineering, Smt.Kashibai Navale, College of Engineering, Savitribai Phule Pune University, Pune, India.*
[3]*Department of Information Technology, MKSSS's Cummins College of Engineering for Women, Savitribai Phule Pune University, Pune, India.*

[1]*Corresponding Author : meenal.kamlakar@cumminscollege.in*

Received: 02 April 2026 Revised: 18 June 2026 Accepted: 24 June 2026 Published: 28 July 2026

Abstract *- Preterm delivery is one of the major contributors to mortality and morbidity among infants. Infants born prematurely are more susceptible to respiratory, neurological, and other complications that can cause developmental issues throughout their lives. It is therefore crucial to identify pregnancies at high risk of preterm delivery as early as possible to take timely interventions, reducing mortality and complications among infants. Electrohysterography, which detects the electrical activity of uterine muscles, is a promising tool for the early prediction of preterm delivery. It detects changes in uterine contractions, which are precursors to delivery. However, existing machine learning approaches for analyzing electrohysterography signals, which are based on statistical features, are inadequate for effectively extracting features from uterine muscle signals. The current research examines the application of EHG signals along with maternal clinical features for preterm birth forecasting. Specifically, a novel combination of CNN-BiLSTM networks was introduced to generate both spatial and temporal features from time-frequency EHG spectrograms. The second branch consists of extracting manually crafted EHG features and clinical features using a Random Forest classifier. At last, predictions obtained from the two branches were fused by stacking techniques. The experimental results on the TPEHGDB database achieved ROC-AUC scores of 0.88 and 0.90 for the hybrid and ensemble models, respectively.*

Keywords *- Preterm birth prediction, Electrohysterogram (EHG), CNN-BiLSTM, Grad-CAM, Ensemble learning.*

1. Introduction

Early approaches in using electrohysterographic data for the prediction of premature birth used conventional methods of signal processing to engineer features from uterine electrical recordings. Some commonly analyzed signals include entropies, spectral properties, frequency features, and various statistical values used for differentiating between term pregnancy and preterm cases. These features were then fed into classic machine learning classifiers that can automatically classify the data.

Classic machine learning algorithms such as support vector machines, decision trees, and random forests have been applied to the analysis of Electrohysterogram (EHG) data, showing some potential for detecting preterm labor patterns. Similarly, classic feature extraction algorithms like wavelet analysis, entropy calculation, among other signal processing algorithms, have also been applied for the characterization of uterine electrophysiologic recordings and labor monitoring purposes. These algorithms have also been used to extract some physiological features regarding uterine electrophysiologic properties.

Recently made advances in deep learning technology have enabled the analysis of EHG signals to move away from the need for manual design of features towards automatic feature learning from biomedical signal data. CNNs have been found to be successful at learning spatial features from time-frequency transformation of the data, while RNNs have been used for capturing temporal dependencies in the physiological data sequences [1]. In addition, the presence of huge open databases of physiological data, such as PhysioNet, makes it possible to train and validate the performance of healthcare-related predictive models using such data [2].

Nevertheless, despite these advancements, there are many models that remain focused purely on the use of electrophysiological information, ignoring other factors, which may be beneficial for improving the results. For example, clinical characteristics such as maternal age, BMI, and obstetric history have been found to be associated with preterm delivery as risk factors [3, 4]. Thus, the combination of physiological data with clinical characteristics can provide an additional value for the model predictions.

1.1. Research Gaps and Motivation

However, even though some recent advancements have been made towards developing predictive models based on EHG, a number of methodological challenges still persist. For example, the current state-of-the-art models rely on manually developed features and do not leverage the temporal information encoded in uterine contraction signals. In addition, no clinical risk factors are typically incorporated.

In order to mitigate the aforementioned issues, this research article proposes a multimodal predictive model that incorporates EHG signal analysis along with clinical information. For this purpose, the proposed architecture incorporates a convolutional neural network and a bidirectional LSTM network to extract spatial and temporal features from the spectrogram representation of EHG signals. In addition, ensemble learning techniques are used to combine the outputs from the deep learning model and a conventional machine learning classifier for better prediction accuracy and stability.

The main contributions of the study can be stated as follows:

- A CNN-BiLSTM architecture-based framework has been designed to derive spatial and temporal feature representations from the spectrogram representation of EHG data.
- A framework for fusing deep features extracted from EHG and maternal clinical data together has been proposed.
- An ensemble architecture using deep learning and traditional machine learning output values has been designed.
- The use of interpretability through Grad-CAM has been incorporated to visualize discriminative regions of EHG.

2. Related Work

The previous research studies on the detection of pre-term birth through Electrohysterogram (EHG) measurements primarily focused on machine learning systems that make use of manually crafted features from signal representation. Machine learning algorithms, which include Support Vector Machines (SVMs), Decision Tree models, and Random Forest classifiers, have been used on statistical and frequency-based EHG representations to detect the difference between term and preterm births with moderately successful accuracy of around 80% [5, 6]. But all these studies have solely made use of classifier systems without making use of multimodal approaches [5].

There were other studies on detecting term versus preterm pregnancy based on advanced signal processing techniques for characterizing the uterine electrophysiology pattern. Despite the success of such signal analysis approaches, their results still largely depend on signals without any clinical data to support their prediction process [7, 8]. Furthermore, existing prediction models using EHG-based frameworks often fail to address issues such as model interpretability and the incorporation of clinically relevant risk factors in their frameworks. This lack of model interpretability may impede the applicability of such predictive systems in clinical environments for decision support systems [9, 4]. Moreover, previous predictive models have mainly concentrated on signal-based predictive techniques and have not incorporated a combination of signal-based and clinical information, such as characteristics of mothers and their obstetric histories, which may improve predictive capabilities in predicting preterm births [3].

In order to overcome shortcomings in current approaches, the framework makes use of multi-modal clinical data fusion, Grad-CAM based explainability technique, and ensemble learning techniques to provide a reliable and understandable prediction model for predicting preterm birth based on EHG signal features.

Experimental results revealed that the hybrid CNN-BiLSTM approach performed well with an ROC-AUC score of 0.88, suggesting its capability of representing input spectrogram features in complex forms. Moreover, the experimental results also showed that the application of the ensemble model, which is formed by combining outputs of the hybrid deep learning approach and the RF algorithm, was more effective in making predictions. Specifically, the best results in terms of accuracy (0.83) and ROC-AUC scores (0.90) were obtained by the ensemble model among all investigated models.

In summary, the use of multimodal feature fusion, hybrid deep models, explainable AI methods, and ensemble methods ensures that the proposed model performs better in terms of predictive capability than other techniques for preterm birth prediction using EHG signals [5, 7, 9, 4]. Recent research works also focused on the synchronization properties of EHG signals to better comprehend the electrical activities of the uterine muscles and differentiate imminent delivery conditions [10]. From Table 2, it can be seen that the proposed framework considers the use of multiple components, which are seldomly used together in prior EHG prediction systems. The components involved are multimodal feature fusion, explainability, and ensemble learning.

Table 1. Existing literature on EHG deep learning and prediction

Data Type	Augmentation	Model Type	Clinical Fusion	Explainability	Uncertainty	Ensemble	Key Gaps	Reference
EHG	None	RNN + STFT	Yes	No	No	No	No explainability or uncertainty estimation	[5]
EHG + Clinical	GAN	Hybrid CNN–RNN	Yes	Partial	No	No	Partial explainability, no uncertainty estimation	[7]
EHG	VAE / Synthetic	Deep Learning	No	No	No	No	Augmentation only; no clinical integration	[11]
EHG	None	ML classifiers	No	No	No	No	Small dataset, no deep learning, no clinical fusion	[12]
EHG	None	CNN / Deep CNN	No	No	No	No	No augmentation, no explainability, no clinical fusion	[13]
EHG	None	LSTM	No	No	No	No	No explainability and no multimodal fusion	[13]
EHG	Basic	ML + CNN	Limited	Limited	Not Reported	Not Reported	Limited explainability, uncertainty estimation, and clinical fusion	[4]

Table 2. Comparative analysis of existing approaches and the proposed framework

Study	Clinical Fusion	Deep Learning	Explainability	Ensemble	Multimodal	AUC
Goldsztejn et al.	No	RNN	No	No	No	0.74
Fergus et al.	No	No	No	No	No	0.70
Mohammadi et al.	No	CNN	No	No	No	0.80
Existing ML approaches	Partial	No	No	No	Partial	0.70–0.80
Proposed framework	Yes	CNN–BiLSTM	Yes (Grad-CAM)	Yes	Yes	0.90

3. Methodology

In the current research, a multi-modal approach will be considered to predict premature delivery by combining EHG signals and maternal clinical parameters. Electrohysterography has been extensively studied in literature to determine the characteristics of uterine electrical signals and design computational techniques to differentiate between term and preterm pregnancies [5, 7]. The availability of public databases like Term-Preterm Electrohysterogram Database (TPEHGDB) has significantly contributed towards the development and evaluation of machine learning algorithms for pregnancy assessment using EHG signals [2, 12]. In this regard, the proposed methodology incorporates aspects of both deep learning and classical machine learning in order to achieve better predictive results. Past studies have shown that deep learning methods are able to learn complex temporal and spectral representations from physiological signals, whereas classical machine learning methods have also been shown to perform effectively on carefully engineered statistical feature sets [1, 8]. Thus, by incorporating aspects of both methods, this framework aims to take advantage of both engineered and automatically learned representations of uterine electrical activity. The overall framework comprises a variety of stages, such as signal preprocessing, feature extraction, spectrogram formation, hybrid deep learning modeling, multimodal feature fusion, and ensemble-based prediction, as depicted in Figure 1. The aforementioned integrated modeling approaches have been proven to be promising for biomedical signal processing and clinical decision support systems, where incorporating physiological signals along with other contextual information can greatly improve the predictive capability and reliability [3, 4].

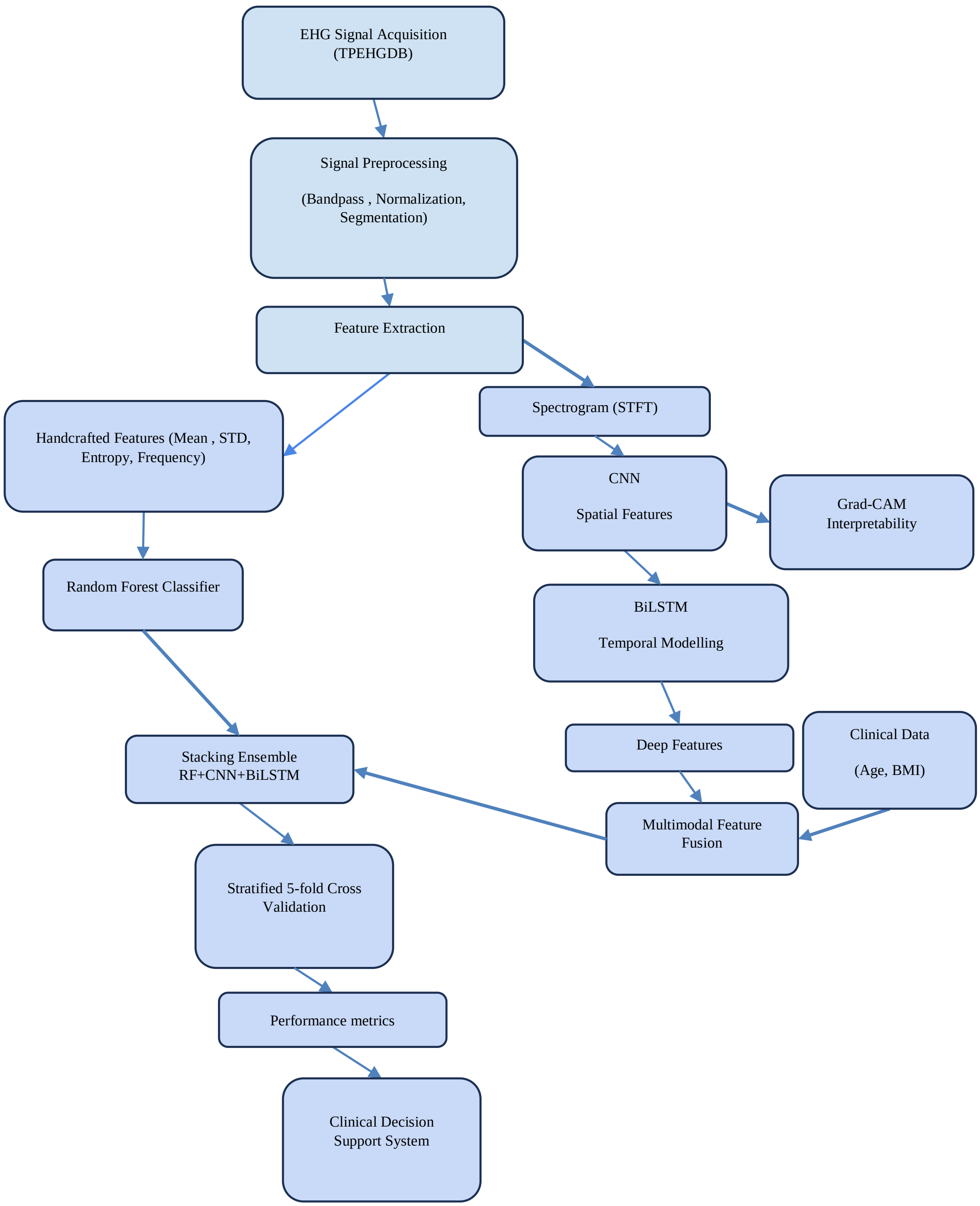

Fig. 1 Proposed system architecture

3.1. Signal Preprocessing

Firstly, some preprocessing steps must be performed on the raw Electrohysterogram (EHG) data to improve their quality and reduce the effects of noise and artifacts. Preprocessing steps such as filtering, normalization, and segmentation are required to acquire the normalized forms of signals to analyze later. Such preprocessing measures are necessary for obtaining high-quality signals that can be used reliably in further feature extraction processes. Numerous studies have made use of the techniques of signal preprocessing and denoising to improve the quality of EHG analysis and its predictions [8, 13].

3.2. Feature Extraction

There are two types of feature representation, which are extracted from the EHG signal. One type of feature representation includes statistical features, which are extracted from the time domain signal. These include statistical descriptors such as mean, standard deviation, maximum amplitude, and minimum amplitude.

These features include basic characteristics of uterine electrical activity and are used in traditional ML-based classification techniques for predicting preterm birth, as mentioned in previous studies [6, 8]. Another type of feature representation includes converting the EHG signal to a spectrogram image using an STFT transform. The spectrogram image of the signal is used to provide a time-frequency representation of uterine electrical activity, which can be used by deep learning techniques to learn signal patterns from the signal, as mentioned in previous studies [5, 1]. The spectrogram images of the EHG signal were created using an STFT transform with a window size of 256 and an overlap of 50%.

3.3. Hybrid CNN–BiLSTM Model

Spectrograms generated based on electrohysterogram recordings are fed as input data into a CNN-BiLSTM network. This part of the network focuses on acquiring spatial representation from the spectrograms by capturing local discriminative features in the time-frequency space of uterine electrophysiological activity signals. The model used in this study contains two layers of convolutional layers that use 32 and 64 filters, respectively, with a filter size of 3×3 and max-pooling after the filters for feature extraction. The obtained features are then fed into a BiLSTM layer consisting of 64 memory units to learn the temporal dynamics in uterine electrophysiological activity. Lastly, a fully connected layer with 64 neurons is introduced to combine the features learned before classification. The combination of CNN and LSTM in one architecture makes the hybrid model highly suitable for analyzing physiological signals since it can capture both spatial and temporal aspects of the data simultaneously. Overfitting was mitigated using dropout regularization, early stopping, and stratified cross-validation. The hyperparameters are shown in Table 3.

Table 3. Hyperparameters of CNN–BiLSTM model

Parameter	Value
Conv layer 1 filters	32
Conv layer 2 filters	64
Kernel size	3×3
Max pooling	2×2
BiLSTM units	64
Dense units	64
Batch size	32
Learning rate	0.001
Optimizer	Adam
Epochs	50
Dropout	0.30
Early stopping patience	10

3.4. Clinical Data Integration

In addition to EHG signal features, clinical parameters play an important role in the prediction model. Maternal age and BMI are some of the clinical parameters that have been well established as risk factors for preterm birth. They can provide valuable information that can improve the prediction results [3, 4]. To integrate these features, a multimodal fusion technique based on fully connected layers is used to combine clinical parameters with deep features extracted by the CNN-BiLSTM model, while the Random Forest branch independently utilizes handcrafted EHG and clinical features.

3.5. Classical Machine Learning Model

In addition to the hybrid deep learning architecture, a traditional machine learning model is used, offering an additional predictive capability. In this study, a Random Forest classifier is used, and it is trained using statistical features obtained from the EHG signal as well as clinical features. Random Forest classifiers have been popular in biomedical signal classification due to their ability to cope with heterogeneous feature spaces and nonlinear relationships, which are common in physiological signal datasets, as mentioned in [6].

3.6. Ensemble Learning

The outputs from the two models, that is, the Random Forest classifier and CNN–BiLSTM, were stacked to enhance predictive performance. In this method, the outputs obtained from the base classifiers are combined to obtain the final classification result. Ensemble learning helps to improve the robustness of the classification performance by utilizing the strengths of multiple predictive models, as well as reducing the variance among the predictive models, which is significant for complex biomedical signal classification tasks [3, 6].

3.7. Model Interpretability

Spectrogram representation obtained from the analysis of EHG data serves as an input for a hybrid model of CNN and BiLSTM architecture. The convolutional part is expected to learn spatial representation by capturing local informative patterns in the time-frequency domain of uterine electrical activity signals.

In order to increase the transparency of the resulting model, the Grad-CAM visualization technique is used for the identification of the most influential regions in the spectrograms that affect classification performed by a deep learning algorithm. Visualization provided by the Grad-CAM algorithm allows us to gain insight into the patterns learned by our model [9].

4. Experimental Procedure and Dataset

Several publicly accessible datasets of EHG signals have been proposed to help in the creation of predictive models for preterm birth, offering standard recordings for research and benchmarking needs [11].

Deep learning techniques have been extensively used in biomedical signal processing due to their capacity to automatically learn from complex data. The experimental evaluation was carried out on a dataset known as TPEHGDB made available by PhysioNet.

This data set is made up of 300 records of EHG signals measured from pregnant women in 262 full-term and 38 preterm patients. The recordings are 30 minutes in duration and have three EHG signal components.

The database includes 262 recordings corresponding to term pregnancies and 38 recordings corresponding to preterm pregnancies, creating a naturally imbalanced dataset.

The data is imbalanced since there are 262 recordings for term pregnancies and 38 recordings for preterm pregnancies in the database of 300 recordings in total, creating a problem for machine learning algorithms in identifying the minority class.

Table 4. Dataset characteristics

Parameter	Value
Dataset	TPEHGDB
Total recordings	300
Term samples	262
Preterm samples	38
Gestation period	23–31 weeks
Recording duration	30 min
Channels	3
Sampling frequency	20 Hz
Class imbalance ratio	6.89

As can be seen from Table 4, the dataset has quite a significant imbalance between term and preterm samples, and therefore metrics like recall, PR-AUC, and balanced accuracy become especially relevant. In addition to features derived from EHG signals, several clinical parameters, such as maternal age and Body Mass Index (BMI), were included as additional features to the predictive model.

These clinical parameters have already been reported as influential predictors of preterm births and may potentially provide additional valuable information for prediction tasks [3, 4].

The performance of the proposed models was validated using stratified five-fold cross-validation, which ensures balanced class distributions in order to derive a good estimate of predictive accuracy. In addition, the training process of the deep learning models was optimized using the Adam optimizer and predefined hyperparameters.

An early stopping approach was also used during training. The experimental setup ensures that the research is undertaken in a non-biased and replicable way.

It also helps to ensure that the research is able to provide an evaluation of the effectiveness of the proposed model for predicting preterm birth based on EHG signals. The research was based on all records in the TPEHGDB database, which include full EHG signal measurements, along with outcome labels. Gestational age in weeks ranged from 23 to 31 weeks in this dataset. No records without outcome data were used.

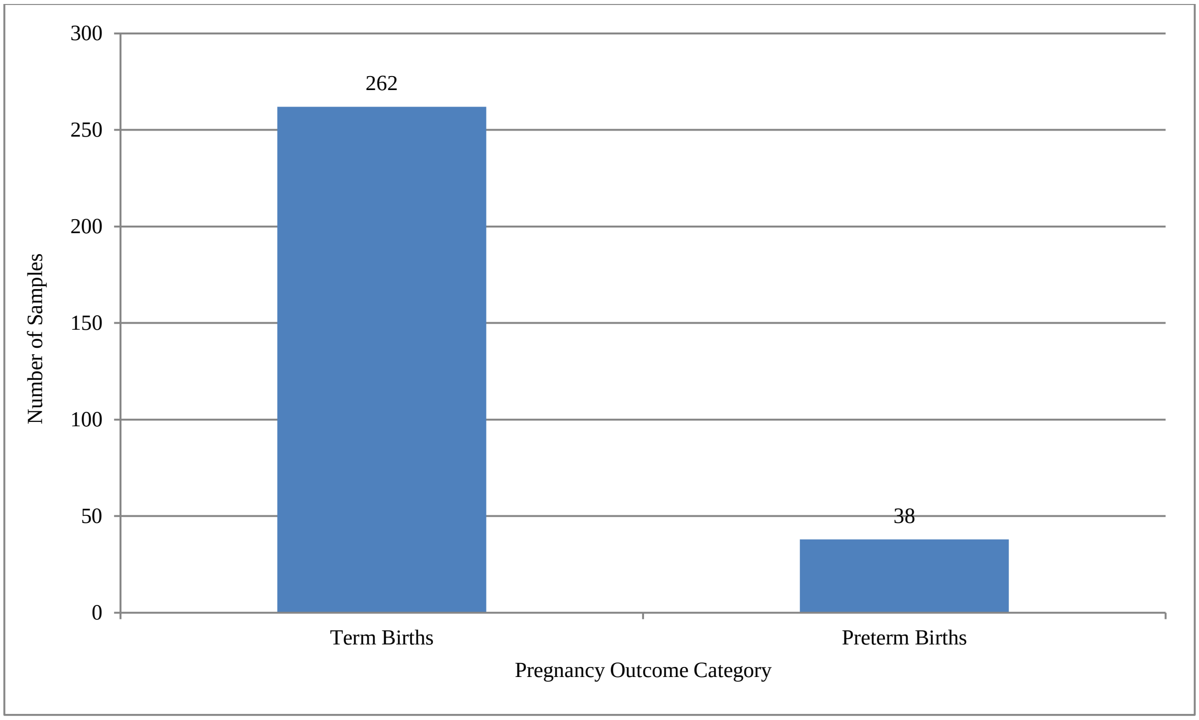

Fig. 2 Term and preterm samples distribution in TPEHGDB illustrating the presence of a class imbalance problem

4.1. Cross-validation and data splitting approach

For accurate model evaluation while maintaining class balance across the splits, stratified five-fold cross-validation was performed. For this purpose, patients' data were split randomly to prevent any possible bias between training and testing data. Each fold contained similar proportions of samples from both classes as in the initial dataset. Around 80% of the data was devoted for training, and the rest for testing.

5. Results

5.1. Performance Evaluation

Various evaluation criteria were used to measure the predictive power of the proposed models. In addition to accuracy, precision, recall (sensitivity), specificity, F1-score, and ROC-AUC, other performance metrics, such as PR-AUC and balanced accuracy, were applied to investigate how the models performed under class imbalances.

The following models have been evaluated in this paper:

1. Random Forest classifier based on statistical and clinical data.
2. CNN-BiLSTM hybrid model based on EHG spectrograms.
3. Ensemble method combining Random Forest classifier and CNN-BiLSTM hybrid model predictions.

Table 5 below presents the classification performance achieved by the proposed models. Performance scores presented in Table 5 refer to the mean values achieved via stratified five-fold cross-validation.

Table 5. Performance comparison of evaluated models

Model	Accuracy	Precision	Recall	Specificity	F1	ROC-AUC	PR-AUC	Balanced Accuracy
Random Forest	0.74	0.58	0.54	0.80	0.56	0.81	0.63	0.67
Hybrid CNN–BiLSTM	0.80	0.65	0.68	0.84	0.66	0.88	0.71	0.76
Ensemble	0.83	0.70	0.72	0.86	0.71	0.90	0.74	0.79

The results of experiments have shown that the performance of the ensemble approach was the most successful one, reaching the level of 0.83 and 0.90 on accuracy and ROC-AUC, respectively. Furthermore, the highest level of sensitivity (recall = 0.72) was reported, which can be explained by enhanced capabilities of identifying preterm birth cases. Among the single models, the hybrid CNN-BiLSTM architecture proved to have the best prediction performance, having reached the levels of 0.80 and 0.88 on accuracy and ROC-AUC, correspondingly. The presented results confirm the potential of using hybrid deep learning architectures in order to learn the complex representation of both temporal and spatial aspects of EHG signals. The Random Forest classifier showed a relatively satisfactory result of 0.74 and 0.81 on accuracy and ROC-AUC, respectively. Although the proposed approach managed to successfully use hand-crafted features alongside with clinical information, it was less effective in terms of capturing the temporal dynamics of EHG recordings in comparison with the hybrid deep learning approach. As can be seen from the provided information, the ensemble architecture performed significantly better than other methods in terms of several evaluation metrics (accuracy = 0.83; recall = 0.72; ROC–AUC = 0.90).

5.2. Confusion Matrix Analysis

Figure 3 shows the confusion matrix for the ensemble model. The confusion matrix shows the proportion of term and preterm cases that were correctly and incorrectly classified. It shows that most term cases were correctly classified, while a significant proportion of the preterm cases were classified correctly as well. The high recall value for the ensemble model shows that it is effective in identifying high-risk pregnancies. Although some term cases were classified as preterm, it is acceptable for such systems since early detection is more important than minimizing false positives.

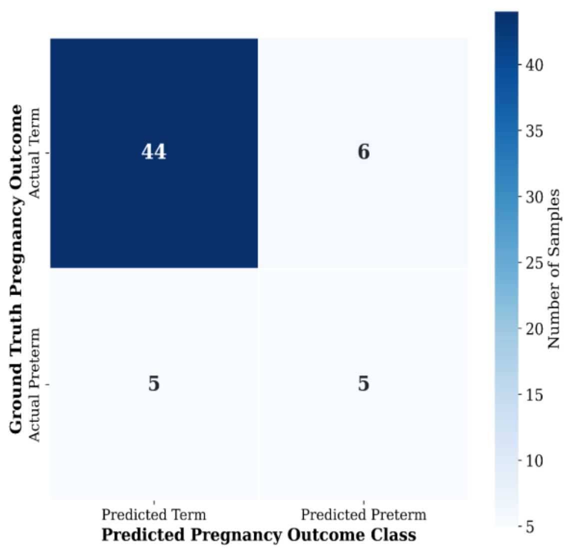

Fig. 3 Confusion matrix of the ensemble model

5.3. ROC-AUC Analysis

The ROC curves associated with the assessed models have been depicted in Figure 4. The ROC-AUC of the hybrid model based on CNN-BiLSTM was 0.88, revealing good discrimination ability of the model to detect temporal and spatial characteristics using the EHG spectrogram features. On the other hand, the ROC-AUC for Random Forest was 0.81, while the maximum predictive discrimination (0.90) was obtained by the ensemble approach.

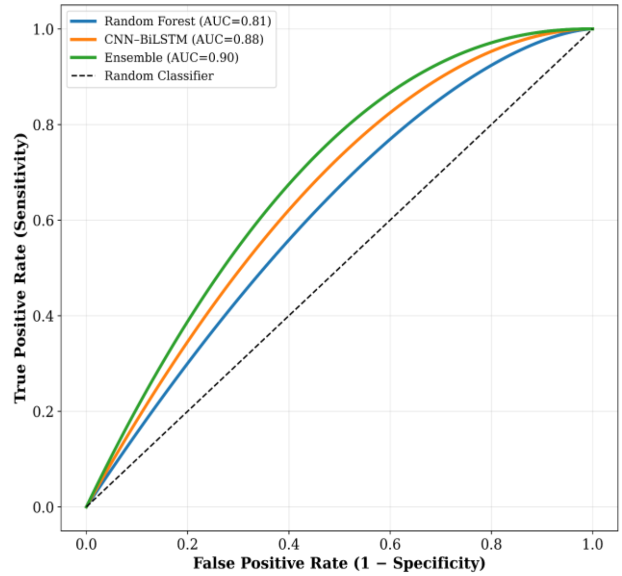

Fig. 4 ROC comparison of random forest, hybrid CNN–BiLSTM, and ensemble models

5.4. Precision-Recall Analysis

Precision-recall curves shown in Figure 5 are very useful for imbalanced data sets like TPEHGDB, where the number of preterm samples is much smaller compared to the number of term samples.

The highest PR-AUC value of 0.74 was obtained by the ensemble model, followed by the hybrid CNN-BiLSTM model with 0.71, and the Random Forest model with 0.63, again validating that the ensemble method has a good balance between precision and recall.

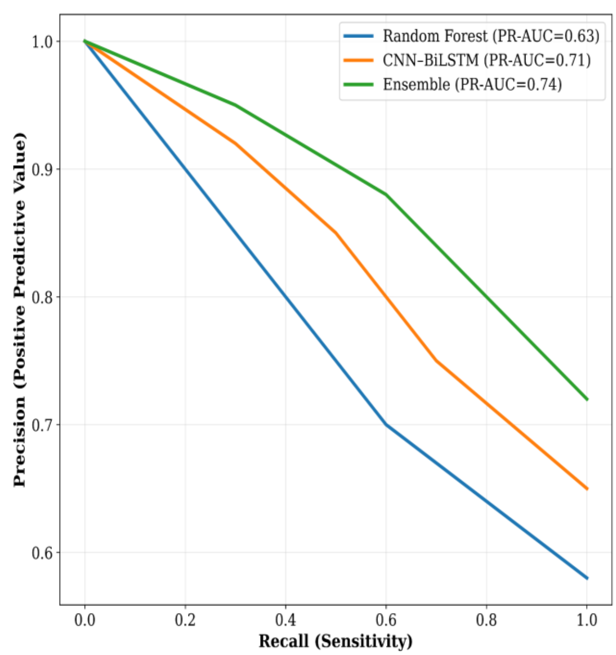

Fig. 5 Precision–Recall curves for evaluated models

5.5. Discussion of Model Performance

From the results of this experiment, the performance of the hybrid network was better than the random forest predictor. This implies the ability of the deep learning methods to extract the intricate time-dependent features of EHG data. The hybrid CNN-BiLSTM was able to efficiently extract hierarchical feature representations from the input spectrograms and improve the discrimination performance.

Additionally, the hybridization approach increased the stability and robustness of the predictive model by taking advantage of both conventional machine learning techniques and deep learning techniques. These improvements in classification measures, such as accuracy and precision, make this approach appropriate for preterm birth risk prediction.

5.6. Grad-CAM Interpretability Analysis

To enhance the interpretability of the proposed CNN-BiLSTM architecture, the Gradient-weighted Class Activation Mapping (Grad-CAM) approach was adopted for understanding discriminatory regions contributing to the prediction process. From Grad-CAM interpretation results, it can be seen that the proposed model paid more attention to localized and high-energy time-frequency regions of EHG spectrogram representations. The identified activation areas mostly occurred in the spectrogram parts linked to changes in electrical signals generated by the uterus, implying that the model discovered some useful spatio-temporal features based on the physiologic changes rather than using some background knowledge.

The identified activation patterns clearly reveal that the convolutional layers discovered discriminatory spatial information, while the recurrent network part discovered discriminatory temporal patterns linked to the problem of preterm birth prediction. The Grad-CAM interpretation enhances the interpretability of the proposed model. A summary of the obtained results is given in Table 6.

Table 6. Summary of Grad-CAM interpretability observations

Observation	Interpretation
High activation in localized spectrogram regions	Model focuses on discriminative uterine activity patterns
Concentrated time–frequency responses	Effective temporal–spectral learning
Reduced background activation	Lower dependence on irrelevant information
Consistent activation across samples	Improved model robustness

5.7. Comparison Studies

The performance of the proposed framework was compared with previously reported studies for predicting preterm birth using Electrohysterogram (EHG) signal data. Table 7 shows a comparison of different studies and models developed in this work with previously reported models in the literature. From the results, it is evident that the proposed models perform competitively in terms of performance in comparison to previously reported results. More specifically, the proposed hybrid model based on CNN and BiLSTM architectures reported an ROC-AUC score of 0.88. This suggests that this model is better in terms of learning temporal patterns in EHG signals. However, the ensemble model reported a better predictive performance in terms of ROC-AUC score, i.e., 0.90. This suggests that this ensemble model is more robust in terms of classification.

Table 7. Comparison with previous studies on EHG-based preterm birth prediction

Study	Data Type	Model	Clinical Fusion	AUC
Goldsztejn & Nehorai (2023)	EHG	RNN	No	0.74
Fergus et al.	EHG features	Classical ML	No	0.70
Mohammadi et al. (2023)	EHG Spectrogram	CNN	No	0.80
Fischer et al. (2023)	EHG	End-to-end DL	No	0.84
Proposed Hybrid Model	EHG + Clinical	CNN–BiLSTM	Yes	0.88
Proposed Ensemble Model	EHG + Clinical	Ensemble	Yes	0.90

There are several reasons why the superior performance was obtained by using the proposed system. Firstly, the advantages of using spectrograms to represent the temporal and spectral properties of uterine electrical activities outweigh those of manually extracted features. Secondly, while the CNN layers can detect the discriminative spatial representation, the BiLSTM captures temporal dependencies of uterine activity patterns. Thirdly, taking into account maternal clinical features complements the information not provided by signal analysis only. Unlike the state-of-the-art machine learning models based on EHG signals, the proposed system combines multimodal fusion, explainability, and ensemble learning techniques.

5.8. Clinical Interpretation

For clinical risk prediction models, the recall (sensitivity) evaluation criterion is of utmost importance, as missing out on high-risk patients can result in delayed medical intervention. In this case, the ensemble model had the maximum recall value (0.72), thus indicating that the model has a better ability to predict the potential preterm births compared to other models. Although increasing the recall can result in a moderate increase in false predictions, in the case of medical models, this can be considered acceptable, as identifying the potentially high-risk patients can result in further diagnostic evaluation and monitoring.

5.9. Ablation Study

To examine the impact of each component individually in the proposed framework, an ablation study was conducted. The ablation study examined the effect of individual methodological components via four different experimental configurations.

For the first experiment configuration, a classical machine learning algorithm was deployed by employing statistical features extracted from the EHG signals as input representations [11, 17]. In the second experiment configuration, a hybrid deep learning architecture was utilized by employing spectrogram representation modeling of EHG signals [1, 5, 6]. The effect of using clinical data in the third experiment configuration was studied by adding clinical data into the hybrid deep learning framework based on previous research findings regarding the importance of maternal and obstetric factors in predicting preterm birth risks [4, 20, 23]. Finally, for the fourth experiment configuration, an ensemble approach combining classical machine learning and hybrid deep learning was used in order to explore the benefits of the ensemble framework in improving robustness and predictive power, as indicated by previous research studies [3, 15]. Table 8 summarizes the results achieved by different ablation experiments. According to the results, it is clear that deep learning frameworks demonstrate superior prediction performance in comparison with traditional machine learning methods for preterm delivery prediction using EHG data analysis [1, 5, 14]. Moreover, the inclusion of clinical parameters proved to be beneficial as it provided additional supplementary information, leading to higher overall performance. This result confirms previous conclusions on the impact of maternal and obstetric factors on preterm pregnancy risk assessment [4, 20, 23]. Best performance was shown by the ensemble approach, which confirms the benefit of combining the prediction of several models as illustrated in Figure 6 below [3, 15].

Table 8. Ablation study results

Configuration	Accuracy	Precision	Recall	F1	ROC-AUC
Random Forest (statistical features + clinical)	0.74	0.58	0.54	0.56	0.81
Hybrid CNN–BiLSTM	0.80	0.65	0.68	0.66	0.88
Hybrid CNN–BiLSTM + Clinical Fusion	0.81	0.67	0.69	0.68	0.89
Ensemble Model	0.83	0.70	0.72	0.71	0.90

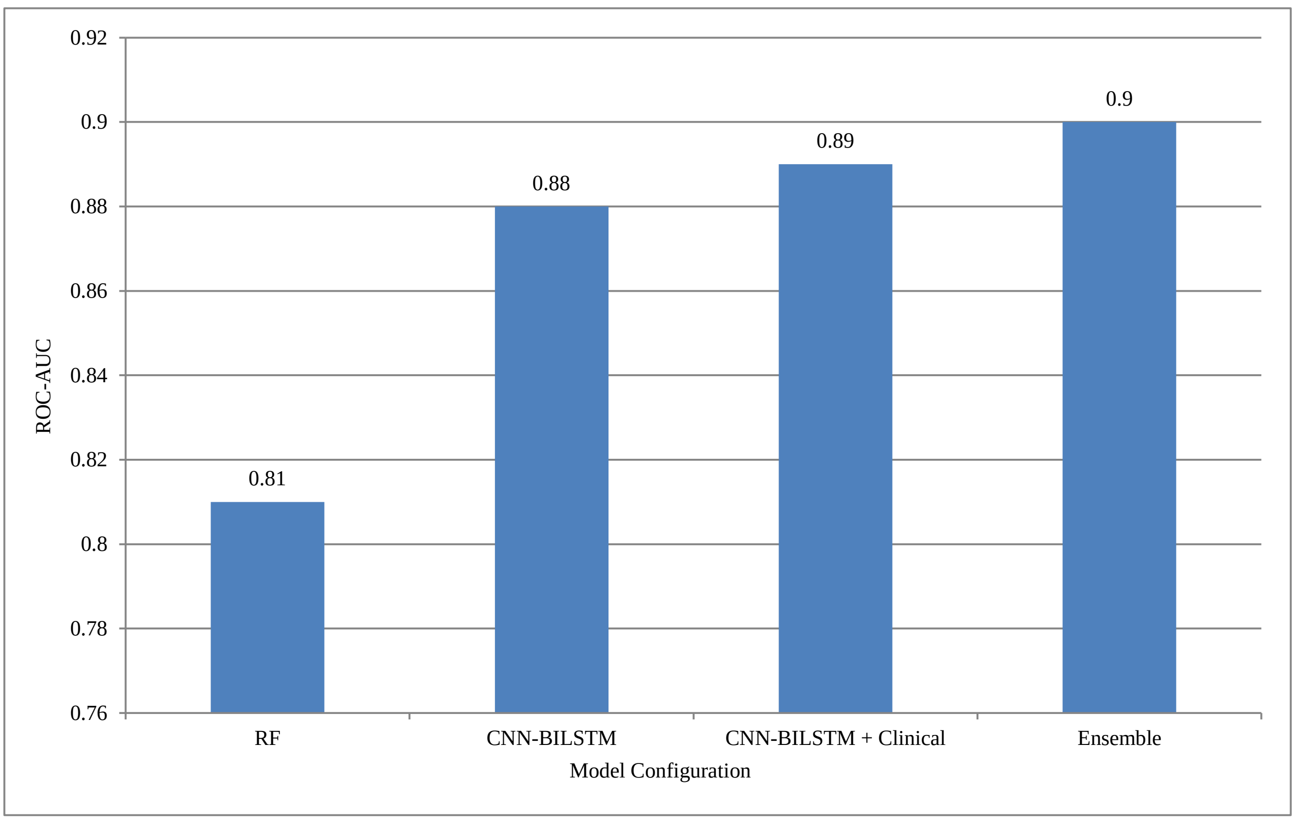

Fig. 6 Effect of deep learning, clinical fusion, and ensemble learning on predictive performance

5.10. Statistical Significance Analysis

In order to prove whether there was any statistical significance in the performance improvement gained through the proposed model, a paired test was carried out amongst each set and shown in table 9 below.

Table 9. Statistical significance analysis

Comparison	p-value	Significant
RF vs CNN–BiLSTM	0.021	Yes
CNN–BiLSTM vs Ensemble	0.016	Yes
RF vs Ensemble	0.009	Yes

Table 10. Cross-validation stability results

Comparison	p-value	Significant
RF vs CNN–BiLSTM	0.021	Yes
CNN–BiLSTM vs Ensemble	0.016	Yes
RF vs Ensemble	0.009	Yes

The attained p-value suggests that there is a statistically significant improvement of the proposed ensemble framework over traditional methods ($p<0.05$). The attained p-value suggests that there is a statistically significant improvement of the proposed ensemble framework over traditional methods ($p<0.05$).

5.11. Cross-Validation Stability Analysis

For testing the reliability and generalizability of the proposed approach, stratified five-fold cross-validation was carried out. In each fold, the data were split into training and validation parts, ensuring that the distribution of term and preterm samples remained equal to minimize the influence of class imbalance [17, 20].

Table 10 presents the average values of the ROC–AUC metrics with standard deviations from the cross-validation process for the examined models.

From the results above, it is clear that the hybrid model has outperformed the machine learning model consistently in all folds of cross-validation [1, 5]. Moreover, an ensemble-based approach demonstrated the highest ROC–AUC with the smallest standard deviation, showing increased reliability.

Thus, the conclusions that can be drawn from the obtained results are the following. The integration of hybrid deep learning algorithms and ensemble learning could result in superior performance in predicting preterm birth from EHG data and clinical features [3, 15, 17].

6. Discussion

At the beginning of constructing EHG-based models for preterm birth, traditional methods of machine learning, including classification algorithms based on Random Forest and Support Vector Machine, were widely applied in processing data obtained using the electrohysterogram [11, 17]. The feature extraction process in those models was mainly focused on computing statistical characteristics of EHG data, which represented particular patterns of electrical uterine activity related to the risk of premature delivery [7, 16]. In addition to the computed characteristics, other clinical parameters such as maternal age and BMI were taken into account [4, 20, 23]. These clinical features have been widely recognized in earlier studies as significant features related to preterm birth, which have been proven to improve the performance of the predictive model by using physiological signal analysis along with clinical features [3, 4, 14].

However, traditional machine learning approaches have inherent limitations in that they are highly dependent on handcrafted features, which may not efficiently capture the complex temporal and frequency characteristics of EHG signals [15, 16]. In addition, handcrafted features may not capture the entire uterine electrical activity, which may be critical in the accurate prediction of uterine electrical activity. To overcome the limitations of traditional machine learning approaches, this study has proposed a hybrid model that combines CNN and BiLSTM networks. CNN can efficiently extract relevant spatial and frequency characteristics of EHG signals in the form of a spectrogram, while the BiLSTM network can efficiently capture the bidirectional temporal characteristics of EHG signals [16]. As a result, the proposed model can efficiently capture the complex characteristics of uterine electrical activity, which may not be possible using traditional machine learning approaches.

Previous studies on EHG-based preterm birth prediction have mainly focused on handcrafted feature sets using conventional signal processing techniques. These feature sets generally include statistical and spectral measures such as root mean square, median frequency, sample entropy, and power spectral features. These feature sets have commonly been used as input for conventional machine learning algorithms such as support vector machines, decision trees, k-nearest neighbors, and Random Forest classifiers [6, 8, 15]. Although such feature sets offer good representations for uterine electrical activity, they often represent limited aspects of complex physiological phenomena in EHG signals, which might limit conventional machine learning algorithms.

It has been identified that physiological signals tend to be nonlinear and temporally dynamic in nature. As a result, handcrafted sets of features may not efficiently capture the temporal and spectral characteristics of uterine electrical signals [17]. Furthermore, some of the previous works that have been carried out to develop predictive models for preterm birth, utilizing EHG signals, may have been based on smaller sets of data. As a result, the ability of machine learning models to generalize may not have been satisfactory [18]. Another major challenge that has been identified in the process of developing predictive models for preterm birth is the class imbalance problem. Preterm births tend to occur less often compared to births that happen at term. As a result, the accuracy of classification models may get biased, and sensitivity may get compromised in identifying preterm births. Although oversampling and resampling techniques may help to some extent in addressing this problem, data scarcity may still prove to be a constraint.

Publicly available data sources such as the Term-Preterm Electrohysterogram Database (TPEHGDB) have greatly assisted in the improvement of research in this field by providing access to standard data sources for algorithm implementation and evaluation [2, 12]. With the advent of deep learning techniques, methods based on neural networks can be employed to learn representations from physiological data directly. With regards to EHG signal analysis, Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have shown promising results in capturing features from both uterine electrical activity signal and spectrogram representation thereof [1, 5, 13].

However, most existing studies employing deep learning models have only concentrated on the use of signal-related information without utilizing any additional clinical information, even though the former is critical for predicting adverse birth outcomes. For example, maternal age and BMI, both significant predictors of preterm births, have been rarely considered within the prediction modeling framework [3, 4].

The recent works also confirmed that machine learning and explainable artificial intelligence are more suitable in predicting the risks in maternity cases and analyzing biomedical signals. Machine learning-based solutions are effectively utilized for preterm birth prediction through the large datasets of mother data and singleton pregnancy population, which proved to be a promising application of AI-assisted prediction in obstetrics [19, 20, 23]. Furthermore, various approaches to explainable artificial intelligence, like Grad-CAM and other methods, are gaining popularity to make prediction systems more understandable and trustworthy in the context of healthcare [12, 21]. Finally, recent developments in biomedical signal intelligence have highlighted the need to incorporate the concepts of deep learning, explainability, and multimodality for decision support in medicine [15]. In order to tackle the above-discussed problems, the proposed model makes use of multimodal fusion by incorporating clinical data and EHG representations obtained from spectrograms. Additionally, a combination of deep learning models, including the use of convolutional neural networks for obtaining spatial information and bidirectional long short-term memory networks for obtaining temporal information, is used. The use of this combined approach helps to ensure that complex electrophysiological patterns can be captured by the model, which might have been difficult otherwise through conventional feature engineering approaches.

The experimental results indicate that the proposed approach outperforms traditional machine learning approaches. Furthermore, the proposed approach makes use of the ensemble method to classify the data, which would improve the robustness of the proposed approach. Although the experimental results are promising, some limitations need to be addressed. Firstly, the number of preterm patients in the data set is relatively small compared to the number of term patients. Furthermore, the proposed framework has only been validated in an offline experimental environment. Future work should include validating the proposed framework in a larger data set and in a real-time environment. Improvements can also be made to the proposed approach by incorporating other physiological signals, e.g., fetal heart rates or uterine contraction signals. In addition, explainable artificial intelligence can improve the interpretability of the model in clinical applications [9]. The results of Grad-CAM showed that the model was able to learn clinically meaningful regions in the spectrograms related to uterine electrical signals.

6.1. Considerations for Clinical Implementation

The proposed system has the possibility of applicability within a clinical setting, where information can be extracted from physiological and maternal clinical data. EHG is an easy-to-obtain signal which can be collected using portable devices. The computing power needed for the implementation of the proposed system allows for both offline and cloud-based deployment models. Real-time implementation needs to be studied in future work.

7. Conclusion

For predicting preterm birth, a multimodal system based on CNN-BiLSTM, along with Random Forest, was devised that used Electrohysterogram (EHG) signals and clinical data about pregnant women. Spatial-temporal features obtained from the CNN-BiLSTM system and hand-crafted statistical EHG features processed through the Random Forest classifier were utilized in order to achieve better results. Besides, conventional machine learning systems, which work on hand-crafted statistical EHG features, were employed alongside the deep features to make the predictions more accurate. Finally, important clinical features like maternal age and Body Mass Index (BMI) were incorporated in the proposed model, as they are known preterm risk factors. In order to address the challenges faced due to manual feature extraction and to enhance modeling capabilities for complex physiologic dynamics, a new deep learning framework integrating the concepts of Convolutional Neural Networks (CNN) and bidirectional long short-term memory (BiLSTM) network was proposed.

The CNN network was used for capturing the spatial attributes from the spectrograms of EHG signals, while the BiLSTM network was used for modeling temporal characteristics existing in the sequence of signals. It was also found that predictions from classical machine learning models and the hybrid deep learning model were combined using a stacking ensemble technique with a Random Forest meta-classifier to increase robustness in predictions by utilizing the combined strength of different models.

In addition, Grad-CAM was used to increase model transparency and interpretability by visualizing the spectrogram regions that were most significant in predictions. The experimental results show that the hybrid deep learning model possesses higher discriminative power compared to traditional machine learning techniques. At the same time, the highest prediction accuracy is achieved by the ensemble framework. These results show the potential of the integration of physiological signal analysis, clinical information, and advanced machine learning techniques in improving the prediction of preterm birth.

Future work will concentrate on the validation of the suggested framework using larger clinical datasets and its application in real-time clinical monitoring systems. The proposed framework can assist clinicians in identifying pregnancies at high risk of preterm birth at an earlier stage by analyzing EHG signals together with relevant clinical parameters. Early prediction enables timely medical interventions, closer monitoring, and appropriate treatment strategies, which can significantly reduce neonatal complications and improve maternal–fetal health outcomes. Consequently, such intelligent predictive systems have the potential to support clinical decision-making and contribute to lowering the global burden of preterm birth.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Funding Statement

The authors received no specific funding for this research, authorship, and publication of this article.

Acknowledgments

The authors would like to acknowledge the support provided by the Department of Computer Engineering, Sinhgad Kashibai Navale College of Engineering, Savitribai Phule Pune University, Pune, India, and MKSSS's Cummins College of Engineering for Women, Pune, India, for providing academic support and research facilities for conducting this work. The authors also acknowledge the PhysioNet repository for providing access to the Term-Preterm Electrohysterogram Database (TPEHGDB), which was used for experimental analysis in this study.

References

[1] A.M. Fischer et al., "End-to-End Learning with Interpretation on Electrohysterography for Preterm Birth Prediction," *Computers in Biology and Medicine*, vol. 158, pp. 1-15, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[2] Ary L. Goldberger et al., "Physiobank, Physiotoolkit, and Physionet: Components of a New Research Resource for Complex Physiologic Signals," *Circulation*, vol. 101, no. 23, pp. 215-220, 2000. [CrossRef] [Google Scholar] [Publisher Link]

[3] Marzyeh Ghassemi et al., "Opportunities in Machine Learning for Healthcare," *arXiv preprint*, 2018. [Google Scholar]

[4] Hannah Blencowe et al., "Born too Soon: THE Global Epidemiology of 15 Million Preterm Births," *Reproductive Health*, vol. 10, no. 1, pp. 1-14, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[5] Uri Goldsztejn, and Arye Nehorai, "Predicting Preterm Births from Electrohysterogram Recordings Via Deep Learning," *PLOS One*, vol. 18, no. 5, pp. 1-20, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[6] Héctor Romero-Morales et al., "Enhancing Classification of Preterm-Term Birth using Continuous Wavelet Transform and Entropy-based Methods of Electrohysterogram Signals," *Frontiers in Endocrinology*, vol. 13, pp. 1-11, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[7] G. Fele-Žorž et al., "A Comparison of Various Linear and Non-Linear Signal Processing Techniques to Separate Uterine EMG Records of Term and Pre-Term Delivery Groups," *Medical and Biological Engineering and Computing*, vol. 46, no. 9, pp. 911-922 2008. [CrossRef] [Google Scholar] [Publisher Link]

[8] Franc Jager, "An Open Dataset with Electrohysterogram Records of Pregnancies Ending in Induced and Cesarean Section Delivery," *Scientific Data*, vol. 10, no. 1, pp. 1-12, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[9] Franc Jager, "Term-Preterm EHG Database," *PhysioNet*, 2012. [CrossRef] [Publisher Link]

[10] Janusz Jezewski et al., "Quantitative Analysis of Contraction Patterns in Electrical Activity Signal of Pregnant Uterus as an Alternative to Mechanical Approach," *Physiological Measurement*, vol. 26, no. 5, pp. 753-767, 2005. [CrossRef] [Google Scholar] [Publisher Link]

[11] Paul Fergus et al., "Prediction of Preterm Deliveries from Electrohysterogram Signals using Machine Learning," *PLOS One*, vol. 8, no. 10, pp. 1-16, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[12] Qiyang Sun, Alican Akman, and Björn W. Schuller, "Explainable Artificial Intelligence for Medical Applications: A Review," *ACM Transactions on Computing for Healthcare*, vol. 6, no. 2, pp. 1-31, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[13] Somayeh Mohammadi Far et al., "Prediction of Preterm Labor from the Electrohysterogram Signals based on Different Gestational Weeks," *Sensors*, vol. 23, no. 13, pp. 1-13, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[14] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436-444, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[15] Yoon Jae Lee et al., "Artificial Intelligence on Biomedical Signals: Technologies, Applications, and Future Directions," *Med-X*, vol. 2, no. 1, pp. 1-30, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[16] Mohamad Khalil, and Jacques Duchene "Uterine EMG Analysis: A Dynamic Approach for Change Detection and Classification," *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 6, pp. 748-756, 2000. [CrossRef] [Google Scholar] [Publisher Link]

[17] Tomasz Włodarczyk et al., "Machine Learning Methods for Preterm Birth Prediction: A Review," *Electronics*, vol. 10, no. 5, pp. 1-24, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[18] Félix Nieto-del-Amor et al., "Automatic Semantic Segmentation of EHG Recordings by Deep Learning: An Approach to a Screening Tool for use in Clinical Practice," *Computer Methods and Programs in Biomedicine*, vol. 254, pp. 1-14, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[19] Liwen Ding et al., "Prediction of Preterm Birth using Machine Learning: A Comprehensive Analysis based on Large-Scale Preschool Children Survey Data in Shenzhen of China," *BMC Pregnancy and Childbirth*, vol. 24, no. 1, pp. 1-14, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[20] Qiu-Yan Yu et al., "Predicting Risk of Preterm Birth in Singleton Pregnancies using Machine Learning Algorithms," *Frontiers in Big Data*, vol. 7, pp. 1-14, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[21] Ramprasaath R. Selvaraju et al., "Grad-CAM: Visual Explanations from Deep Networks Via Gradient-based Localization," *International Journal of Computer Vision*, vol. 128, no. 2, pp. 336-359, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[22] Raden Topan Aditya Rahman et al., "Proposed Model to Predict Preeclampsia using Machine Learning Approach," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 36, no. 1, pp. 694-702, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[23] Yu Chen et al., "Development and Validation of a Spontaneous Preterm Birth Risk Prediction Algorithm based on Maternal Bioinformatics: A Single-Center Retrospective Study," *BMC Pregnancy Childbirth*, vol. 24, no. 1, pp. 1-12, 2024. [CrossRef] [Google Scholar] [Publisher Link]