ARTICLE

Greylag Goose Optimization and Deep Learning-Based Electrohysterogram Signal Analysis for Preterm Birth Risk Prediction

Anis Ben Ghorbal[1,*], **Azedine Grine**[1], **Marwa M. Eid**[2,3,*] **and El-Sayed M. El-Kenawy**[4,5]

[1]Department of Mathematics and Statistics, Faculty of Science, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, 11432, Saudi Arabia
[2]Faculty of Artificial Intelligence, Delta University for Science and Technology, Mansoura, 11152, Egypt
[3]Jadara University Research Center, Jadara University, Irbid, 21110, Jordan
[4]Faculty of Engineering, Design and Information & Communications Technology (EDICT), School of ICT, Bahrain Polytechnic, Isa Town, P.O. Box 33349, Bahrain
[5]Applied Science Research Center, Applied Science Private University, Amman, 11931, Jordan
*Corresponding Authors: Anis Ben Ghorbal. Email: assghorbal@imamu.edu.sa; Marwa M. Eid. Email: marwa.3eeed@gmail.com
Received: 23 May 2025; Accepted: 05 August 2025; Published: 31 August 2025

ABSTRACT: Preterm birth remains a leading cause of neonatal complications and highlights the need for early and accurate prediction techniques to improve both fetal and maternal health outcomes. This study introduces a hybrid approach integrating Long Short-Term Memory (LSTM) networks with the Hybrid Greylag Goose and Particle Swarm Optimization (GGPSO) algorithm to optimize preterm birth classification using Electrohysterogram signals. The dataset consists of 58 samples of 1000-second-long Electrohysterogram recordings, capturing key physiological features such as contraction patterns, entropy, and statistical variations. Statistical analysis and feature selection methods are applied to identify the most relevant predictors and enhance model interpretability. LSTM networks effectively capture temporal patterns in uterine activity, while the GGPSO algorithm finetunes hyperparameters, mitigating overfitting and improving classification accuracy. The proposed GGPSO-optimized LSTM model achieved superior performance with 97.34% accuracy, 96.91% sensitivity, 97.74% specificity, and 97.23% F-score, significantly outperforming traditional machine learning approaches and demonstrating the effectiveness of hybrid metaheuristic optimization in enhancing deep learning models for clinical applications. By combining deep learning with metaheuristic optimization, this study contributes to advancing intelligent auto-diagnosis systems, facilitating early detection of preterm birth risks and timely medical interventions.

KEYWORDS: Preterm birth prediction; electrohysterogram signals; LSTM; time-series analysis; metaheuristic optimization; auto-diagnosis; clinical decision support

1 Introduction

Preterm birth, defined as delivery before 37 weeks of gestation, represents one of the most critical challenges in modern obstetric care, serving as a leading cause of neonatal morbidity and mortality worldwide. Every year, approximately 15 million children are born preterm, placing them at substantially elevated risk for a spectrum of life-threatening and long-term health complications. These complications typically manifest as respiratory distress syndrome, increased susceptibility to infections, neurodevelopmental impairments, and persistent developmental delays that can extend well into adulthood [1]. The profound impact of preterm birth extends beyond immediate medical concerns, creating substantial burdens for families through

prolonged hospitalizations, intensive medical interventions, and significant financial strain associated with specialized neonatal care [2]. With preterm birth affecting approximately one in every ten pregnancies globally, the condition represents a major public health challenge that demands urgent development of effective predictive and preventive strategies [3].

1.1 Clinical Importance of Early Prediction and Intervention

Early and accurate prediction of preterm birth risk carries immense clinical significance, as timely identification enables implementation of evidence-based interventions that can dramatically improve maternal and neonatal outcomes. When preterm birth risk is detected sufficiently early, clinicians can administer antenatal corticosteroids to accelerate fetal lung maturation, provide neuroprotective magnesium sulfate therapy, and arrange for in-utero transfer to tertiary care centers equipped with specialized neonatal intensive care units. These interventions have been demonstrated to significantly reduce the incidence of respiratory distress syndrome, intraventricular hemorrhage, and necrotizing enterocolitis, while simultaneously decreasing the overall duration and cost of neonatal hospitalization. Furthermore, early prediction facilitates enhanced maternal monitoring, allowing for optimal timing of delivery and preparation of appropriate medical resources, ultimately contributing to reduced neonatal mortality rates and improved long-term developmental outcomes.

1.2 Challenges in Current Predictive Approaches

The assessment and prediction of preterm birth risk present a multifaceted challenge characterized by complex interactions among biological, environmental, and sociodemographic factors [4]. Traditional clinical approaches have predominantly relied on subjective physician assessment combined with basic statistical methodologies, which, while clinically valuable, fail to capture the intricate multidimensional relationships that characterize preterm birth risk factors. These conventional methods often result in suboptimal sensitivity and specificity, leading to either delayed recognition of high-risk pregnancies or unnecessary interventions in low-risk cases. The inherent limitations of traditional approaches underscore the critical need for more sophisticated predictive models capable of analyzing complex, multimodal data types and identifying subtle patterns that may escape conventional analysis [5]. Advanced predictive systems have the potential to detect early warning signs and complex variable interactions that traditional methods cannot recognize, thereby enabling more precise identification of high-risk patients and facilitating targeted preventive interventions.

1.3 Transition to Advanced Computational Approaches

The emergence of artificial intelligence and machine learning technologies offers unprecedented opportunities to address the limitations of traditional preterm birth prediction methods. Unlike conventional statistical approaches that assume linear relationships and require predefined variable interactions, machine learning algorithms can automatically discover complex patterns within high-dimensional datasets, making them particularly well-suited for analyzing the multifactorial nature of preterm birth risk. The ability of these systems to process diverse data types—including physiological signals, clinical measurements, demographic information, and environmental factors—while simultaneously learning intricate feature relationships positions them as powerful tools for advancing predictive accuracy in obstetric care [6].

1.4 Existing Machine Learning Approaches and Their Limitations

Previous research efforts in preterm birth prediction have explored various machine learning approaches, including decision trees, support vector machines, random forests, and basic neural networks,

each offering distinct advantages and limitations. Decision tree algorithms provide high interpretability but often struggle with complex non-linear relationships and are prone to overfitting with limited data. Support vector machines demonstrate robust performance with small datasets but lack the ability to effectively model temporal dependencies inherent in physiological time-series data. Random forest approaches offer improved generalization through ensemble learning but may not capture the sequential patterns crucial for analyzing uterine activity signals. Basic neural networks, while capable of learning non-linear relationships, often fail to retain long-term dependencies necessary for modeling the temporal evolution of physiological processes leading to preterm birth [7].

1.5 Justification for LSTM Networks and Optimization Techniques

Long Short-Term Memory networks emerge as particularly promising solutions for preterm birth prediction due to their specialized architecture designed to capture temporal dependencies in sequential data. LSTM networks address the fundamental limitation of traditional recurrent neural networks—the vanishing gradient problem—through sophisticated gating mechanisms that enable selective retention and forgetting of information across extended time sequences. This capability is especially relevant for analyzing Electrohysterogram signals, where uterine contraction patterns evolve dynamically over time, and subtle temporal changes may indicate increasing preterm birth risk. The ability of LSTM networks to learn from time-series data makes them uniquely suited for processing physiological signals that exhibit temporal dependencies characteristic of preterm birth development [8].

However, the effectiveness of LSTM networks is critically dependent on optimal hyperparameter configuration, including the number of hidden units, learning rate, dropout rate, and batch size. Traditional hyperparameter optimization approaches, such as grid search or random search, often fail to explore the complex hyperparameter space efficiently, particularly in medical applications where computational resources may be limited and model reliability is paramount [9]. The challenge is further compounded by the inherent characteristics of medical datasets, which frequently exhibit class imbalance, limited sample sizes, and high noise levels that require sophisticated preprocessing and feature selection strategies.

1.6 Advanced Optimization through Metaheuristic Algorithms

Metaheuristic optimization algorithms, inspired by natural processes and collective behaviors observed in biological systems, offer powerful solutions for addressing the complex optimization challenges associated with deep learning model development. These algorithms, including Greylag Goose Optimization, demonstrate superior capability in navigating vast solution spaces while avoiding entrapment in local optima, making them particularly valuable for optimizing LSTM hyperparameters [10]. The integration of multiple optimization strategies, such as combining Greylag Goose Optimization with Particle Swarm Optimization, creates hybrid approaches that leverage the complementary strengths of different algorithms, potentially achieving superior optimization performance compared to individual methods [11]. When applied to LSTM networks, these integrated metaheuristic optimization techniques can systematically fine-tune hyperparameters, enhance convergence rates, and reduce overfitting, ultimately improving the predictive performance and generalization capability of deep learning models [12].

1.7 Identified Research Gaps

Although recent advances in machine learning and deep learning have been applied to preterm birth prediction, several persistent limitations continue to constrain model performance, clinical applicability, and generalizability. These gaps highlight the need for more robust, interpretable, and optimization-driven

approaches to improve outcomes in this domain. Specifically, the following research gaps have been identified:

- Existing LSTM-based approaches often lack systematic hyperparameter optimization, leading to suboptimal performance.
- Hybrid metaheuristic optimization strategies are rarely applied in the context of preterm birth prediction.
- Many studies fail to address the challenge of small, imbalanced medical datasets using appropriate augmentation techniques.
- Most models lack interpretability, limiting their clinical transparency and acceptance.
- Comparative benchmarking against traditional and hybrid models is frequently missing or inconsistent.

1.8 Main Contributions of This Study

In response to these limitations, this study presents a comprehensive deep learning framework that incorporates optimization, interpretability, and robust data preprocessing to improve the prediction of preterm birth from Electrohysterogram (EHG) signals. The major contributions of this work are outlined as follows:

- A GGPSO-optimized LSTM model is proposed for preterm birth prediction using Electrohysterogram signals.
- Clinically relevant features such as entropy and contraction-related parameters are extracted for model input.
- Synthetic Minority Oversampling Technique (SMOTE) is applied to address class imbalance in the limited dataset.
- The proposed model is benchmarked against conventional machine learning and other optimized LSTM models.
- SHapley Additive exPlanations (SHAP) and Random Forest feature importance are used to interpret model predictions and enhance transparency.
- The model demonstrates high predictive performance on real-world clinical data, showing potential for practical application.

The following sections include Related Works, which reviews prior research on preterm birth prediction methods, followed by Materials and Methods that describe the dataset used, preprocessing steps, and the proposed LSTM model and optimization methods. The Experimental Results section assesses model performance using different metrics, comparing different machine learning methodologies, and the Conclusion and Future Directions summarize the results and future directions for study.

2 Related Works

Approximately 10% of babies are born preterm, defined as delivery before 37 weeks of pregnancy, a condition associated with high child mortality and long-term neuro-logical impairments. As outlined in the study of [13], while signs of imminent preterm labor can often be detected, accurately predicting preterm births more than a week in advance remains a significant challenge. A novel deep learning approach was developed to address this issue by utilizing electrical uterine activity measurements taken from pregnant mothers around 31 weeks of pregnancy. The method incorporates a recurrent neural network and short-time Fourier transforms of these recordings combined with clinical data from public datasets. With an area under the receiver-operating characteristic curve of 0.78 (95% confidence interval: 0.76–0.80), the study highlighted the predictive strength of frequency patterns over time-based features, indicating the feasibility of automated

prediction of preterm births from short uterine activity recordings. This advancement could facilitate timely interventions to reduce the prevalence and adverse outcomes of preterm births.

Preterm birth is a global health crisis with significant health, financial, and economic impacts. As reported by [14], machine learning methods have been used to combine uterine contraction signals with predictive models to assess premature birth risk. This study focused on enhancing predictions using physiological signals, including uterine contractions and maternal and fetal heart rates, in South American women in active labor. The Linear Series Decomposition Learner improved prediction accuracy for both supervised and unsupervised models, with supervised models achieving high metrics after signal preprocessing. Unsupervised models effectively categorized patients using uterine signals but performed less accurately with heart rate data.

The vaginal microbiome has been linked to preterm birth, but its predictive accuracy is low due to individual variability in microbial communities. As reported by [15], a study involving 150 Korean pregnant women (54 preterm and 96 full-term) aimed to improve predictions using machine learning models. Cervicovaginal fluid samples, demographic data, white blood cell counts, and cervical length were analyzed, with subjects divided into training and test sets. Univariate analysis identified key bacterial markers, including Lactobacillus species, *Gardnerella vaginalis*, *Ureaplasma parvum*, and Prevotella timonensis. Combining these markers with clinical data, the random forest model achieved the highest area under the curve (0.84). Single tree models, like GUIDE, offered biologically interpretable insights with a similar predictive rate. This study highlights the potential of integrating microbiome data and clinical factors to enhance preterm birth predictions.

Uterine electromyography offers a noninvasive method to detect preterm birth, focusing on noncontraction signals to address irregular uterine activity during pregnancy. As shown by [16], 53 term and 47 preterm signal segments were analyzed using features like activity, mobility, and complexity. Preterm signals showed decreased mobility and increased complexity. A support vector machine achieved 84.3% accuracy and an F1-score of 82.8%, further improved by 3% through a decision fusion approach combining predictions from multiple channels. This method highlights non-contraction signals as potential biomarkers for reliable preterm birth prediction.

Monitoring uterine contractions provides critical insights into labor progression, with preterm deliveries posing risks to both mother and fetus. As outlined in the study by [17], electrohysterography offers a non-invasive method to detect preterm delivery and mitigate adverse effects. A novel method was proposed to predict preterm births using electrohysterography signals through three phases: preprocessing, feature extraction, and prediction. Noise and artifacts in the signals were removed using a band-pass filter and wavelet transform, followed by the extraction of features such as Shannon energy and median frequency. Prediction utilized an enhanced sheep flock optimized hybrid extreme artificial neural learning network, combining neural learning with optimization techniques for improved accuracy. Tested on the term-preterm electrophotography database, the method demonstrated superior performance in accuracy, recall, specificity, and f-measure, highlighting its potential for precise term and preterm birth prediction.

Maternal and fetal factors are known to influence preterm births, though all contributing variables remain undefined. As evidenced in the study by [18], maternal chronodisruption, including altered circadian rhythms due to factors like night light exposure and sleep duration, may negatively impact fetal development. This study analyzed a cohort of 380 births, classified as preterm or term, using machine learning methods to explore the influence of maternal habits, light exposure, and other gestational variables. Statistical analysis confirmed significantly lower cervix dilation, estimated fetal weight, and birth weight in preterm cases ($p < 0.05$). Machine learning further identified non-obvious relationships, with features such as bedroom light exposure and nighttime light through windows emerging as key predictors in a decision tree

model. These findings suggest that chronodisruption factors could enhance preterm birth prediction and prevention efforts.

Globally, over 1 in 10 babies are born prematurely, often facing lifelong health challenges such as learning difficulties, hearing impairments, and vision loss. As re-ported by [19], monitoring of uterine contractions allows physicians to see how the pregnancy is progressing and to diagnose labor, which in turn can prevent complications by ensuring timely hospital visits. A machine learning and deep learning-based system was developed to predict labor and diminish the impact of premature birth of the mother and fetus. The system's reliability was tested, resulting in a high accuracy rate of 0.98, showing the possibility of such advanced computational methods to improve pregnancy outcomes.

Heart rate variability data from preterm and full-term infants can serve as a basis for estimating functional maturational age, offering insights into infant development. As outlined in the study by [20], a machine learning model was developed to predict functional maturational age and deviations from postmenstrual age using heart rate variability features, which may serve as a maturational progress measure. The feature selection was performed by filtering and genetic algorithms, where data from 50 healthy infants born between 25 and 41 weeks of gestational age were analyzed. The ensemble machine learning algorithm applied through a linear + random forest regression had a mean absolute error of 0.93 weeks with 95% confidence intervals. Similarly, bradycardia and respiration rate variability data were used to derive similar results. This could be used noninvasively in neonatal intensive care units to monitor maturation in real life to improve neonatal care outcomes.

This study developed an explainable algorithm for predicting lateonset sepsis in preterm infants using continuous multi-channel physiological signals in a neonatal intensive care unit. As demonstrated by [21], the algorithm analyzed electrocardiogram and chest impedance derived data of heart rate variability, respiration and motion, including gestational age and birth weight. Features were extracted from the 24 h preceding sepsis onset or matched control time point for 127 preterm infants (59 who developed sepsis and 68 without), and data were analyzed. However, an extreme gradient boosting classifier produced the best prediction performance using all signal channels with an area under the curve of 0.88, positive predictive value of 0.80, and negative predictive value of 0.83 in the 6 h before sepsis onset. Motion data inclusion added to predictive accuracy, and visualization of the feature impacts helped to increase algorithm interpretability for possible clinical intervention.

Preterm birth, defined as delivery before 37 weeks of pregnancy, is a leading cause of newborn mortality and long-term health issues in children. As reported by [22], a novel machine learning approach was developed to predict preterm birth using cervical electrical impedance spectroscopy data collected at 20–22 weeks of pregnancy. The method involves selecting the optimal impedance spectrum using a filter and predicting preterm birth with polynomial feature-based models. Using data from 438 patients, the model achieved an area under the curve of 0.76 with a random forest predictor and 0.74 with logistic regression. For patients without treatment interventions, performance improved to 0.8 and 0.79, respectively. This approach outperformed existing methods, highlighting its potential clinical utility in early preterm birth prediction.

Preterm birth is the leading cause of newborn mortality, with survivors often facing severe health complications. Threatened preterm labor is the primary reason for hospitalization in late pregnancy. As reported by [23], current diagnostic methods, such as the Bishop score and cervical length, have high negative predictive values but lack strong positive predictive abilities. This study evaluated classification algorithms based on electrohysterographic recordings to predict imminent labor (<7 days) in women with threatened preterm labor. Algorithms such as random forest, extreme learning machine, and K-nearest neighbors were assessed using temporal, spectral, and nonlinear electrohysterographic parameters with and without obstetric data. The extreme learning machine achieved the highest F1-score (90.2% ± 4.43%) and sensitivity

when using the 50th percentile of parameters and obstetric inputs. These findings high-light the potential of computationally efficient models to improve labor predictions and inform clinical decision-making.

Table 1 shows a detailed summary of recent studies addressing preterm birth prediction using machine learning and signal processing techniques. The research spans diverse data types, including uterine electromyography, heart rate variability, cervical impedance, microbiome composition, and maternal environmental exposure. A wide array of methods—ranging from deep neural networks to decision trees and hybrid optimizers—were employed. The studies emphasize the growing integration of computational intelligence in healthcare, with promising results for early and non-invasive detection of preterm birth risk.

Table 1: Summary of related works

Study	Data type	Method	Key findings
[13]	Uterine Electromyography (EMG) + clinical data at advanced gestation	Recurrent Neural Network (RNN) + Short-Time Fourier Transform (STFT)	Frequency-domain features provided better discrimination compared to time-domain features
[14]	Contractions and heart rate signals	Linear Series Decomposition (Supervised/Unsupervised)	Supervised learning showed improved performance with appropriate preprocessing
[15]	Vaginal microbiome and clinical indicators	Random Forest + Generalized, Unbiased, Interaction Detection and Estimation (GUIDE) Tree	Microbial signatures significantly enhanced predictive capability
[16]	Non-contraction uterine Electromyography (EMG) signals	Support Vector Machine (SVM) with Decision Fusion	Preterm labor signals exhibited higher complexity, aiding classification
[17]	Electrohysterography (EHG) signals	Hybrid Neural Network optimized by Sailfish Optimizer (SFO)	Model demonstrated high accuracy in predicting labor onset
[18]	Light exposure and sleep behavior data	Decision Tree	Disruptions in light exposure patterns were associated with increased preterm risk
[19]	Uterine contraction monitoring data	Combined Machine Learning (ML) and Deep Learning (DL) architecture	The system showed strong predictive performance for labor detection
[20]	Infant heart rate variability and respiration	Linear regression with Random Forest	Framework effectively estimated functional age of neonates
[21]	Neonatal Intensive Care Unit (NICU) data including Heart Rate Variability (HRV), motion, Electrocardiogram (ECG)	Extreme Gradient Boosting (XGBoost)	Motion data significantly improved the predictive accuracy of the model

Table 1 (continued)

Study	Data type	Method	Key findings
[22]	Cervical impedance measured during early gestation	Random Forest + Logistic Regression	Demonstrated utility as an early screening tool for risk assessment
[23]	Electrohysterography (EHG) signals combined with obstetric factors	Extreme Learning Machine (ELM), Random Forest, K-Nearest Neighbors (KNN)	Models were highly effective in classifying cases with threatened preterm labor

Despite the promising advances in preterm birth prediction research, several significant gaps remain unaddressed in the existing literature. Most notably, previous studies have largely overlooked the critical role of advanced optimization techniques in enhancing deep learning model performance, typically relying on default hyperparameter settings or basic grid search methods rather than sophisticated optimization strategies. Furthermore, most previous research fails to achieve an optimal balance between model interpretability and predictive performance, often sacrificing one for the other. The field lacks systematic exploration of hybrid metaheuristic approaches and robust validation strategies that can effectively handle class imbalance issues in medical datasets. These methodological gaps highlight the need for more sophisticated approaches that combine advanced optimization techniques with careful consideration of medical data characteristics and clinical requirements.

3 Materials and Methods

This section outlines the methodology employed for analyzing Electrohysterogram signals to predict the risk of preterm birth. The analysis involves transforming the raw signals into structured features, followed by the application of deep learning techniques for classification. A Long Short-Term Memory (LSTM) network was selected due to its strength in modeling temporal dependencies inherent in physiological time-series data. To enhance model performance, a Hybrid Greylag Goose and Particle Swarm Optimization (GGPSO) algorithm was used specifically for hyperparameter tuning. This optimization process was designed to search for the most effective combination of LSTM parameters—such as the number of hidden units, learning rate, dropout rate, and batch size—thereby improving the model's accuracy and generalization ability. The overall framework integrates statistical analysis, deep learning, and methodology to develop a robust and interpretable classifier for early detection of preterm birth.

Fig. 1 illustrates the complete methodology pipeline of our proposed GGPSO-LSTM framework for preterm birth prediction. The workflow processes EHG data through preprocessing and SMOTE augmentation, extracts five key physiological features, and splits data into training-testing sets. The LSTM network analyzes temporal patterns while GGPSO optimization fine-tunes hyperparameters, ultimately producing preterm birth classification results with comprehensive performance evaluation.
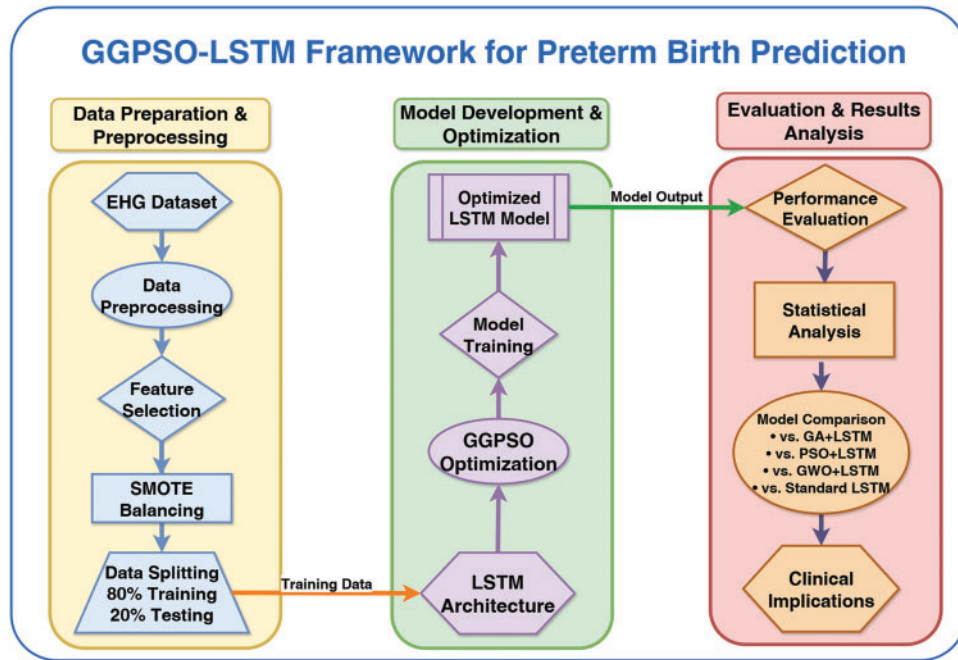
Figure 1: Proposed GGPSO-LSTM framework for preterm birth prediction

3.1 Dataset

The original dataset used in this study was derived from Electrohysterogram (EHG) signals, which are biomedical time-series recordings used to monitor uterine electrical activity during pregnancy. These signals provide critical insight into contraction behavior and can help determine whether a pregnancy is at risk of preterm birth. The dataset is binary classified and consists of 58 instances, each corresponding to a single 1000-second EHG recording. This transformation process involved mining key statistical and temporal features from each signal, resulting in the following variables:

- **Count Contraction:** The number of discrete uterine contractions detected within the signal window.
- **Length of Contraction:** The average duration of each contraction event, reflecting the temporal characteristics of uterine activity.
- **Standard Deviation (STD):** The standard deviation of the signal amplitude, which captures the variability and fluctuation of contractions over time.
- **Entropy:** The Shannon entropy of the signal, quantifying its complexity and irregularity to assess the unpredictability in uterine contraction patterns.
- **Contraction Time:** The total cumulative time of all contraction events observed within the signal, indicating overall contraction burden.
- **Preterm:** The binary target label, where 1 indicates a preterm birth case and 0 denotes a non-preterm case, used for classification purposes.

These extracted features reflect both physiological and signal-based properties relevant to uterine dynamics and preterm birth prediction. The resulting tabular format made the dataset suitable for machine learning classification tasks using both deep learning and optimization-based models.

Due to the limited size of the dataset, special care was taken to maintain balance and reduce the risk of overfitting during model evaluation. The data was split into 80% for training and 20% for testing, ensuring that both classes (preterm and non-preterm) were proportionally represented in each subset. All

models, including LSTM and its optimized variants, were trained exclusively on the training set. Performance metrics—including accuracy, sensitivity, specificity, precision, and F1-score—were computed on the held-out test set. Each experiment was repeated 10 times with different random seeds to reduce the variance in evaluation, and the average results were reported. While the dataset size imposes inherent limitations, this evaluation procedure provides a consistent and fair basis for model comparison.

3.2 Data Analysis and Feature Exploration

The role of data analysis includes determining how structured, related, and patterned the dataset can be used for predictive modeling. The key features are investigated with different statistical and visualization approaches to understand their distribution, the correlations among various features, and how they may impact the classification of preterm birth in this study. In this case, it is helpful to identify strong and weak correlations among features, pick the most helpful features, and relegate the redundant ones. Secondly, the distribution of the features can help us to understand how variable data is, that central tendencies, including outliers might affect the model's performance. These patterns are considered to improve the quality of feature selection, the interpretability of the model, and the assurance that the machine learning approach presents the underlying characteristics of preterm birth prediction [24].

This study uses a binary class dataset with 58 instances extracted from 1000-second-long Electrohysterogram signals. These signals give the physicians crucial information about the kind of uterine activity and help them conclude whether the baby is preterm or not. Key physiological indicators detailed under contraction patterns, signal entropy and statistical variations are tied to preterm birth classification and are included in the dataset. The Electrohysterogram signal data is preprocessed and structured to enter a tabular format to make it suitable for machine learning applications. As the dataset comes from signal processing techniques, it also learns such temporal patterns that have the potential to improve predictive accuracy. This study uses these features to build a robust machine learning model to accurately identify preterm birth risks.

Fig. 2 presents a Pearson correlation matrix among all extracted features, including the binary target variable 'Preterm'. The matrix helps illustrate the degree of linear relationship between features and with the target class. Notably, features such as Entropy, Contraction Times, and Count Contraction show moderate positive correlation with the Preterm label. The stronger correlations are represented by darker shades of red or blue and weaker correlations by lighter shades. Entropy and Contraction Times are correlated more with the preterm outcome than Entropy and Contraction Times and perhaps should be considered more in predictive modeling. In contrast, the negative correlation of "STD" with the preterm status is relatively weak, suggesting a less direct prediction. Correlation matrix at an overall level is a basic understanding of how these features interact with each other and this target variable, for feature selecting before developing the model.

Fig. 3 depicts the density plot of the feature "Entropy" per 'preterm' status, showing different plotting for 'Non-Preterm' and 'Preterm' groups. The entropy distribution of the "Pre-term" group is much flatter and broader, whereas that for the "Non-Preterm" group is sharply peaked and therefore indicates a more concentrated range of entropy values around a certain point. The contrast may indicate that 'Entropy' is essential to distinguishing between the two classes. In certain regions where the two curves somewhat overlap, this indicates ambiguity in classifying the data, and refining the features or thresholds in the data might help improve classification accuracy.

Fig. 4 plots the features extracted from the Electrohysterogram signals, including 'Count Contraction,' 'Length of Contraction,' 'STD,' 'Entropy,' and 'Contraction Times.' Every subplot illustrates how a particular feature provides contents to the dataset and its degree of variability. The distributions for 'Features' such as 'Entropy' and 'Contraction Times' are tightly clustered, indicative of a consistency of values. However,

"Length of Contraction" and "STD" have broader ranges, indicating more variance. These features have distinct peaks and spreads which might give clues of its discriminatory power in the preterm classification model. The patterns across the distributions point to the fact that preprocessing and normalization enhance the robustness of predictions.
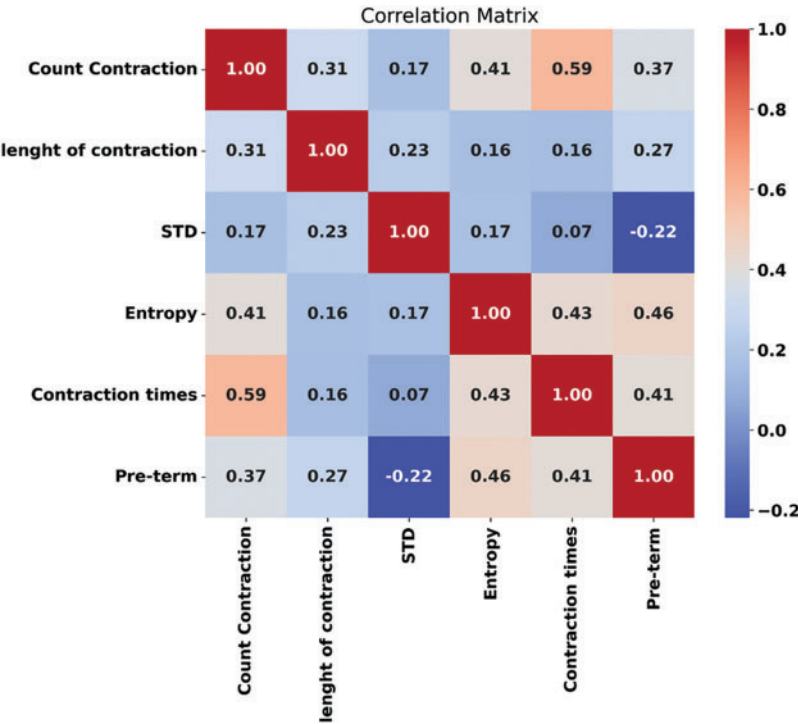

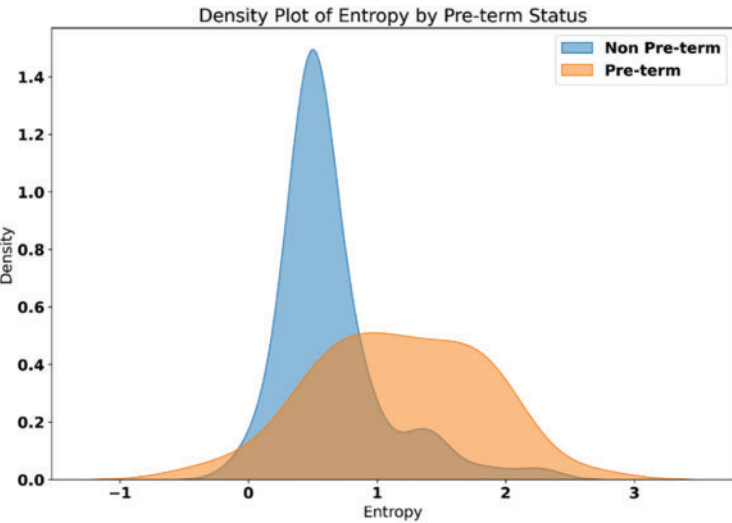**Figure 2:** Correlation matrix of features for preterm birth prediction
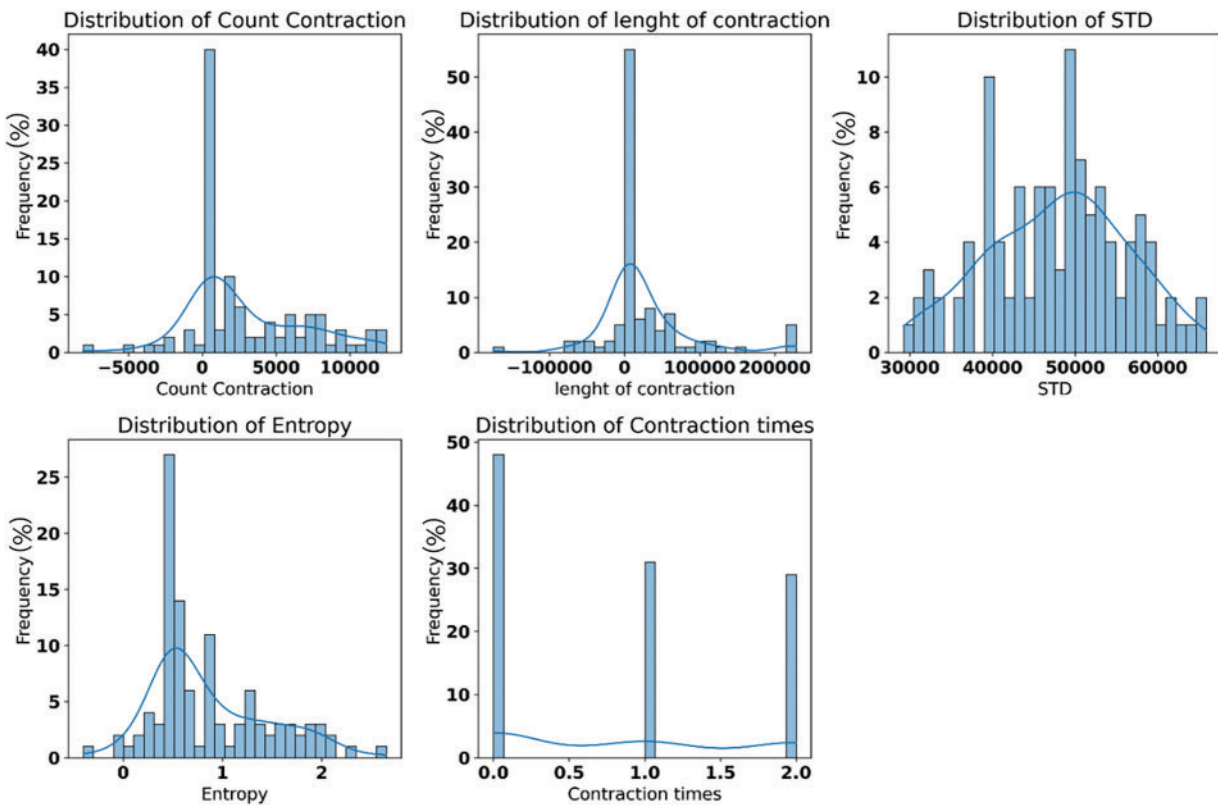

Figure 3: Density plot of entropy by preterm status

Figure 4: Frequency distributions of key features from electrohysterogram signals

Fig. 5 reveals the inherent challenge that researchers face when working with real-world medical data—the problem of class imbalance. In our original dataset, we can clearly see that nature doesn't provide us with perfectly balanced samples. The chart shows that we have 39 cases of non-preterm births compared to only 19 cases of preterm births. This 2:1 ratio is actually quite typical in medical datasets, where certain conditions or outcomes are naturally less frequent than others. While this reflects the real-world prevalence of preterm births, it creates a significant challenge for machine learning algorithms, which tend to become biased toward the majority class. This imbalance means that without proper handling, our model might become very good at identifying non-preterm cases but struggle to accurately detect the preterm cases that we're most concerned about preventing.

Fig. 6 demonstrates the transformative power of the SMOTE (Synthetic Minority Oversampling Technique) algorithm in addressing our class imbalance challenge. What we see here is a perfectly balanced dataset with exactly 156 samples in each category—a dramatic improvement from our original uneven distribution. SMOTE works its magic by intelligently creating synthetic examples of the minority class (preterm births) rather than simply duplicating existing samples. This technique examines the characteristics of real preterm cases and generates new, realistic samples that share similar features but aren't exact copies. The result is a dataset that gives our machine learning model an equal opportunity to learn the patterns associated with both preterm and non-preterm births. This balanced approach is crucial for developing a fair and accurate predictive model that won't overlook the critical preterm cases that clinicians need to identify early.
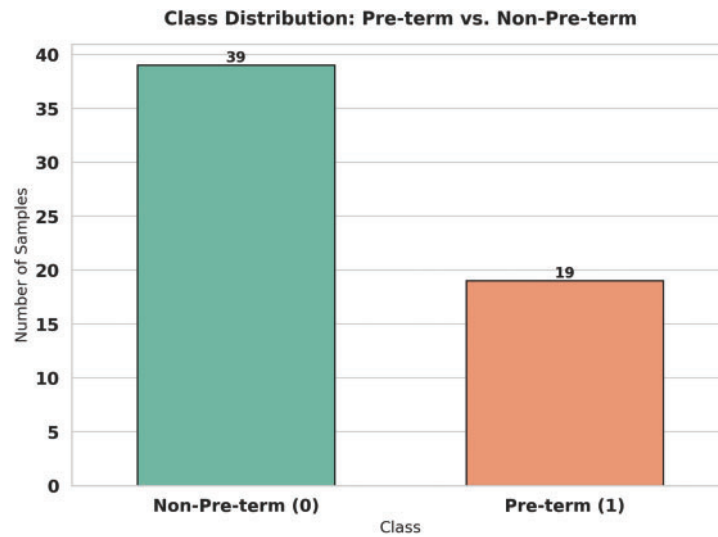
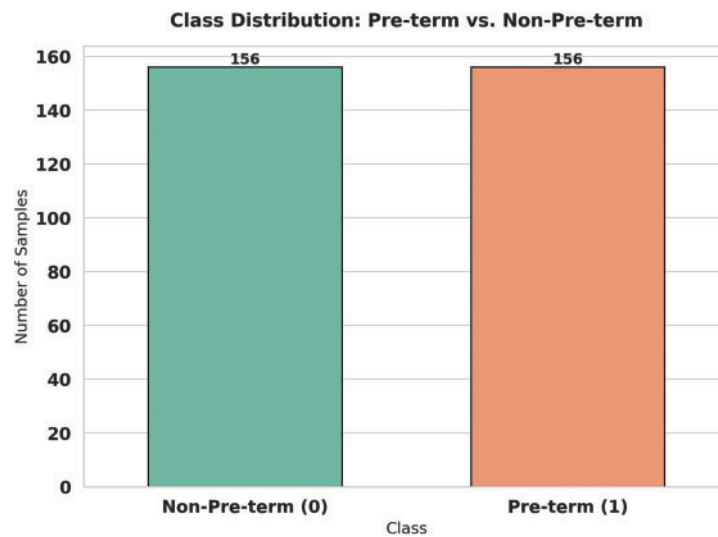**Figure 5:** Correlation matrix of features for preterm birth prediction

Figure 6: Correlation matrix of features for preterm birth prediction

3.3 Long Short-Term Memory (LSTM)

The Long Short-Term Memory (LSTM) is a specialized type of Recurrent Neural Network (RNN) that takes care of the vanishing gradient problem associated with the original approach of the RNN when dealing with a long sequence of data. However, traditional RNNs find it difficult to retain information for extended time intervals due to a phenomenon during training where the gradient values steadily diminish or explode, making learning impossible. This problem can be overcome by LSTMs, which enable them to selectively remember and forget information, thus making them very helpful for time series prediction, sequential data classification, and pattern recognition tasks [25].

Recognizing the fact that the analysis of Electrohysterogram signal is crucial in cases of preterm birth classification because the patterns of time dependent contractions are present in the Electrohysterogram signals. Unlike traditional machine learning models based on static feature sets, LSTM networks can learn

temporal dependencies in the data, mainly processing subtle but crucial changes over time. This inherent ability of LSTMs is beneficial for separating preterm and non-preterm cases, where a physiological signal evolves dynamically.

An LSTM unit consists of a cell state and three gates—the forget gate, input gate, and output gate—which work together to control the flow of information through the network. The forget gate determines which information from the previous cell state should be discarded. It takes the previous hidden state (h_{t-1}) and the current input (x_t) and applies a sigmoid activation function, producing a value between 0 and 1:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{1}$$

where W_f and b_f are the weight matrix and bias, respectively.

The input gate decides which new information should be stored in the memory cell. It consists of two parts: a sigmoid function that determines the significance of incoming information and a tanh activation function that generates candidate values to be added to the memory cell:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{2}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{3}$$

where $\tilde{C}_t$ represents the candidate values that could be added to the cell state.

The cell state is updated by combining the retained information from the previous state and the newly selected input information:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \tag{4}$$

The output gate determines what information should be passed forward as the hidden state to the next layer. It first applies a sigmoid function to determine the significance of the cell state:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{5}$$

The hidden state is then updated by applying a tanh function to the cell state and multiplying it by the output gate:

$$h_t = o_t \cdot \tanh(C_t) \tag{6}$$

The hidden state (h_t) is then used in the next step, allowing the LSTM to capture long-term dependencies effectively.

Time series problems can be solved with LSTMs, but the set of hyperparameters needed to make them perform well includes the number of hidden units, learning rate, batch size, and dropout rate. Such problems as overfitting, slow convergence and low accuracy can happen in terms of tuning so that it can go off the mark. Metaheuristic optimization algorithms can be used to explore potential solutions and deal with them efficiently. Some optimization techniques fine-tune the hyperparameter, reduce the computational cost, and enhance the model's generalization for LSTM training. The model will reduce over-fitting, convergence speed, and with higher accuracy by integrating LSTM with hybrid optimization methods.

3.4 Hybrid Greylag Goose and Particle Swarm Optimization (GGPSO) Algorithm

GGPSO is created based on the Greylag Goose Optimization (GGO) [26], which mimics the aggregation movement behavior and optimize on the base of velocity strategy of PSO. It is a blend of the exploitation

of current knowledge and the repeated acquisition of new information in the process of optimizing and avoiding the problem of getting fixed in local optima.

In this method, the population of search agents represents candidate solutions to the optimization problem. Each agent's performance is evaluated using a fitness function F_n, which determines the best agent position (P) at each iteration. The population is divided into two groups: an exploration group (n_1) that searches new regions of the solution space and an exploitation group (n_2) that fine-tunes solutions around promising candidates.

The position update rules in GGPSO alternate between GGO-based exploration and PSO-based exploitation, depending on whether the current iteration is even or odd.

(1) GGO-Based Exploration Step

During even iterations $(t\%2 == 0)$, the algorithm mimics the flocking behavior of geese, where agents adjust their positions based on leader dynamics and environmental factors. If a random condition $(r_3 < 0.5)$ is met and the absolute distance $|A|$ is small, the agent updates its position using:

$$X(t+1) = X'(t) - A_j \cdot C \cdot X'(t) - X(t) \tag{7}$$

where $X(t)$ is the agent's current position, $X'(t)$ is the best-known updated position, A_j and C introduce random variations to enhance search diversity.

If this condition is not met, movement is determined by a weighted combination of three randomly selected search agents $(X_{paddle1}, X_{paddle2}, X_{paddle3})$. The update factor z is calculated using an exponential decay function:

$$z = 1 - \left(\frac{t}{t_{max}}\right)^2 \tag{8}$$

The new position is updated using:

$$X(t+1) = \frac{w_1}{w_1 + w_2 + w_3} X_{paddle1} + z \cdot \frac{w_2}{w_1 + w_2 + w_3} (X_{paddle2} - X_{paddle3}) + (1 - z) \cdot \frac{w_3}{w_1 + w_2 + w_3}$$
$$(X - X_{paddle}) \tag{9}$$

where w_1, w_2, w_3 are weight factors, X_{paddle} represents a reference agent influencing movement.

This GGO-inspired movement ensures a broad search space, preventing stagnation in local optima.

(2) PSO-Based Exploitation Step

During odd iterations $(t\%2 \neq 0)$, the algorithm switches to a PSO-based velocity update to refine solutions. The search agents adjust their positions based on personal best (P_b) and global best (P) solutions:

$$X(t+1) = X(t) + w_k U(t) + c_1 r_1 (P_b - P_{old}) + c_2 r_2 (P - P_{old}) \tag{10}$$

where w_k is the inertia weight controlling the effect of previous velocities, $U(t)$ is the velocity update term, c_1, c_2 are acceleration coefficients, r_1, r_2 are random numbers between 0 and 1.

This equation allows agents to converge towards optimal solutions while maintaining some randomness to escape local optima.

(3) Individual Position Updates

For individual agents, a different position update rule applies:

$$X(t+1) = X(t) + D(1+z) \cdot w \cdot (X - X_{Flock1}) \tag{11}$$

where D is a distance parameter and X_{Flock1} represents the leading flock's position. This allows agents to gradually align with the best-discovered solutions.

When $t\%2 == 0$ in the exploitation group, the position update follows:

$$
\begin{aligned}
X_1 &= X_{\text{Sentry}_1} - A_1 \cdot \left| C_1 \cdot X_{\text{Sentry}_1} - X \right|, \\
X_2 &= X_{\text{Sentry}_2} - A_2 \cdot \left| C_2 \cdot X_{\text{Sentry}_2} - X \right|, \\
X_3 &= X_{\text{Sentry}_3} - A_3 \cdot \left| C_3 \cdot X_{\text{Sentry}_3} - X \right|.
\end{aligned}
\tag{12}
$$

Through optimizing LSTM hyperparameters using GGPSO in this study, it realizes a better classification accuracy of preterm birth, so the deep learning model is more reliable and interpretable. The actual flowchart of the proposed algorithm, the GGPSO, is presented in Algorithm 1 to depict the functioning of the search agents in an environment, how they fine-tune their algorithms, and how they aim at the best performance.

Algorithm 1: Pseudo-code of greylag goose particle swarm optimization (GGPSO)

Require: Population size n, maximum iterations $t_{\max}$, objective function F_n, parameters for GGO and PSO
Ensure: Optimal solution P
　1: **Initialize** population X_i, evaluate fitness F_n, select best P
　2: **Divide** into exploration (n_1) and exploitation (n_2) groups
　3: **for** iteration $t = 1$ to $t_{\max}$ **do**
　4:　　**For exploration group:**
　5:　　**if** t even **then**
　6:　　　Update positions using GGO rules based on random condition:
　7:　　　**if** local condition met **then**
　8:　　　　Perform local update
　9:　　　**else**
10:　　　　Update using random paddle agents
11:　　　**end if**
12:　　**else**
13:　　　Update positions using PSO rules
14:　　**end if**
15:　　**For exploitation group:**
16:　　**if** t even **then**
17:　　　Update positions using "sentry" agents' information
18:　　**else**
19:　　　Update positions based on flock alignment
20:　　**end if**
21:　　Evaluate new fitness, update global best P
22:　　**If** no improvement over two iterations, adjust group sizes (n_1, n_2)
23: **end for**
24: **return** Optimal solution P

These methodologies have been very well worked to give a correct and dependable preterm birth prediction in this study. The method of the study is proposed to identify the important characteristics of Electrohysterogram signal from the area of statistics, signal processing and feature selection and machine learning algorithms. By incorporating the adopted GGPSO, hyperparameters in the model are optimized,

thereby enhancing the model efficiency through LSTM networks. This makes the classification of the model far more accurate and guarantees the ability of the model to generalize from one data set to another. This framework would provide, in the long term, an interpretable and reliable way to detect preterm birth at an early stage and can also be beneficial in clinical decision-making or in maternal health care practices.

The incorporation of these two methodologies provides a strong background to the development of intelligent healthcare systems, which would, in one way or another, assist physicians in coming up with suitable decisions concerning high-risk preterm births.

4 Experimental Results

The examination of the results shows that state-of-the-art machine learning and optimization approaches can deliver outstanding classification accuracy. Successfully applied to capture temporal characteristics, the LSTM model achieved high outcomes for such vital signs as accuracy, sensitivity, and specificity. Hence, these results demonstrate the ability to accomplish classification tasks with higher reliability and accuracy. Moreover, the integration of the proposed GGPSO optimization thoughtfully enhanced the model by optimizing the already proven LSTM, which in turn helped the model to reach new levels of accuracy and sensitivity together with the generalization of different scenarios. The GGPSO algorithm was configured with 20 candidate solutions per population, executed over 100 iterations, and repeated across 30 independent runs to ensure robust performance evaluation. The complementarity of optimization algorithms and machine learning models thus delineates their ability to solve multi-variable classification problems with a high degree of accuracy and reliability in which applications are critical and a tight constraint on error margin is required.

To ensure a robust and fair comparison, several well-established machine learning models were implemented along with the proposed framework. These included the Gated Recurrent Unit (GRU), Convolutional Neural Network (CNN), and Artificial Neural Network (ANN), each chosen for their recognized strengths in time-series classification and biomedical signal analysis. The GRU model was configured with a single recurrent layer containing 64 hidden units and a dropout rate of 0.2, optimized using the Adam optimizer. The CNN architecture [27] employed a 1D structure with convolutional and pooling layers, followed by a dense classification layer. The ANN consisted of two fully connected hidden layers with ReLU activation and dropout regularization, optimized with binary cross-entropy loss. Additionally, several hybrid models were considered to evaluate the impact of metaheuristic optimization strategies. These included GA + LSTM, PSO + LSTM, and GWO + LSTM, where Genetic Algorithm (GA) [28], Particle Swarm Optimization (PSO), and Grey Wolf Optimizer (GWO) [29] were respectively applied to fine-tune LSTM hyperparameters such as the number of units, learning rate, and dropout values. All models were trained under a consistent experimental protocol using 10-fold cross-validation, stratified data splitting, and early stopping to prevent overfitting. This comparative strategy highlights the effectiveness of the proposed GGPSO + LSTM model within a standardized evaluation framework, reinforcing its competitive advantage in predictive performance.

4.1 Comprehensive Feature Analysis and Model Interpretability

Table 2 presents the statistical validation results for all extracted EHG signal features using p-value analysis with a significance threshold of $\alpha = 0.05$. The analysis reveals that four out of five features demonstrate strong statistical significance, with Entropy showing the most exceptional discriminative power ($p = 1.06 \times 10^{-20}$). Contraction-related features (timing, count, and duration) all exhibit extremely high significance levels, confirming their clinical relevance for preterm birth prediction. Notably, the Standard Deviation (STD) feature shows no statistical significance ($p = 0.769$), indicating its lack of discriminative value and supporting its exclusion from the final model. This statistical foundation validates the feature selection methodology and provides robust evidence for the clinical applicability of the identified biomarkers.

Table 2: Statistical significance and clinical relevance of features

Feature	p-value	Significance level	Statistical interpretation	Clinical relevance
Entropy	1.06×10^{-20}	*** (Extremely Significant)	Highest discriminative power	Signal complexity indicator
Contraction times	2.64×10^{-18}	*** (Extremely Significant)	Very strong predictor	Timing pattern analysis
Count Contraction	2.28×10^{-13}	*** (Extremely Significant)	Strong discriminative ability	Frequency assessment
Length of contraction	1.23×10^{-5}	** (Highly Significant)	Moderate predictor strength	Duration measurement
STD	0.769	NS (Not Significant)	No discriminative value	Should be excluded

Fig. 7 demonstrates the SHAP (SHapley Additive exPlanations) feature importance analysis, revealing the relative contribution of each EHG signal feature to the model's prediction decisions. The analysis shows that Entropy achieves the highest SHAP importance value, followed closely by Count Contraction, confirming these features as the primary drivers of preterm birth classification. Contraction times and Length of contraction exhibit moderate importance levels, while STD shows minimal contribution to model predictions. This SHAP analysis provides essential evidence for model interpretability, directly addressing clinical requirements for understanding which physiological parameters most significantly influence preterm birth risk assessment. The consistent ranking with statistical significance results (Table 2) validates the robustness of feature selection methodology.
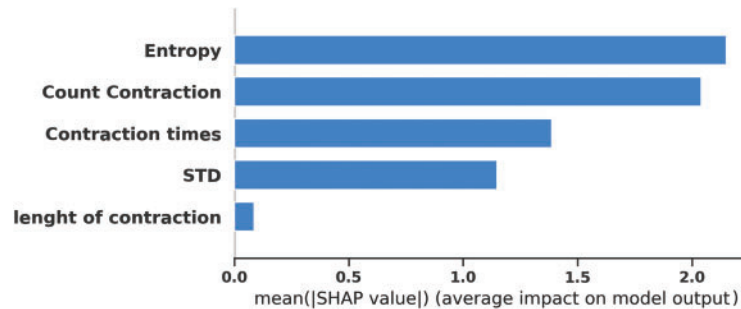

Figure 7: SHAP feature importance analysis

Fig. 8 presents the Random Forest feature importance analysis, providing an ensemble-based perspective on feature significance that complements the SHAP analysis. The Random Forest algorithm assigns the highest importance score (~0.35) to Count Contraction, followed by Entropy (~0.32), demonstrating slight variation in ranking compared to SHAP but maintaining consistency in identifying the top two most influential features. Contraction times and Length of contraction receive moderate importance scores, while STD consistently shows the lowest contribution (~0.04) across both analytical methods. This convergence between SHAP and Random Forest importance rankings strengthens the reliability of feature

prioritization and validates the robustness of the identified biomarkers across multiple machine learning interpretability approaches.
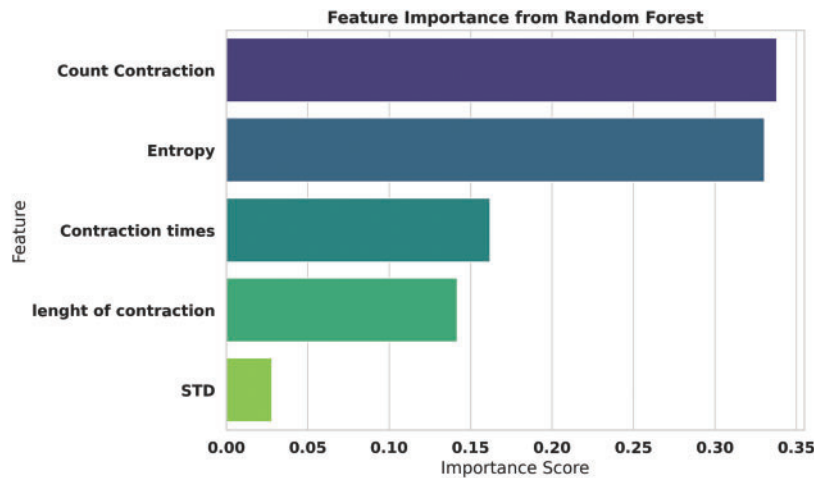

Figure 8: Random forest feature importance analysis

Fig. 9 illustrates the detailed SHAP values distribution, providing granular insights into how individual feature values impact model predictions for each sample in the dataset. The visualization reveals distinct directional patterns where higher Entropy values (represented by red points) consistently contribute positive SHAP values, pushing predictions toward preterm classification, while lower Entropy values (blue points) generate negative SHAP values, supporting non-preterm predictions. Similarly, Count Contraction demonstrates clear threshold effects where specific value ranges strongly influence classification outcomes. This detailed analysis enables clinicians to understand feature-specific decision boundaries and provides actionable insights into the physiological parameter ranges that most significantly indicate preterm birth risk.
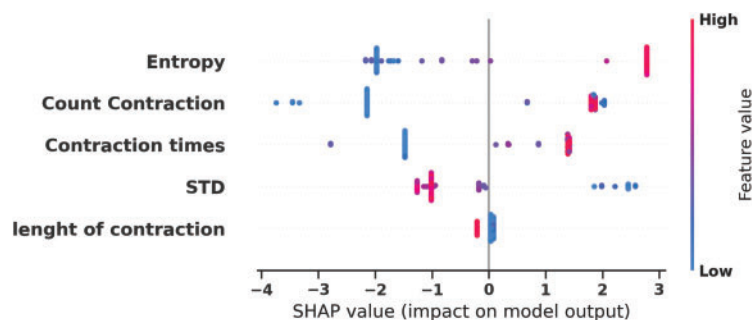

Figure 9: SHAP values distribution for individual feature impact on model predictions revealing directional effects of feature values

Fig. 10 presents the Principal Component Analysis (PCA) projection of the five-dimensional EHG feature space onto a two-dimensional representation, demonstrating exceptional class separability characteristics. The analysis reveals that the first principal component (PC1) captures an remarkable 98.7% of the total variance, indicating that the extracted features contain highly concentrated discriminative information along the primary axis of variation. The second principal component (PC2) accounts for only 1.2% of the variance, suggesting that most classification-relevant information is effectively captured in the first dimension. The visualization clearly shows distinct spatial clustering with preterm cases (blue points) and non-preterm

cases (red points) forming separate groups with minimal overlap, validating the effectiveness of the feature extraction methodology and supporting the model's classification capabilities.
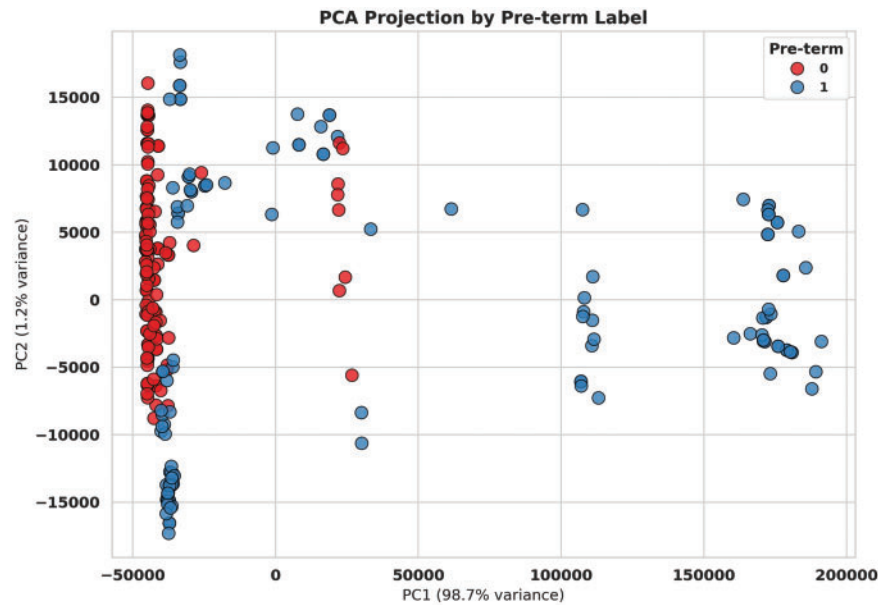

Figure 10: PCA projection visualization

Fig. 11 displays the t-SNE (t-Distributed Stochastic Neighbor Embedding) visualization, revealing complex non-linear clustering patterns that complement the linear PCA analysis and provide deeper insights into the feature space structure. The enhanced two-dimensional projection demonstrates well-defined spatial groupings with preterm cases (orange points) and non-preterm cases (green points) forming distinct clusters across the embedding space. The t-SNE analysis uncovers multiple sub-clusters within each class, suggesting potential phenotypic variations in EHG signal patterns that may correspond to different underlying pathophysiological mechanisms. Despite these sub-groupings, clear inter-class boundaries are maintained, with minimal overlap between preterm and non-preterm regions, confirming the discriminative power of the extracted features in a non-linear context.

Fig. 12 presents a comprehensive correlation analysis focusing on feature pairs that exceed the 0.5 correlation threshold, identifying significant linear relationships between extracted EHG signal character-istics. The analysis reveals that Entropy and Contraction times exhibit an exceptionally strong correlation ($r \approx 1.0$), indicating that signal complexity measures are fundamentally linked to temporal contraction patterns in EHG recordings. Count Contraction demonstrates strong positive correlations with both Entropy ($r \approx 0.9$) and Contraction times, suggesting that contraction frequency is closely associated with both signal complexity and temporal characteristics. These correlation patterns provide valuable insights into potential feature redundancy while maintaining clinical interpretability, as the relationships align with established obstetric knowledge about uterine activity dynamics and support evidence-based dimensionality reduction strategies if needed for model optimization.
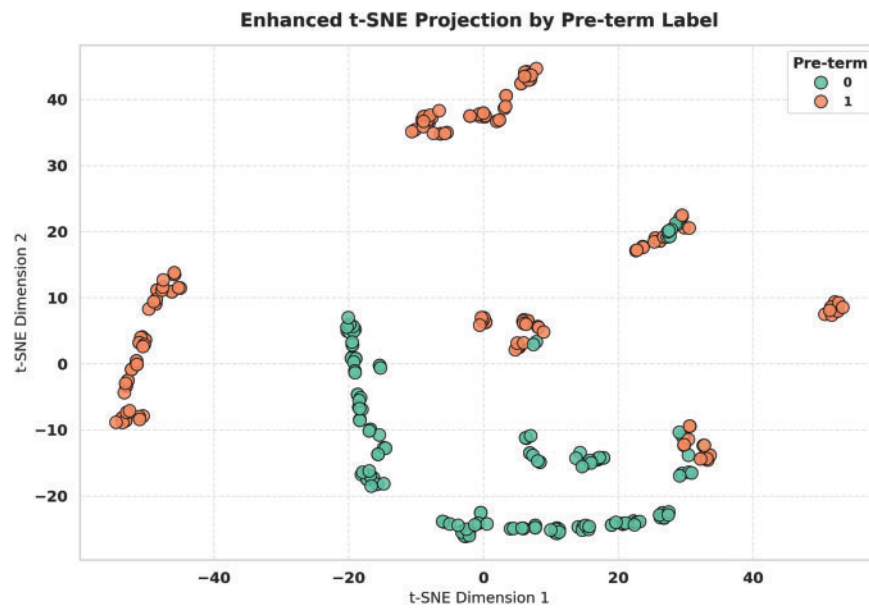
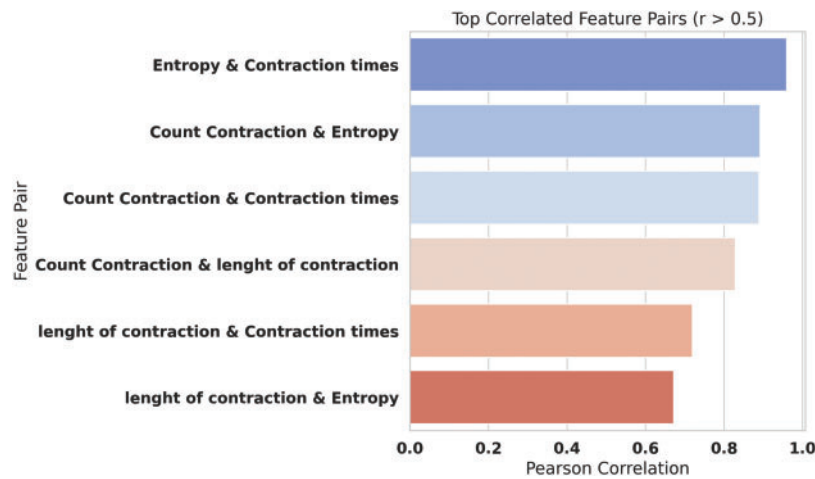**Figure 11:** t-SNE clustering visualization revealing non-linear patterns and distinct class groupings

Figure 12: Top correlated feature pairs analysis identifying relationships between EHG signal characteristics (r < 0.5)
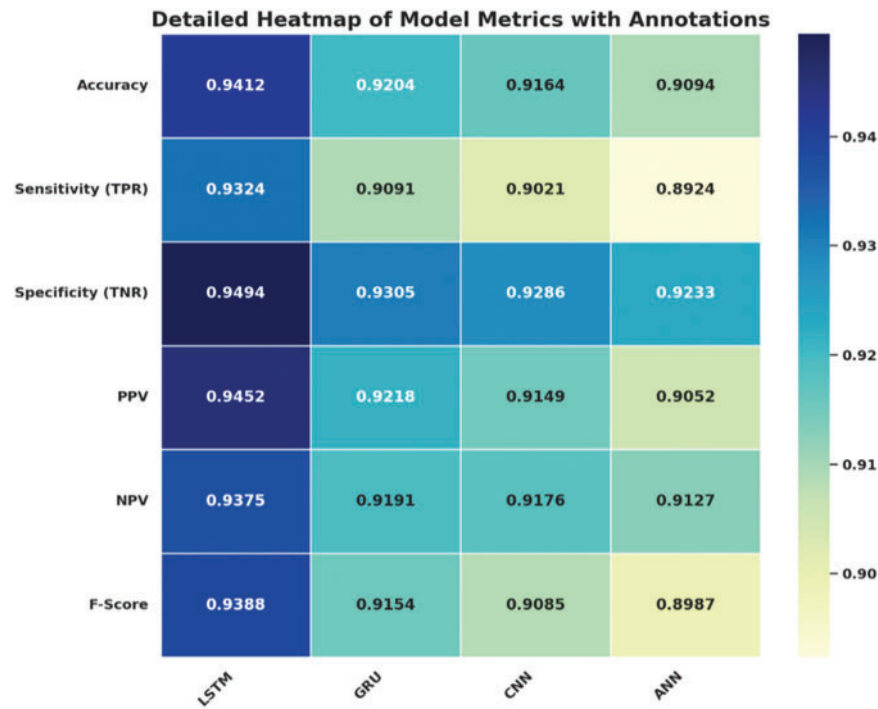
4.2 Machine Learning Models Results

The performance of the best model, the Long Short-Term Memory (LSTM) model, is an accuracy of (0.94118), and the sensitivity True Positive Rate (TRP) of (0.93243) is depicted in Table 3 below. It was found that the LSTM model outperformed the rest and could classify positive instances effectively with high sensitivity. Further, evidence of high True Negative Rate (TNR) and Positive Predictive Value (PPV) holds good specificity and less chance of having false positives in the result. It seems that the proposed LSTM model is very useful in situations where classification is critical, and the model needs to be stable and reliable. This optimal performance across several evaluation measures indicates that it has a good generalization ability.

Table 3: Machine learning models performance evaluation

Models	Accuracy	Sensitivity (TPR)	Specificity (TNR)	PPV	NPV	F-Score
LSTM	0.94118	0.93243	0.94937	0.94521	0.93750	0.93878
Gated recurrent units (GRUs)	0.92037	0.90909	0.93052	0.92179	0.91912	0.91540
CNN	0.91640	0.90210	0.92857	0.91489	0.91765	0.90845
ANN	0.90938	0.89235	0.92334	0.90517	0.91274	0.89872

Fig. 13 provides the heatmap of model metrics with proper annotations as can be observed LSTM model has higher scores than others in all evaluation criteria. It is the one that recognizes the highest accuracy, which makes it the best in the classification and the most reliable. In this work, LSTM appears most sensitive, measured by TPR, showing its effectiveness in pinpointing positive cases and specific or with high TNR in discerning negative cases. Furthermore, a high PPV means a very low chance of misdiagnosis, while a high NPV means a very low chance of confined false negatives. The F-Score provides one more piece of evidence for the skillful balance of the metrics as the model maintains high equalized values for precision and recall, indicating its superiority over other models being examined.

Figure 13: Heatmap of model performance metrics with annotations

According to the evaluation of the machine learning models, it is observed that the LSTM model represents immense reliability and has performed better in most of the evaluation parameters. This figure shows that it is effective in classifying instances correctly, as can be seen by its high accuracy. The sensitivity (TPR) part determines how effectively the model is able to find true positive instances. The robust specificity (TNR) aspect of the proposed model can effectively account for the strong negative predictive accuracy, thus

minimizing false positives. In addition, an increased value for PPV and NPV highlights higher accuracy in terms of predictions; all positive cases and, at the same time, all negatives are well distinguished. The F-Score further equalizes the LSTM model for the normal distribution of performance concerning precision and recall, especially with relation to data variance.

4.3 Optimization Results

In Table 4 below, the results of the best model, GGPSO + LSTM, are represented in terms of Accuracy (0.97341) and Sensitivity (True Positive Rate) (0.96910). This model was better in all aspects because it gave better Area Under the Curve (AUC), better kappa value, less standard error, fewer false positives, and better sensitivity. The model also tested high reliability in terms of specificity (TNR) and positive predictive value (PPV) which strengthens the elimination of true positives and negatives among the study. These outcomes show that the proposed GGPSO optimization successfully improved the LSTM model performance. Due to its generalize performance and high efficiency, it is very suitable for classification/Input/Output problems.

Table 4: Optimization results with LSTM models

Models	Accuracy	Sensitivity (TPR)	Specificity (TNR)	PPV	NPV	F-Score
GGPSO + LSTM	0.97341	0.96910	0.97741	0.97549	0.97150	0.97229
GA + LSTM	0.96311	0.95745	0.96816	0.96397	0.96236	0.96070
PSO + LSTM	0.96131	0.95459	0.96694	0.96028	0.96217	0.95743
GWO + LSTM	0.95744	0.94880	0.96440	0.95551	0.95897	0.95214

Fig. 14 provides the detailed heatmap of model metrics with annotations as follows: the best overall output is generated from the GGPSO + LSTM model. This decreases variability while at the same time increasing the precision, thus making this model achieve the highest accuracy hence thus being in a position to classify instances the most. The first, TPR (True Positives Rate), communicates how well it is doing in correctly identifying cases that are positive. While the second, TNR (True Negative Rate), shows how accurately it is in identifying negative cases. Further, the PPV and NPV validate the model's accuracy for managing affirmative and negative prognoses.

Fig. 15 presents the accuracy performance comparison of LSTM models optimized by different algorithms for preterm birth prediction. The results demonstrate the superiority of GGPSO optimization, achieving 97.34% accuracy compared to GA+LSTM (96.31%), PSO+LSTM (96.13%), and GWO+LSTM (95.74%). This consistent performance advantage validates the effectiveness of the hybrid GGPSO approach in enhancing LSTM model capabilities for clinical applications.

The optimization results focus on the great success of the integration between GGPSO and the LSTM model. This has acted as proof of better accuracy since it has exhibited high instance classification credentials. It gives a clear picture of the true positive rate (TPR) model efficiency to detect true cases, and the non-overlapping false positive rate (FPR) indicates negative instances with high specificity (TNR). Also, the rise of the PPV and NPV value goes on to show the effectiveness of the model towards the outcome that was predicted in the positive and negative hypothesis. Thus, as the F-Score values are balanced, it indicates this approach is well-prepared to maintain both precision and recall constantly. These outcomes confirm the efficiency of the proposed strategy of employing GGPSO to improve LSTM, which undoubtedly identifies LSTM as one of the most effective techniques suited for classification problems with high precision, for which acceptable complex level may cause.
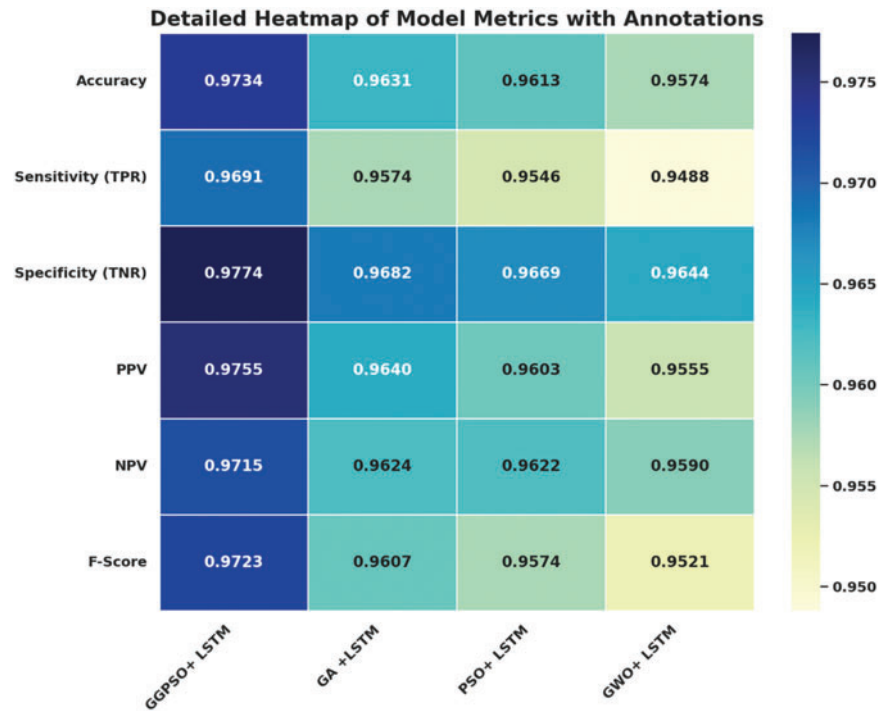
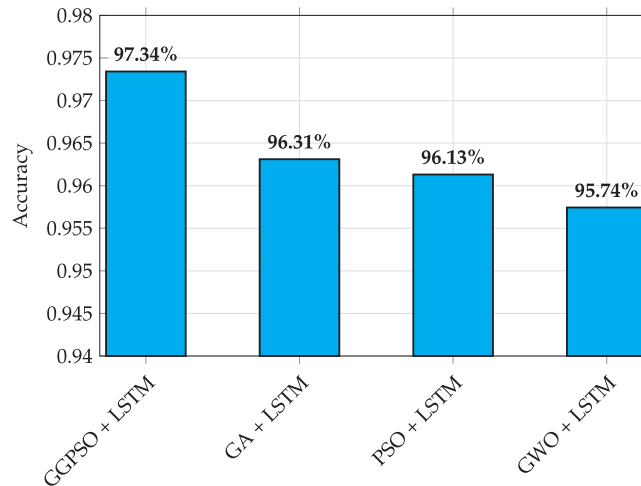**Figure 14:** Heatmap of optimized LSTM model metrics with annotations

Figure 15: Accuracy performance comparison of LSTM models with different optimization algorithms

5 Discussion

The findings of this study highlight the considerable potential of hybrid metaheuristic optimization in enhancing deep learning models for biomedical prediction tasks, specifically preterm birth classification. The proposed GGPSO-optimized LSTM model achieved a notably high accuracy of 97.34%, outperforming both traditional machine learning methods and existing optimized deep learning models reported in the literature. Previous studies on similar Electrohysterogram (EHG) datasets typically report accuracy values in the range of 0.78 to 0.90 [13,16,22], indicating a substantial improvement attributable to our hybrid approach.

This increased performance can be explained by the fact that the GGPSO algorithm successfully strikes the balance between exploration and exploitation in the hyperparameter search space. Whereas traditional optimizers such as GA and PSO have been known to either early converge or become unable to get out of local optima, the GGPSO algorithm has been able to couple the benefits of the social dynamics of Greylag Goose Optimization with the adaptiveness of Particle Swarm Optimization in search. The result is more effective training of LSTM networks and better convergence of models, which can be achieved by this hybrid architecture that enables strong global search with fine-tuned local adjustment.

In addition to raw performance, one of the most critical contributions of this study is model interpretability. By employing clinically relevant feature extraction from EHG signals—including entropy, contraction count, and contraction duration—we ensured that the input space reflects meaningful physio-logical information. Our SHAP and Random Forest analyses confirmed that entropy and contraction-related variables were the most influential features, findings that are consistent with obstetric literature describing uterine behavior in preterm labor [15,17]. This strengthens the clinical validity of our model and promotes transparency, addressing one of the common shortcomings of deep learning systems in healthcare.

From a data perspective, this study successfully overcame the challenge of dataset imbalance, a common issue in medical machine learning. The use of SMOTE to balance the minority class expanded the original dataset from 58 to 312 samples, enabling the model to learn equally from both preterm and non-preterm cases. This approach, coupled with repeated cross-validation, demonstrates how even small, imbalanced datasets can be leveraged effectively with proper preprocessing and evaluation strategies.

The clinical implications of these results are significant. High sensitivity (96.91%) ensures that most at-risk pregnancies are correctly flagged, enabling timely interventions such as corticosteroid administration or NICU preparation. Simultaneously, the high specificity (97.74%) minimizes false alarms, thereby reducing unnecessary interventions in low-risk cases. These performance metrics are well-balanced, enhancing the model's practical utility in real-world clinical decision-making where both false positives and false negatives carry substantial consequences.

However, several limitations need to be mentioned. The small sample size is a drawback, despite the SMOTE technique's ability to reduce it. However, external validation of the suggested model on broader and more heterogeneous information is necessary to demonstrate the model's generalizability. The EHG signal features employed in this work were good but restricted to only a few statistical and temporal descriptors; further predictive accuracy could be achieved by the inclusion of other signal modalities (e.g., maternal vitals, biochemical markers). Additionally, we have introduced SHAP to achieve explanations. Still, one should continue the study using more advanced explainable AI methods (e.g., integrated gradients or counterfactual analysis) to have a more insightful view of the logic behind model decisions.

This work addresses gaps in the literature compared to previous studies. The majority of current research above is either non-complex optimizing or compromised in terms of interpretation, with generated output. Our hybrid model has been constructed, proving that both objectives can be fulfilled simultaneously. It also points to the little-studied opportunity of nature-inspired algorithms in a time-series classification problem, such as in medicine, in which more traditional optimizers have dominated. This contribution is a methodological novelty that can serve not only to the area of understanding prediction of preterm birth but also to a more general area of intelligent clinical decision support.

To conclude, the given research provides a structurally integrated methodology that strikes the balance of predictive performance, transparency of the model, and clinically relevant conclusions. It establishes a precedent for new future studies into the implementation of hybrid metaheuristic-deep learning based models in the maternal-fetal medium as well as in other fields. The following steps to practical application

in obstetric care will be the validation of the model in the real-time monitoring settings, as well as in a diverse population.

6 Conclusion and Future Directions

The study focuses on how advanced machine learning techniques, which include the LSTM model, can result in high classification performance as seen on the various evaluation standards. LSTM-optimized networks using the GGPSO optimization provided a boost to the model, thereby corresponding to higher accuracy, sensitivity, and specificity. These enhancements argue the role of the optimization process in enhancing the quality of the models to provide accurate and precise results when working with complicated classification problems. The results obtained prove the possibilities of integrating machine learning models with optimization algorithms to solve issues in various fields of application when accurate and stable classification is crucial. Because of the high precision and reliability of the proposed algorithm, the GGPSO-optimized LSTM is suitable for implementation on actual sites.

In further studies, various avenues can be taken to expand and build on the already existing framework. One potential avenue involves an investigation of any alternative or higher-order metaheuristic optimization algorithms, including differential evolution, ant colony optimization, or novel hybrid swarm intelligence models, to determine whether they provide better convergence speed or predictive performance than GGPSO. In addition, heterogeneous hybrid optimization methods that can take advantage of many algorithms can be applied to improve the robustness and flexibility of the hyperparameter tuning process. The next significant direction is verifying the proposed model on the extended, multi-center data, which has diverse demographic and clinical features, as it is necessary to check its generality and stability factors in real-world conditions of the healthcare setting. Moreover, although preliminary interpretability has been fulfilled, further research should utilize more elaborate explainable AI methods, like SHapley Additive Explanations (SHAP), LIME, or integrated gradients, to ensure more transparency on the decision paths used by the model and enable the acceptance of the model by clinicians. Lastly, the actual performance of the model in real-life applications—such as continuous fetal monitoring systems/mobile health platforms development—will be essential to assess its performance in a dynamic environment, as well as ensure that the model is successful in adoption to practical, time-sensitive clinical applications.

Although the outcomes of the study are encouraging, it cannot be said that there are no limitations that should be considered. Whereas the proposed GGPSO-optimized LSTM model demonstrates a high level of predictive potential, its present assessment is limited to a relatively small pool of data and an experimental environment. In this regard, additional verification in multifaceted clinical settings, patient populations, as well as health care systems is needed to evaluate the robustness and generalizability of the model. Besides, although attempts to address interpretability were made earlier, a detailed analysis of the decision-making process in the model by applying the more sophisticated explainable AI methodologies of the second generation could also enhance clinical trust and transparency. Lastly, testing the model's performance in real-time, dynamic clinical pathways remains a research priority to achieve practical relevance within the clinical context.

Author Contributions: Methodology, Anis Ben Ghorbal and Azedine Grine; Software, Marwa M. Eid and El-Sayed M. El-Kenawy; Investigation, Anis Ben Ghorbal, Marwa M. Eid, and El-Sayed M. El-Kenawy; Writing—original draft, Anis Ben Ghorbal and Marwa M. Eid; Writing—review & editing, Azedine Grine, Marwa M. Eid, and El-Sayed M. El-Kenawy; Supervision, El-Sayed M. El-Kenawy; Project administration, El-Sayed M. El-Kenawy. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are openly available in Kaggle at https://www.kaggle.com/datasets/ahmadalijamali/preterm-data-set/data (accessed on 15 May 2025).

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Olack B, Santos N, Inziani M, Moshi V, Oyoo P, Nalwa G, et al. Causes of preterm and low birth weight neonatal mortality in a rural community in Kenya: evidence from verbal and social autopsy. BMC Pregnancy Childbirth. 2021;21(1):536. doi:10.1186/s12884-021-04012-z.

2. Noroña-Zhou AN, Ashby BD, Richardson G, Ehmer A, Scott SM, Dardar S, et al. Rates of preterm birth and low birth weight in an adolescent obstetric clinic: achieving health equity through trauma-informed care. Health Equity. 2023 12;7(1):562–9. doi:10.1089/heq.2023.0075.

3. Singer D, Pauline Thiede L, Perez A. Adults born preterm. Deutsches Ärzteblatt International. 2021;118(31–32):521–7.

4. Vanes L, Fenn-Moltu S, Hadaya L, Fitzgibbon S, Cordero-Grande L, Price A, et al. Longitudinal neonatal brain development and socio-demographic correlates of infant outcomes following preterm birth. Dev Cogn Neurosci. 2023;61:101250. doi:10.1016/j.dcn.2023.101250.

5. Gandhi A, Adhvaryu K, Poria S, Cambria E, Hussain A. Multimodal sentiment analysis: a systematic review of history, datasets, multimodal fusion methods, applications, challenges and future directions. Inf Fusion. 2023;91(3):424–44. doi:10.1016/j.inffus.2022.09.025.

6. Singha S, Arha H, Kar AK. Healthcare analytics: a techno-functional perspective. Technol Forecast Soc Change. 2023;197(6):122908. doi:10.1016/j.techfore.2023.122908.

7. Naser MZ. An engineer's guide to eXplainable Artificial Intelligence and Interpretable Machine Learning: navigating causality, forced goodness, and the false perception of inference. Autom Constr. 2021;129(1):103821. doi:10.1016/j.autcon.2021.103821.

8. Jager F, Libenšek S, Geršak K. Characterization and automatic classification of preterm and term uterine records. PLoS One. 2018;13(8):e0202125. doi:10.1371/journal.pone.0202125.

9. Włodarczyk T, Płotka S, Szczepański T, Rokita P, Sochacki-Wójcicka N, Wójcicki J, et al. Machine learning methods for preterm birth prediction: a review. Electronics. 2021;10(5):586. doi:10.3390/electronics10050586.

10. Lameesa A, Hoque M, Alam MSB, Ahmed SF, Gandomi AH. Role of metaheuristic algorithms in healthcare: a comprehensive investigation across clinical diagnosis, medical imaging, operations management, and public health. J Comput Design Eng. 2024;11(3):223–47. doi:10.1093/jcde/qwae046.

11. Shami TM, El-Saleh AA, Alswaitti M, Al-Tashi Q, Summakieh MA, Mirjalili S. Particle swarm optimization: a comprehensive survey. IEEE Access. 2022;10(4):10031–61. doi:10.1109/access.2022.3142859.

12. Keyimu R, Tuerxun W, Feng Y, Tu B. Hospital outpatient volume prediction model based on gated recurrent unit optimized by the modified cheetah optimizer. IEEE Access. 2023;11(11):139993–40006. doi:10.1109/access.2023.3339613.

13. Goldsztejn U, Nehorai A. Predicting preterm births from electrohysterogram recordings via deep learning. PLoS One. 2023;18(5):e0285219. doi:10.1371/journal.pone.0285219.

14. Nsugbe E, Reyes-Lagos JJ, Adams D, Samuel OW. On the prediction of premature births in Hispanic labour patients using uterine contractions, heart beat signals and prediction machines. Healthc Technol Lett. 2023;10(1–2):11–22. doi:10.1049/htl2.12044.

15. Park S, Moon J, Kang N, Kim YH, You YA, Kwon E, et al. Predicting preterm birth through vaginal microbiota, cervical length, and WBC using a machine learning model. Front Microbiol. 2022;13:912853. doi:10.3389/fmicb.2022.912853.

16. Selvaraju V, Karthick PA, Swaminathan R. Detection of preterm birth from the noncontraction segments of uterine emg using hjorth parameters and support vector machine. J Mech Med Biol. 2023;23(6):2340014. doi:10.1142/s0219519423400146.

17. Kavitha SN, Asha V. Predicting risk factors associated with preterm delivery using a machine learning model. Multimed Tools Appl. 2024;83(30):74255–80. doi:10.1007/s11042-024-18332-7.

18. Díaz E, Fernández-Plaza C, Abad I, Alonso A, González C, Díaz I. Machine learning as a tool to study the influence of chronodisruption in preterm births. J Ambient Intell Humaniz Comput. 2022;13(1):381–92. doi:10.1007/s12652-021-02906-6.

19. Allahem H, Sampalli S. Automated labour detection framework to monitor pregnant women with a high risk of premature labour using machine learning and deep learning. Inform Med Unlocked. 2022;28(3):100771. doi:10.1016/j.imu.2021.100771.

20. León C, Cabon S, Patural H, Gascoin G, Flamant C, Roué JM, et al. Evaluation of maturation in preterm infants through an ensemble machine learning algorithm using physiological signals. IEEE J Biomed Health Inform. 2022;26(1):400–10. doi:10.1109/jbhi.2021.3093096.

21. Peng Z, Varisco G, Long X, Liang RH, Kommers D, Cottaar W, et al. A continuous late-onset sepsis prediction algorithm for preterm infants using multi-channel physiological signals from a patient monitor. IEEE J Biomed Health Inform. 2023;27(1):550–61. doi:10.1109/jbhi.2022.3216055.

22. Tian D, Lang ZQ, Zhang D, Anumba DO. A filter-predictor polynomial feature based machine learning approach to predicting preterm birth from cervical electrical impedance spectroscopy. Biomed Signal Process Control. 2023;80(10):104345. doi:10.1016/j.bspc.2022.104345.

23. Prats-Boluda G, Pastor-Tronch J, Garcia-Casado J, Monfort-Ortíz R, Perales Marín A, Diago V, et al. Optimization of imminent labor prediction systems in women with threatened preterm labor based on electrohysterography. Sensors. 2021;21(7):2496. doi:10.3390/s21072496.

24. Preterm EDA and classification [Internet]. [cited 2025 Apr 13]. Available from: https://kaggle.com/code/parmajha/preterm-eda-and-classification.

25. Manohar B, Das R, Lakshmi M. A hybridized LSTM-ANN-RSA based deep learning models for prediction of COVID-19 cases in Eastern European countries. Expert Syst Appl. 2024;256(7):124977. doi:10.1016/j.eswa.2024.124977.

26. El-kenawy ESM, Khodadadi N, Mirjalili S, Abdelhamid AA, Eid MM, Ibrahim A. Greylag Goose Optimization: nature-inspired optimization algorithm. Expert Syst Appl. 2024;238(22):122147. doi:10.1016/j.eswa.2023.122147.

27. Sun Y, Xue B, Zhang M, Yen GG, Lv J. Automatically designing CNN architectures using the genetic algorithm for image classification. IEEE Trans Cybern. 2020;50(9):3840–54. doi:10.1109/tcyb.2020.2983860.

28. Mirjalili S. Genetic algorithm. In: Evolutionary algorithms and neural networks. Cham, Switzerland: Springer International Publishing; 2019. p. 43–55. doi:10.1007/978-3-319-93025-1_4.

29. Mittal N, Singh U, Sohi BS. Modified grey wolf optimizer for global engineering optimization. Appl Comput Intell Soft Comput. 2016;2016(1):7950348. doi:10.1155/2016/7950348.