Nat. Volatiles & Essent. Oils, 2021; 8(4): 14269-14283

Extraction Of Features Of Uterine Emg Signals And Their Correlation

Suma K V[1,*], Mamtha Mohan[1] , Pooja[1]

[1]Department of Electronics and Communication, Ramaiah Institute of Technology, Bengaluru, India

Abstract:

Prediction of premature labor is of great significance in prevention of deaths of infants, or the risk of health ensuing. The uterine Electromyography signals have been very encouraging in the study of uterine contractions. Here, we have considered TPEHG DB (Term-Preterm Electrohysterogram Database) dataset having 300 records, of which 262 are term records while 38 are preterm records. Initially the raw uterine EMG signal is pre-processed and then various statistical,non-linear and linearfeatures are extracted. The features extracted are applied to different machine learning classifiers. Further, Bayesian Hyperparameter Optimization technique was employed on these classifiers to improve their classification accuracy. Support vector machine (SVM) classifier with Bayesian Hyperparameter Optimization technique was tested by employing a 10-fold cross-validation. This was conducted on 38 preterm records and it deliveredaccuracy of 96.667%. This is useful in early detection of preterm delivery in pregnant women and helps in avoiding infant fatality.

Keywords:uterine Electromyography signals, Bayesian Hyperparameter Optimization, Electrohysterogram, pre-term delivery, correlation

1. Introduction

Uterine EMG signals (Electro Hysterogram signals) refers to the electrical activity that occurs in the uterus during pregnancy. These are extracted using surface EMG electrodes in a frequency range of 0-5 Hz. The depolarization and repolarization of myometrial smooth cells induces a burst of action potentials. Each contraction is associated with a burst action as shown in Figure 1.

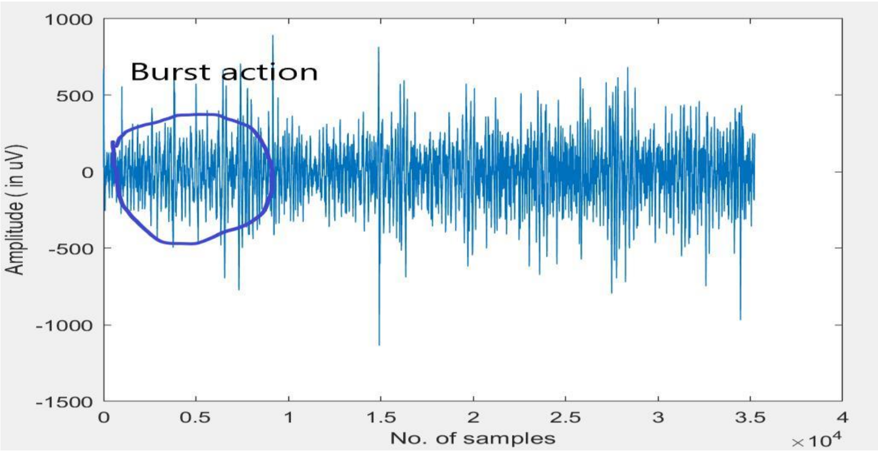

Figure 1: Uterine EMG signal showing Burst action

Thus, the generation and transmission of these burst through junction gaps will increase the electrical activity especially in the last trimester. The frequency, burst action potential duration and total cells simultaneously active is directly proportional to the frequency, time and amplitude of the contraction. Predominantly, the energy of EHG signals lies in 0.2 - 3 Hz band with an amplitude of 100µV to 1.8mV. These EHG signals are useful in determining the information of the uterine contractions. The labor starts before 24 hours for Term delivery and 7-10 days prior for the preterm delivery. Hence there is considerable difference in this EHG activity.

The Uterine EMG activity can be explained by using two components. A low wave and a fast wave. The fast wave is further divided into : Fast Wave Low (FWL) ranging 0.2 - 0.45 Hz and Fast Wave High (FWH) ranging 0.8 - 3 Hz. The FWL is related to propagation of EHG and FWH is related to uterine cell excitation [4]. During Term and Preterm labor, there are frequent uterine EMG bursts activity, large amplitude, pain sensation and drastic changes in intrauterine pressure.

The increased electrical transmission leads to coupling between the cells which will result in better synchronization and coordination among the contractions of uterus. Hence, larger the uterine area synchronized with the burst activity, larger the uterine contractions leading to labor. Hence, predicting the extent of propagation of contractions in the uterus becomes important.

2. Literature Survey

V Selvaraju et.al have consideredthirteen term and thirteen preterm delivery signals from Term-Preterm Electrohysterogram dataset with Tocogram of physionet. Third channel was used for recording the Signals reported artifactfree.uEMG burst signals were used to obtain the analytic signals in the complex plane.The number of bursts from Term and preterm recorded were 53 and 47 respectively.

T AnanthaBabu et. al has worked and extracted features using uterine magnetomyography (MMG).24 MMG signals were employed. The evaluation of the features with Naïve-Bayes, k-Nearest Neighbor

(KNN) and Support Vector Machine (SVM) classifiers were employed to do the classification of MMG signals . KNN classifier confirmed better performance amid the other classifiers.

Punitha. N and Ramakrishnan. S have analysed the preterm condition using EMG signals. Digital Butterworth filter is used to preprocess the signals. Multifractal features namely maximum singularity exponent, peak singularity exponent, strength of multifractality and exponent index are extracted by subjecting the preprocessed signals to MFDFA. Comparing the features they found coefficient of variation is lower for peak singularity exponent.

P. ShanibaAsmi et.al The electrohysterogram (EHG) or uterine electromyogram (Uterine EMG), collected from the abdominal surface is considered as a biomarker for the prediction or preterm labor. fourth order band pass filter was used . Four classifiers were used to two fractal features ,Higuchi Fractal dimension(HFD) and Detrended Fluctuation Analysis (DFA). Elman neural network classifier was found to be the best classifier with classification accuracy (95.7989%) was obtained . Out of 78 labours (women) 38 preterm records and 38 term records were used. The window size for segmentation used was 1200 and overlap was 75% for comparison, after filtering with Butter worth and Elliptic filter respectively features were extracted.

Lili Chen et.al employed Hilbert-Huang transform (HHT) and extreme learning machine (ELM) to do the feature extraction and classification of EHG between pregnancy and labour group. of IMF1 with ELM features delivered higher classification of labour and pregnancy.

3. Methodology

The methodology of this work consists of the following phases.

i. **Data collection:** Using the freely available Physio Net TPEHG (Term-Preterm Electro hysterogram) dataset. This dataset consists of a three-channel uterine EMG records of 300 patients (including both Term and Preterm).
ii. **Signal Processing:** Pre-processing of the signals to remove the unwanted noise. The raw signal is filtered in a particular frequency band and used further for analysis.
iii. **Feature Extraction:** Extracting the relevant information from the signals to be used for classification.

Uterine EMG Dataset Description:

The dataset employed in this study is Term-Preterm Electrohysterogram(TPEHG). It consists of 300 uterine EMG records. These recordings were performed at Obstetrics and GynecologyDepartment, Medical Centre Ljubljana, Ljubljana and they can be accessed freely from Physio Net.The records were collected from the patients diagnosed for preterm labour and also from the general patients. In this dataset, there are total of 300 recordings. Recording was done for a duration of aboutthirty minutes, with 20Hz sampling rate (fs) and a16-bit resolution with amplitude of ±2.5mV. a three-pole Butterworth filter was employed before sampling and it had frequency range of 0 - 5 Hz.

The total 300 EHG records consists of 38 Preterm records (gestation duration at delivery ≤ 37 weeks) and 262 Term (gestation duration at delivery >37 weeks). These records were further divided based upon the week of gestation at the time of recording as shown in figure 2.

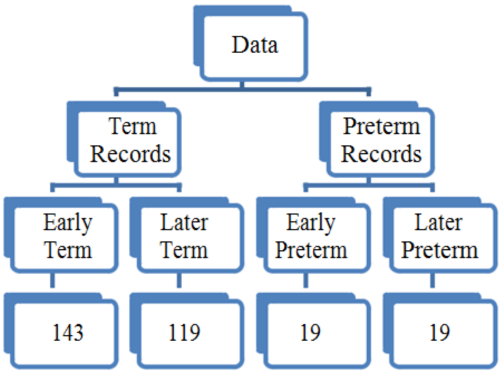

Figure 2: Classification of TPEHG dataset

Pre-Processing:

The raw EHG signals with lower frequencies always contain unwanted noise due to breathing and stretching of skin while recording and so the data needs to be filtered before analysis. To achieve this, a four-pole digital Butterworth filter is used with bandwidth ranging 0.3Hz – 4Hz and 90 seconds of the beginning and end part of the signal is removed to eliminate the transient effects. The major drawback of using digital Butterworth filter is phase-shifting. Hence to obtain zero phase shift, filtering is done twice in both forward and backward directions. The raw and pre-processed signals of Term and Preterm records are shown in the below figures.

Raw Uterine EMG signals:

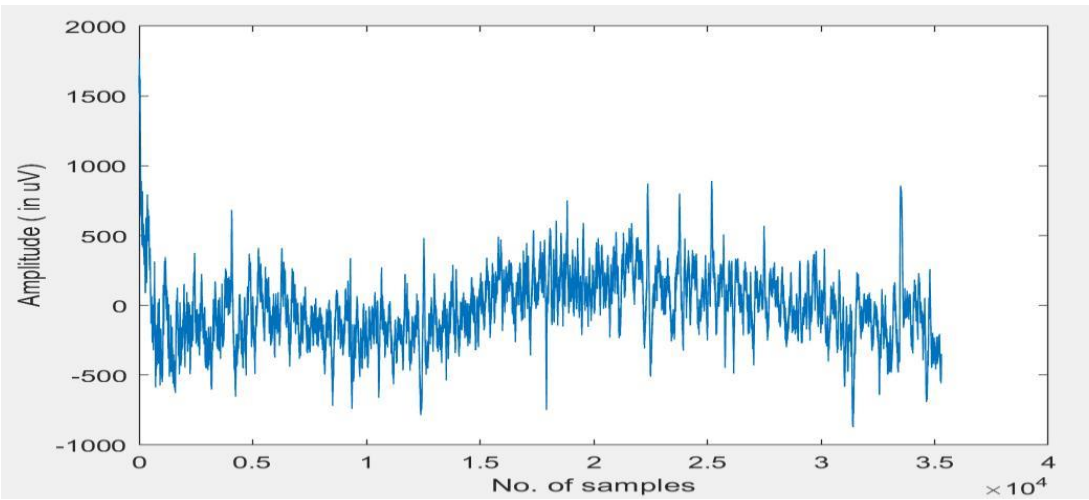

Figure 3: Raw uterine EMG signal of Term Record

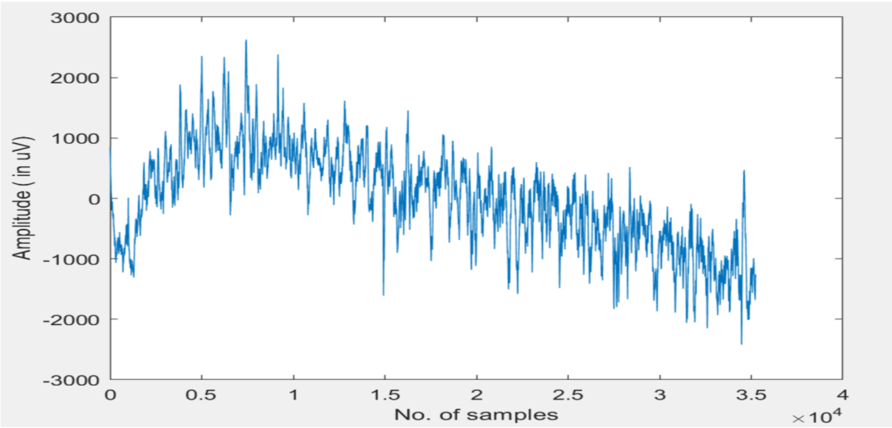

Figure 4: Raw Uterine EMG signal of Preterm Record

Figure 3 and Figure 4 shows the single channel raw uterine EMG signal of Term-Preterm records. X-axis represents the sample size and y-axis being the amplitude in micro volts. The raw uterine EMG is obtained from the Physio Net database.

Filtered Uterine EMG signals:

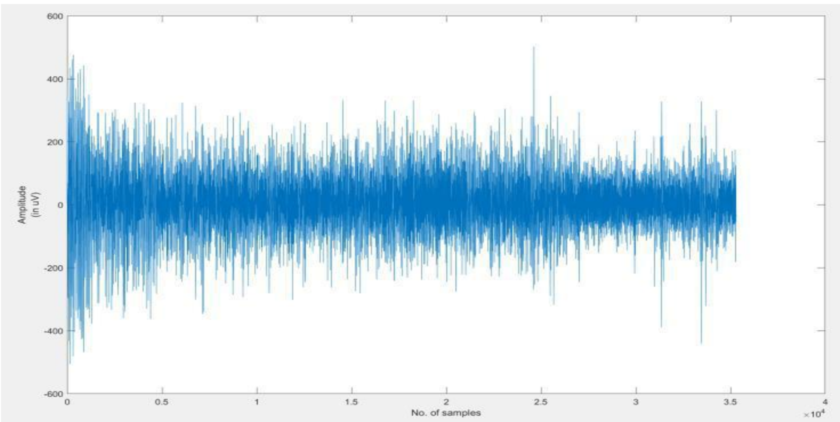

Figure 5: Filtered EHG Signal of Term Record

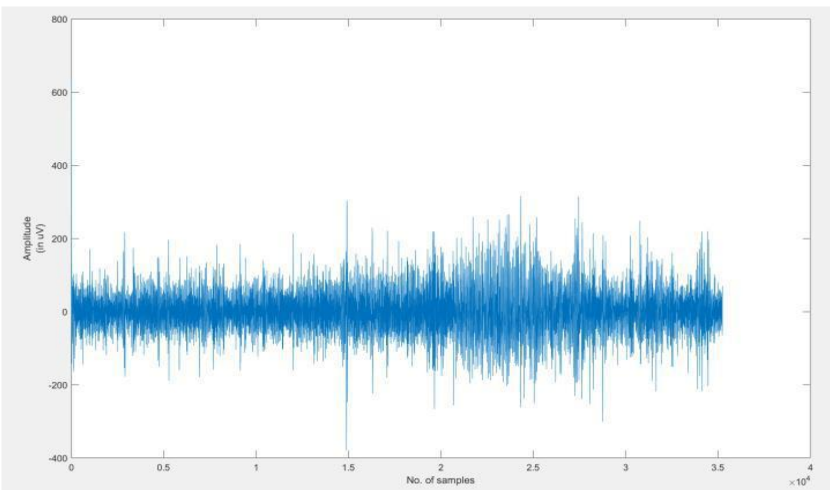

Figure 6: Filtered EHG Signal of Preterm Record

Figure 5 and Figure 6 shows the single channel filtered uterine EMG signals of Term - Preterm record. The signal filtering is done using a four-pole Butterworth filter with bandwidth 0.3 – 4 Hz and sampling frequency 20 Hz. The signal filtering bandwidth range should be kept as short as possible since filtering in the wide range can result in loosing of information and may not provide the required results/feature data in the further steps.

Feature Extraction:

Feature Extraction transforms raw signals into an informative attribute of a specific domain whichaids in combining different classes. This step is most important in pattern recognition. These extracted features must be very effective since the classification accuracy mainly depend on them. For the classifiers to achieve acceptable classification performance, features that are highly correlated with the class value must be chosen. In addressing this issue, a variety of feature extraction techniques are considered

In this study, total of ten features from both frequency domain as well astime domain belonging to statistical, non-linear and linear group are extracted for the sorting of Term-Preterm uterine EMG records.

Linear features:

1: Root mean square (RMS) value

The RMS value is defined as the square root of mean of the square of all the samples in a signal. It is given by equation (1)

RMS= $\sqrt{\frac{1}{N}\sum_{i=0}^{N-1} x(i)^2}$ ....(1)

Where, x(i) = input signal

N= Sample size

2: Peak frequency

The frequency at which the maximum peak occurs is called as the peak frequency. The peak frequency fmax is calculated by equation (2)

$\mathrm{fmax} = \arg\left(\frac{f_S}{N}\max_{i=0}^{N-1} P(i)\right)$.........(2)

Where, fs = sampling frequency

N = sample size

P= frequency power spectrum

3: Median frequency

It is the frequency at which half of the total power within the epoch is reached. It is given by equation (3)

$f_m = i_m \frac{f_s}{N}, \quad \sum_{i=0}^{i=i_m} P(i) = \sum_{i=i_m}^{i=N-1} P(i)$**……….(3)**

Where, fs= sampling frequency

N = sample size

P= frequency power spectrum

4: Waveform length

Waveform length (WL) is the cumulative length of the wave over a period of time. It is given by equation (4)

$wL = \sum_{n=1}^{N-1} |x_{n+1} - x_n|$…………(4)

Where, xn= input signal

N = sample size

5: Zero Crossings

It is the number of times the waveform crosses the zero y-axis.

It is formulated as in equation (5)

$$ZC=\sum_{n=1}^{N-1}[\,\mathrm{sgn}(x_n * x_{n+1}) \cap |x_n - x_{n+1}| > threshold \quad \text{………..(5)}$$

Where, xn= input signal

N = sample size

6: Peak location

Peak location denotes the sample number at which the signal exhibits maximum amplitude.

Non-Linear features:

1: Sample entropy

Sample entropy is used for estimating the complexity of the time series signal. The main advantage of using sample entropy is that it is independent of the length of the data and makes the implementation trouble-free. It is given by equation (6)

$$SampEn=\begin{cases} -\log(c_m)/C_{(m-1):c_m \neq 0 \wedge C_{(m-1)} \neq 0} \\ -\log\left(\frac{N-m}{N-m-1}\right) \quad : \; c_m = 0 \vee c_{(m-1)} = 0 \end{cases} \text{…. (6)}$$

Where, N= sample size

Statistical features:

1. Mean absolute value (MAV)

It is calculated by taking the average of the absolute value of the signal as given by equation (7)

$$MAV=\frac{1}{N}\sum_{n=1}^{N}|x_n| \quad \text{.........................} \quad (7)$$

Where, xn= input signal

N= sample size

2. Variance

The variance is the mean value of the square of the deviation of that variable. It is calculated by equation (8)

$$VAR =\frac{1}{N-1}\sum_{n=1}^{N}x_n^2 \quad \text{........................} \quad (8)$$

Where, xn= input signal

N= sample size

3. Standard Deviation

The standard deviation ~~SD of a signal is~~ formulated as by equation (9)

$$SD =\sqrt{\frac{1}{N-1}\sum_{n=1}^{N}x_n^2} \text{.................................} \quad (9)$$

Where, xn= input signal

N= sample size

Once the features are extracted from the uterine EMG records, the following steps are followed for analysis:

Feature correlation:

Feature correlation is used to understand the relationship between two variables (features). It helps to know the extent of dependency of one feature on the other feature. The correlation can be positive or negative.In positive correlation, with increase in the value of one variable, there is increase in the value of another variable also. It means that both the variables have linear relationship. In negative correlation the variables are inclined to move in opposite direction. Therefore, as the value of one variable increases, then the other variable decreases and vice versa.

The mathematical expression for representing correlation is shown below in equation (10)

$$r = \frac{1}{n-1}\sum\left(xi - \frac{X}{sx}\right)\left(yi - \frac{Y}{sy}\right) \ldots\ldots\ldots\ldots\ldots\ldots.(10)$$

Where, r is the Pearson correlation and 1/(n-1) denotes the covariance.

Where, r > 0 for positive correlation and r < 0 for negative correlation.

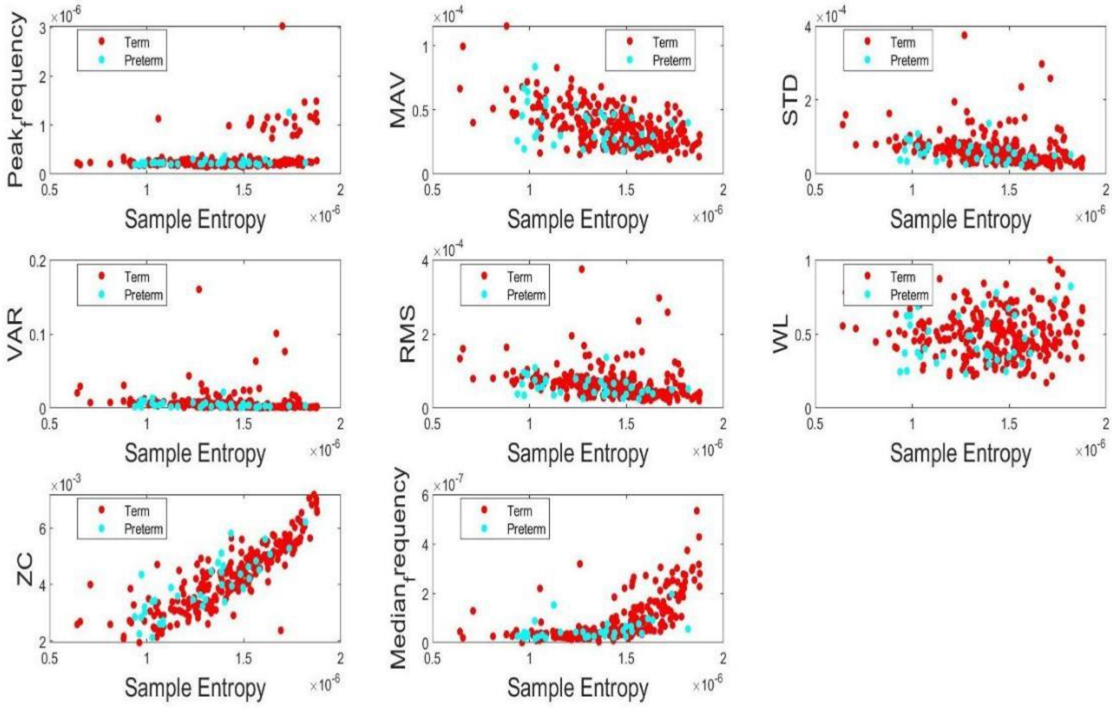

Figure 7: Feature Correlation Plot

The scatter plot of correlation between the features is presented in figure 7. The red colored dots indicate the feature values of Term records and the blue colored dots indicate the feature values of Preterm records. Since the sample entropy being the non-linear feature is efficient in separating the records, all the remaining features are correlated with the sample entropy.The combination of the peak frequency and variance with sample entropy does not correlate much since maximum records are in the range of 1 – 1.5 for sample entropy and 0.1~0.3 for peak frequency. The scatter plot is almost flat for this combination.The correlation of mean absolute value, waveform length, zero crossings and median frequency with sample entropy are positively correlated. This shows the strong linear relationship between these features.

Further, the standard deviation and root mean square feature values with the sample entropy shows almost negative correlation. This indicates a strong weak relationship between them.

Principal Component Analysis (PCA):

Feature reduction is a process of reducing the feature vector dimensions. It helps in avoiding over-fitting. Over-fitting is a problem in large datasets which arise due to missing values in the datasets. Also, over-fitting can also result when the data is not free from redundant values and noise. Hence this leads to results that are not reliable.Therefore, to eliminate feature redundancy, Principal component Analysis (PCA) is used. PCA calculates Eigenvalues, Eigenvectors, and Scores of the features.

The variations in each principle component, with first component being the largest is measured by Eigen values. The combination of original variables and its corresponding Eigenvalue is called Eigen vector and Scores represent the data and illustrates how close the features lie with respect to the first and second principal component.

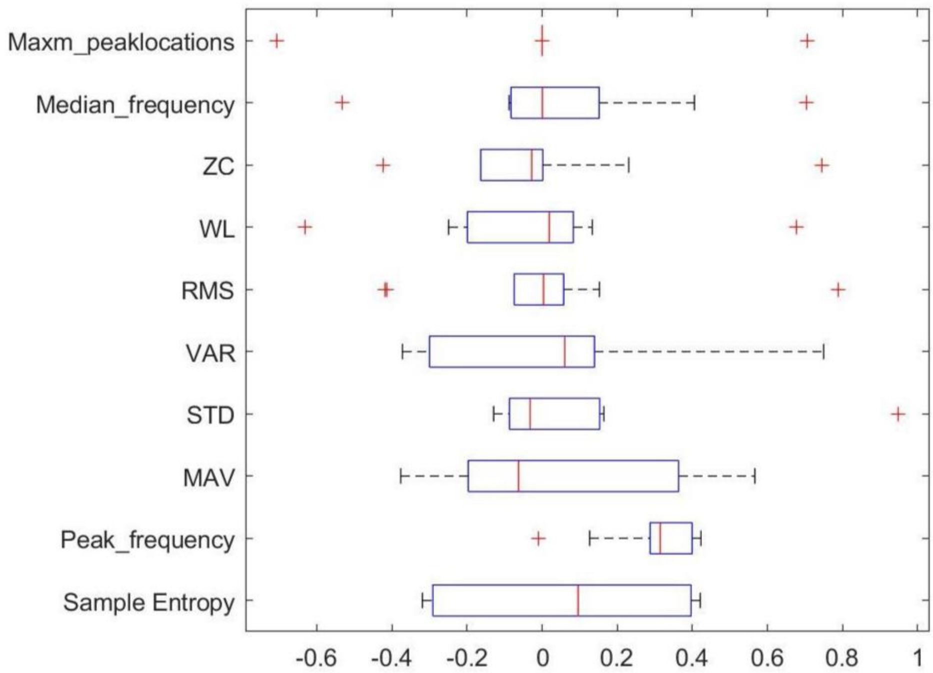

Figure 8: Principal Component Analysis of features

The figure 8 presents the Box plot of PCA of TPEHG dataset. All the features vector dimensions have been reduced within the range of -0.3 – 0.4. The sample entropy plays a significant role since it occupies the largest range of -0.3 – 0.4. Next being the mean absolute value with range -0.2 – 0.39 followed by variance, waveform length, median frequency, standard deviation, zero crossings and root mean square value. Peak frequency occupies lowest dimensional range 0.3 – 0.4. Therefore, PCA increases the interpretability but at the same time minimizes information loss.

Results and discussion:

The relevant features are extracted from the uterine EMG records in the bandwidth range of 0.3 – 4 Hz. Total of 10 features which include statistical, non-linear and linear features are used. The Mean ± Standard deviation values of all the features are tabulated in this section. The feature values of the Term records for the three channels is presented in Table 1.

Table 1: Features of EHG signals of Term Record

Feature name	Channels of Preterm Record		
	CH1	CH2	CH3
Median frequency (Hz)	0.20±0.11	0.15±0.06	0.20±0.09
Maximum Peak location	2126±1248	1671±1043	2023±1192

Root mean square (mV)	91.8±56.4	77.4±47.8	71.2±47.3
Waveform length (m)	0.88±0.29	0.54±0.25	0.56±0.17
Zero crossings	5090.8±1271.9	3921±1153	4886±1220
Peak frequency (Hz)	0.55±0.41	0.38±0.15	0.48±0.34
Mean absolute value	58.83±25	50.5±24.8	40±16.39
Variance	12180±22113	8450±14278	7451±1573
Standard deviation	93.8±54.4	76.4±45.8	70.2±46.3
Sample entropy	1.76±0.28	1.46±0.31	1.74±0.27

Table 1 gives the Mean ± Standard Deviation values of all the features of the 262 Term uterine EMG records.Dispersion of dataset relative to its mean value is given by the Mean ± Standard Deviation value. The above features are extracted in the frequency range 0.3 – 4 Hz.From the Table.4.1, it can be inferred that all the feature values for channel 1 are higher than the other two channels. Mean absolute value, variance, standard deviation and Root mean square values of channel 2 are greater than that of channel 3. Channel 3 seems to dominate over channel 2 for the remaining features. The feature values of the Preterm records for the three channels is presented in Table 2.

Table 2 presents Mean ± Standard Deviation values of all the features of 38 Preterm uterine EMG records. Similar to the Term records, these features are also extracted in the frequency range 0.3 – 4 Hz. From the Table 2, it is observed that all the feature values of channel 1 are greater than other two channels. Mean absolute value, Root mean square, standard deviation andvariance, median frequency values and waveform length values of channel 2 are higher than that of channel 3. Channel 3 is seen to dominate over channel 2 for the remaining features.

Table 2: Features of EHG signals of Preterm Record

Feature name	**Channels of Preterm Record**		
	CH1	**CH2**	**CH3**
Median frequency (Hz)	0.18±0.07	0.16±0.05	0.15±0.04
Maximum Peak location	2091±1260	1488±1001	1862±1163
Root mean square (mV)	83.5±35.4	73.2±32.4	64.9±27
Waveform length (m)	0.76±0.26	0.52±0.21	0.51±0.17

Zero crossings	4902±1225	3873±1173	4480±1132
Peak frequency (Hz)	0.44±0.28	0.36±0.05	0.41±0.19
Mean absolute value	56.3±25.5	50±22.2	41.8±16.9
Variance	8380±7424	6460±6055	5110±4646
Standard deviation	84.5±35.5	74.2±31.4	65.9±28
Sample entropy	1.73±0.27	1.45±0.30	1.73±0.26

On comparison, there is a huge difference in peak frequency, root mean square value,standard deviationand variance of the feature values of both the Term and Preterm records. But it is also observed that the median frequency values are sufficient to represent frequency domain characteristics. Sample entropy values for Preterm records are lower than the Term records. The maximum peak location and zero crossing values can be effectively used in classification. Even though the differences in values among all the features are noticeable, they are dispersed.

Conclusion and future work:

In this modern era, health and illness are also one of the factors responsible for possibility of Premature delivery. The f uterine electrical activity has been effectively employed to determine the Preterm birth. The machine learning algorithms are used for extracting the biological patterns of the data. Further, EHG records arefiltered in the band 0.3 – 4 Hz and features are extracted in statistical, non-linear and linear domains. Theseextracted features are fed to 4 classifiers and tested using a 10 - fold cross validation on 38 preterm records. In order to improve the classification accuracy Bayesian Hyperparameter Optimization technique is employed to the machine learning classifiersand it is observed to give accuracy of 96.667%.Thus, the uterine Electromyography signals have shown encouraging results to monitor the Preterm delivery and progress the health of the society.

References:

[1] Lili Chen, YaruHao ,"Feature Extraction and Classification of EHG between Pregnancy and Labour Group Using Hilbert-Huang Transform and Extreme Learning Machine", Hindawi Publishing Corporation Computational and Mathematical Methods in Medicine, Vol. 2017.

[2] P. ShanibaAsmi, KamalrajSubramaniam and Nisheena V. Iqbal," Classification of Fractal Features of Uterine EMG Signal for the Prediction of Preterm Birth", Biomedical & Pharmacology Journal, Vol. 11(1), 369-374 (2018).

[3] Punitha. N and Ramakrishnan. S," Analysis of uterine electromyography signals in preterm condition using multifractal algorithm",ProcInstMechEng H. 2019 Mar;233(3):362-371

[4] T Ananda Babu, Dr. P Rajesh Kumar," Features Extraction and Classification of Uterine Magnetomyography Signals", International Journal of Current Engineering and Scientific Research (Ijcesr) Vol.5(3), 2018

[5] V. Selvaraju, P. Namadurai and R. Swaminathan," Analysis of Electromyography Burst Signals using Topological Feature Extraction for Diagnosis of Preterm Birth", IEEE SPMB 2020 vol.1.0, 2020.

[6] Suma K. V., PunyaPrabha V., Pooja Arun Kumar M., Venkatesh D. (2020). Preterm Delivery Prediction by Analysis of Multi-channel Uterine EMG signals. International Journal of Advanced Science and Technology, Vol. 29, No. 4, (2020), pp.9819 – 9826