Wearable Devices for Real-time Monitoring the Labor: A Potential Predictor of Preterm Labor

Burak Erdil Biçer*, Mehmet Efdal Özel*, İbrahim Burak Karaüzüm*
*Electronics and Communications Engineering, Istanbul Technical University, Istanbul, Turkey

***Abstract*—Preterm labor, occurring prior to 37 weeks of gestation, is a significant global health issue closely related to neonatal mortality and morbidity. Notwithstanding medical advancements, preterm labor statistics have not significantly decreased, partially due to inadequate monitoring techniques. Biosignals, including electrocardiography (ECG), electromyography (EMG), and electroencephalography (EEG), can yield critical information regarding the health of both mother and infant. However, the lack of accessibility, continuity, and real-time monitoring in current methods renders them insufficient for timely interventions. A wearable device capable of constant, real-time, and precise monitoring of these biosignals would mark a breakthrough in obstetrics. This paper explores the limitations of existing preterm labor prediction methods, the potential of fetal heart rate monitoring through non-stress tests, the utility of EMG in diagnosing preterm labor, and the promising use of uterine electrical signals via electrohysterography (EHG). Additionally, the literature review section critically evaluates three recent studies on real-time biosignal monitoring approaches during pregnancy, assessing their merits and shortcomings. The core of this research, however, is the development of a machine learning-based model to predict preterm labor using EMG signals. The approach encompasses data collection, preprocessing, analysis, and the creation of a user-friendly interface for displaying real-time predictions. Ultimately, this study aims to develop a novel, wearable technology for the early detection of preterm labor, potentially transforming prenatal care and significantly reducing neonatal risks associated with preterm births.**

***Index Terms*—Wearable devices, real-time monitoring, labor, preterm labor, biosignals, electromyography (EMG), prenatal care, electrohysterography (EHG), machine learning.**

I. Introduction

Preterm labor, defined as labor before 37 weeks of pregnancy, is a major health concern that poses a significant risk to newborns. It's one of the leading causes of neonatal mortality and morbidity worldwide. Despite advances in obstetric and neonatal care, the incidence of preterm labor remains high. The problem is that current methods of monitoring the biosignals of mother and child during pregnancy are not adequate or accessible for many people. Biosignals are the electrical or mechanical signals that reflect the activity of living cells or tissues, such as the electrocardiography (ECG), electromyography (EMG) and electroencephalography (EEG) as it can be seen from Fig. 1. These signals can provide important information about the health and well-being of both the mother and the baby. However, many people do not have regular access to prenatal care or ultrasound exams that can measure these signals. Moreover, these methods are not continuous or real-time, meaning that they cannot detect sudden changes or emergencies that may occur between visits. This could lead to some baby deaths because of late diagnosis of preterm labor or other complications. Preterm labor is when the baby is born before 37 weeks of pregnancy, which can cause serious health problems for the baby, such as breathing difficulties, bleeding in the brain, infection, or vision loss. Preterm labor can be caused by various factors, such as problems with the uterus, cervix, or placenta, infections, chronic health conditions, or multiple pregnancies. Therefore, there is a need for a better way of monitoring the biosignals of mother and child during pregnancy that is continuous, real-time, accurate, and affordable.

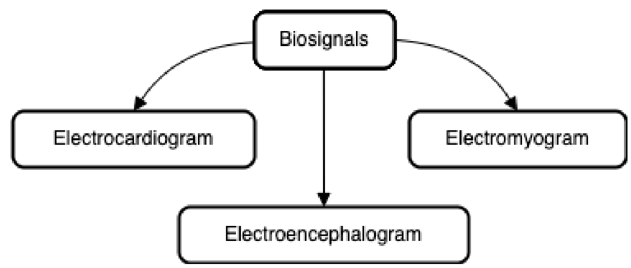


Fig. 1: Overview of the three main types of biosignals: EEG, ECG, and EMG.

The current methods of predicting preterm labor, cervical length measurement and fetal fibronectin testing, provide valuable information in certain situations but have limitations, emphasizing the need for more accurate and real-time wearable devices. Cervical length measurement requires an invasive transvaginal ultrasound, has low sensitivity and its rapid changes can lead to inaccuracies. On the other hand; fetal fibronectin testing, a non-invasive method, has limitations in terms of guaranteeing or ruling out preterm labor and is not recommended for routine use [1]. Thus, a wearable device that offers continuous, accurate monitoring could significantly improve prenatal care and pregnancy outcomes.

Fetal heart rate monitoring is a critical tool for assessing the health of developing fetuses during pregnancy. The fetal heart rate can be an early indicator of preterm labor. One way to predict preterm labor is through non-stress test (NST). An example NST can be observed in Fig. 2. During NST, a device is placed on the mother's abdomen to measure the fetal heart rate and the frequency and duration of uterine contractions. Changes in the fetal heart rate can be a sign of a problem, including changes in fetal oxygen supply that may indicate the onset of preterm labor because in severe cases, a lack of oxygen can lead to fetal death in utero. In cases where fetal distress is detected, preterm delivery may be the best course of action to ensure the health and survival of both the mother and baby. However, NST requires expertise to interpret and is not always accessible. The limitations of NST highlight the need for a tool that's easy to use, accessible and can be used in real-time.

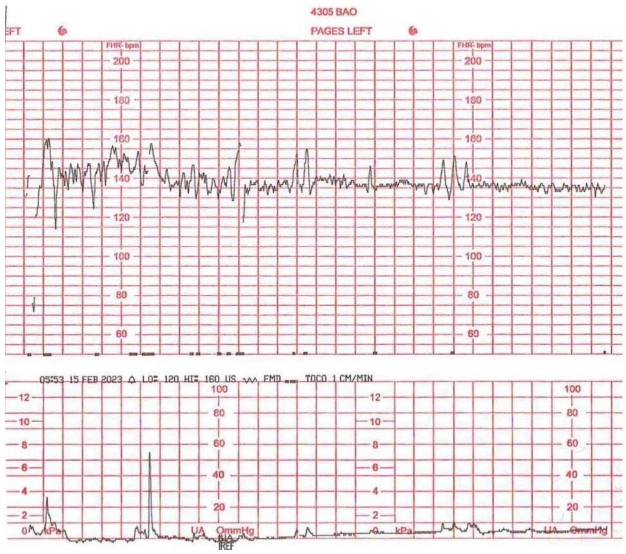


Fig. 2: This figure displays a Non-Stress Test (NST) signal acquired from a pregnant woman at Erciyes Hospital in Kayseri. The variations and patterns shown in this graph provide insights into the fetus's condition and can help healthcare professionals to assess the ongoing pregnancy for potential risks or complications.

Another way to predict the preterm labor is the use of electromyography (EMG), that is a promising diagnostic tool for identifying preterm labor, as it noninvasively records electrical activity of the myometrium, providing real-time data on uterine contractions. Uterine EMG has demonstrated its ability to detect contractions with equal accuracy compared to existing methods. Moreover, alterations in cellular excitability and coupling, essential for efficient contractions leading to delivery, are evident in various EMG parameters. Employing uterine EMG surpasses other currently utilized techniques in clinical settings for identifying patients experiencing genuine labor [2].

Finally, an emerging technique for forecasting preterm deliveries is through the utilization of electrohysterography (EHG), which involves the monitoring of uterine electrical signals. Conventional methods for predicting preterm births have been contingent on subjective assessments, thus, incorporating the evaluation of uterine electrical activity presents a promising avenue for the objective diagnosis of authentic labor and the prediction of preterm deliveries. Leveraging machine learning methodologies that discriminate between term and preterm datasets can further bolster predictive accuracy. Additionally, to address class imbalance, the Synthetic Minority Oversampling Technique (SMOTE) is employed to augment the underrepresented preterm class in the dataset. Moreover, cross-validation procedures are utilized to assess the reliability and validity of the dataset in comparison to analogous studies [3].

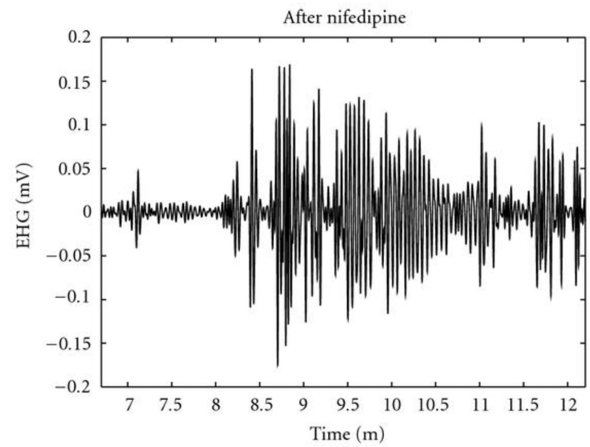


Fig. 3: An example of a preprocessed Electromyography (EMG) signal from the uterus, often referred to as an Electrohysterogram (EHG) [4].

Wearable technology is a promising solution for predicting preterm labor in real-time. The use of wearable technology for predicting preterm labor is a novel approach that has the potential to improve the accuracy and reliability of preterm labor prediction. The development of an easy-to-use, accessible and real-time tool for predicting preterm labor would be a significant advancement in obstetric care. Such a tool could help identify women who are at risk of preterm labor earlier, allowing healthcare providers to take preventive measures to improve neonatal outcomes. By continuously monitoring for preterm labor, healthcare providers can provide appropriate care.

II. Literature Review

The purpose of this literature review is to evaluate the quality and relevance of the existing literature on real-time monitoring of biosignals during pregnancy. This review aims to provide constructive feedback and suggestions for improvement. The scope of this review focuses on three reports that represent different approaches and technologies for real-time monitoring: telemonitoring, wearable technology, and continuous glucose monitoring.

A. Telemonitoring of Pregnant Women at Home

The first report introduces a medical cyber-physical system (MCPS) for telemonitoring of pregnant women at home [5]. The system consists of a body area network (BAN) of sensors that measure fetal heart rate and uterine activity signals, and a personal area network (PAN) that processes and transmits the signals to a central surveillance center. The report focuses on biosignal acquisition and measurement and presents a smart selection of algorithms for abdominal signal analysis in the mobile instrumentation of PAN.

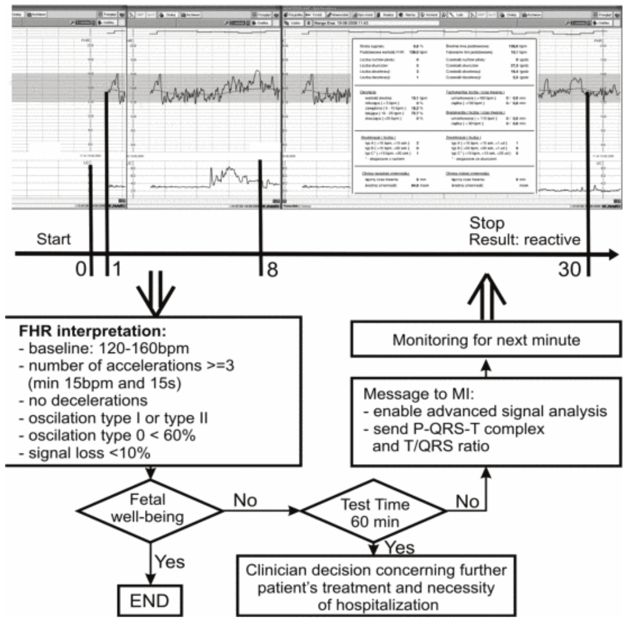


Fig. 4: Overview of fetal health evaluation test during pregnancy suggested in the first report [5].

The strengths of this report lie in its clear and comprehensive description of the system design and implementation. It showcases a high level of technical expertise and innovation in developing a wireless and reliable system for real-time monitoring of biosignals during pregnancy. The potential benefits and applications of the system for improving maternal and fetal health outcomes are also highlighted. However, the report has some weaknesses. It lacks empirical evidence and evaluation of the system's performance and effectiveness. User acceptance and satisfaction, as well as ethical and legal issues related to data privacy and security, are not addressed. Additionally, a comparison with other existing or alternative approaches or technologies is missing.

To improve this report, conducting and reporting a pilot study or a clinical trial with real users and data would be beneficial. Including a user survey or interview to assess user experience and feedback with the system would enhance the report. It should also discuss ethical and legal implications and provide a literature review or background section to situate the system within the context of previous or related work.

B. Wearable Technology for Health Monitoring during Pregnancy

The second report presents a study that evaluates the perception and acceptance of remote fetal ECG monitoring technologies among women of child-bearing age [6]. The study uses a survey method to assess women's willingness, preferences, concerns, and expectations regarding wearable ECG devices. Most women express their willingness to use a wearable ECG device throughout pregnancy to increase the frequency of monitoring maternal and fetal health outside the hospital. Survey result can be observed in Fig. 5.

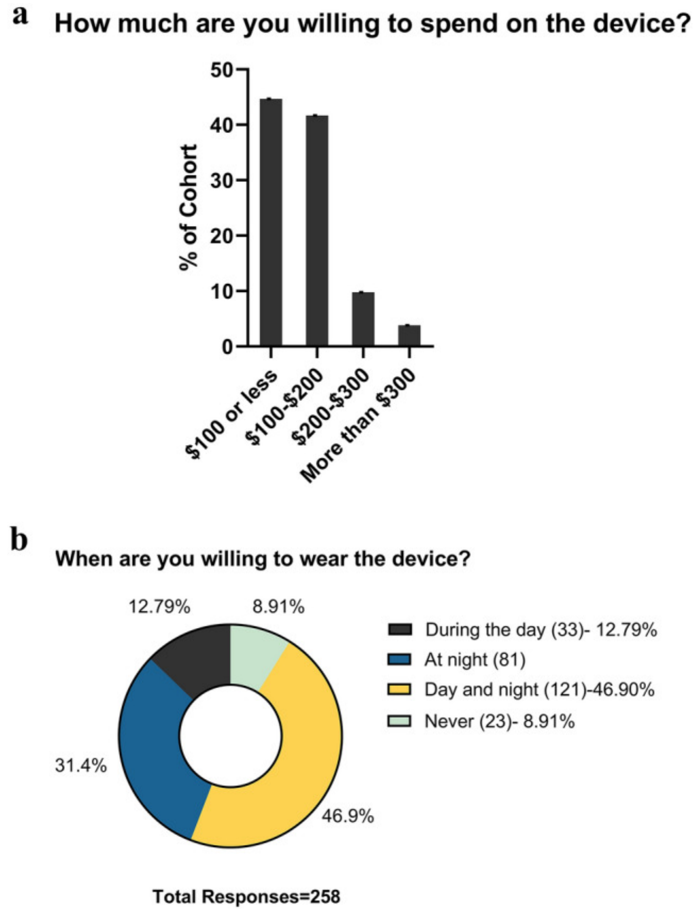


Fig. 5: The figure presents the spending patterns and usage willingness of wearable health monitors among pregnant women. The majority are willing to invest up to $200 for such devices. Most prefer wearing these devices continuously, both day and night, revealing a demand for 24/7 health monitoring [6].

The report's strengths lie in its clear and concise overview of the study design and results. It addresses an important and relevant research question with implications for user-centered design and development of wearable technologies for health monitoring during pregnancy. The report offers valuable insights into user needs, preferences, concerns, and expectations regarding wearable ECG devices.

However, the report lacks depth and detail in describing the study methodology and data analysis. Information about sample size, selection, characteristics, and representativeness is missing. The development, validation, and administration of the survey questions are not explained. The report lacks descriptive or inferential statistics to support the findings, and it does not discuss limitations or biases in the study design or data collection.

To improve this report, more information and justification for the study methodology and data analysis should be provided. Tables or graphs can be included to illustrate the survey results. The report should discuss limitations or challenges in conducting the study, such as sampling bias, response bias, or social desirability bias. Recommendations or implications for future research or practice based on the findings can also be presented.

C. Benefits of Real-Time Continuous Glucose Monitoring in Pregnancy

The third report reviews randomized controlled trials and cohort studies comparing continuous glucose monitoring (CGM) with self-monitoring of blood glucose (SMBG) in pregnant women with diabetes [7]. CGM continuously measures interstitial glucose levels and provides real-time feedback, while SMBG periodically measures blood glucose levels using finger-prick samples. The review demonstrates that CGM improves glycemic control, reduces hypoglycemia risk, and lowers adverse obstetric and neonatal outcomes compared to SMBG.

The report's strengths lie in its comprehensive and systematic review of the literature on CGM versus SMBG in pregnancy. The report follows a rigorous methodology for searching, selecting, appraising, synthesizing, and reporting the evidence. It demonstrates a high level of scientific rigor and validity in evaluating the effectiveness and safety of CGM in pregnancy. The clinical significance and relevance of CGM for improving maternal and fetal health outcomes in pregnant women with diabetes are highlighted.

However, the report lacks originality and novelty in presenting new findings or perspectives on CGM versus SMBG in pregnancy. It does not address gaps or controversies in the literature, nor does it provide critical analysis or discussion of the evidence. Contextual or practical factors that may affect the implementation or adoption of CGM in pregnancy, such as cost, availability, usability, or acceptability, are not considered.

To improve this report, it should provide an original contribution or insight to the literature on CGM versus SMBG in pregnancy. Identifying research questions or hypotheses that have not been adequately addressed or tested by previous studies would add value. Alternatively, offering alternative interpretations or explanations for conflicting or inconsistent findings in the literature could be explored. Additionally, discussing contextual or practical implications or challenges for using CGM in pregnancy, such as cost-effectiveness analysis, user satisfaction evaluation, or ethical considerations, would be beneficial.

In conclusion, this literature review evaluated three reports representing different approaches and technologies for real-time monitoring of biosignals during pregnancy. Each report exhibited strengths and weaknesses in terms of clarity, completeness, accuracy, validity, originality, and significance. Comparison of these reports can be seen in Table I. Future work should address common gaps or challenges across these reports, such as providing more empirical evidence and evaluation, assessing user acceptance and satisfaction, discussing ethical and legal issues, and comparing approaches with existing or alternative methods or devices.

III. Machine Learning for Real-Time Labor Monitoring

A. Our Approach to Preterm Labor Prediction

Our approach is centered around the concept that electromyography (EMG) signals from the abdomen of pregnant women hold valuable insights into uterine activity and potential risks of preterm labor. Condition marked by childbirth prior to the completion of 37 weeks of gestation, which may pose significant health threats to the baby. To mitigate this, our goal is to engineer a device capable of real-time monitoring of a patient's EMG signals, analyzing them through machine learning, and processing the results to communicate whether an emergency situation arises.

To realize this, several steps need to be undertaken. The first is data collection, which involves the acquisition of EMG signals from pregnant women with the help of abdominal sensors, and gathering information concerning their pregnancy outcomes such as delivery times (preterm or term) and any associated complications. This data can be procured either from an established database like the "Term-Preterm EHG Database" or gathered from actual patients. In our case we used "Term-Preterm EHG Database" to conduct a machine learning algorithm [8].

Subsequently, data preprocessing becomes necessary to clean and prepare the EMG signals for scrutiny. This step encompasses the elimination of any noise or artifacts that could interfere with the signal quality, and segmentation and labeling of signals in accordance with their respective pregnancy outcomes. For instance, signals can be tagged as preterm or term based on gestational age at delivery. Once preprocessing is complete, the focus shifts to data analysis. Here, machine learning techniques are employed Logistic Regression, K-Means, Agglomerative Clustering, MeanShift Clustering and SVC Classifier to scrutinize the EMG signals and extract distinguishing features between preterm and term pregnancies. Furthermore, the model is trained, tested, and evaluated using the most fitting methods and metrics. Lastly, in the data processing phase, the trained machine learning model is utilized to analyze fresh EMG signals from pregnant women and predict their pregnancy outcomes. All these steps shown in the Fig. 6. A user-friendly interface is designed to exhibit these predictions and deliver feedback or alerts to the patient or doctor if a risk of preterm labor or other complications is identified.

Overall, this approach not only aims to present an innovative and practical solution for decreasing infant mortalities linked to late diagnosis of preterm labor or other complications but also strives to push the boundaries of wearable technologies for health monitoring during pregnancy.

Approach	Goal	Method	Technology	Outcome	Limitation
Telemonitoring of pregnant women at home	To provide continuous and remote monitoring of fetal heart rate and uterine activity signals	A body area network (BAN) of sensors connected to a personal area network (PAN) that processes and transmits the signals to a central surveillance center	Wireless ultrasound and tocodynamometer devices; embedded algorithms for signal analysis; reliable transmission channel	Accurate and consistent measurement of biosignals; smart alerts and human-machine interface; communication with medical staff	Technical difficulties in signal acquisition and transmission; user acceptance and compliance; ethical and legal issues
Wearable technology for health monitoring during pregnancy	To evaluate the perception and acceptance of remote fetal ECG monitoring technologies among women of child-bearing age	A survey study that asked women about their willingness, preferences, concerns, and expectations regarding wearable ECG devices	A wearable ECG device that can measure fetal heart rate and maternal heart rate continuously and remotely	High willingness to use a wearable ECG device throughout pregnancy; preference for comfort, convenience, and feedback; concern about privacy, reliability, and cost	A hypothetical scenario without actual testing of the device; limited sample size and diversity; lack of clinical outcomes
Benefits of real-time continuous glucose monitoring in pregnancy	To assess the impact of continuous glucose monitoring (CGM) on maternal glycemia and obstetric and neonatal outcomes in pregnant women with diabetes	A review of randomized-controlled trials and cohort studies that compared CGM with self-monitoring of blood glucose (SMBG)	A CGM device that measures interstitial glucose levels continuously and provides real-time feedback	Improved glycemic control; reduced risk of hypoglycemia; lower rates of preterm delivery, cesarean section, macrosomia, neonatal intensive care unit admission, and neonatal hypoglycemia	High cost and limited availability of CGM devices; skin irritation and sensor failure; need for calibration with SMBG

TABLE I: Comparative analysis of three different approaches used for monitoring and assessing health metrics during pregnancy. Each of these techniques is evaluated based on their goals, methods, technology used, outcomes, and limitations.

Fig. 6: A representation of the EMG Signal Analysis and Preterm Labor Prediction process.

B. Results and Demonstration

The TPEHG DB is a database that contains records of Electrohysterograms, which are recordings of uterine electromyography (EMG) signals. These recordings were obtained at the Department of Obstetrics and Gynecology, University Medical Centre Ljubljana, between 1997 and 2005. The data collection took place during regular check-ups, specifically around the 22nd or 32nd week of gestation. In order to improve the quality of the data, a 4-pole band-pass Butterworth filter was applied, which restricted the signal frequencies to a range between 0.3Hz and 3Hz. The TPEHG DB was specifically chosen as the dataset for analysis. The code begins by importing libraries and loading the dataset. The data is then edited and preprocessed. It is divided into training and test sets. The training data is utilized to train a logistic regression model, which is then employed to make predictions on the test set. Model performance is evaluated using various metrics. Additionally, the code applies different clustering methods, namely K-Means, Agglomerative Clustering, and Mean Shift, to the data. An SVM classifier is also trained on selected features. The results are plotted and analyzed. These models can be seen in Fig. 7, 8, 9 and 10 respectively. Real-time demonstration of EHG analysis for predicting preterm labor can be observed after references.

In summary, this code performs classification using logistic regression, clustering using K-Means, Agglomerative Clustering, and Mean Shift, and trains an SVM classifier. It includes data preprocessing, splitting into training and test sets, and

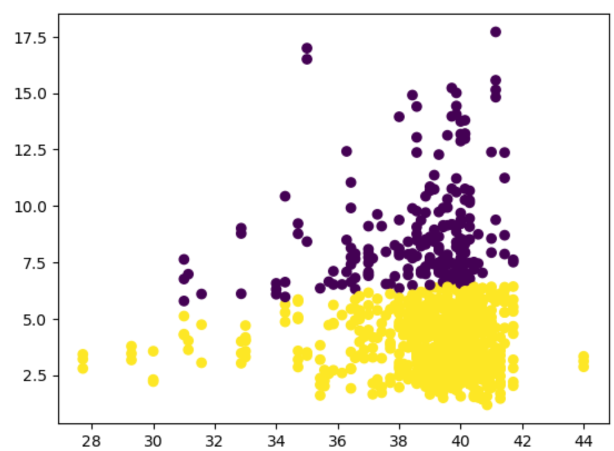


Fig. 7: The k-means model is a popular clustering algorithm used to group data points into distinct clusters based on their similarity. Firstly, the data is classified by the K-Means model. The F-score value of 0.71875 indicates that the k-means model's performance is reasonably good.

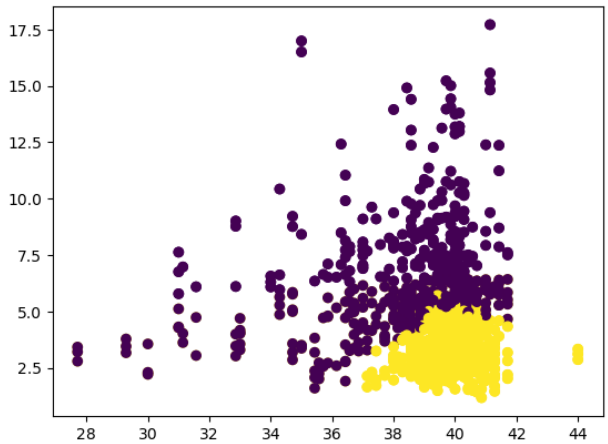


Fig. 8: The data has been clustered using Agglomerative Clustering, but the results are not reliable or consistent. Different runs of the algorithm or variations in parameters may lead to varying cluster assignments, indicating instability in the clustering outcomes.

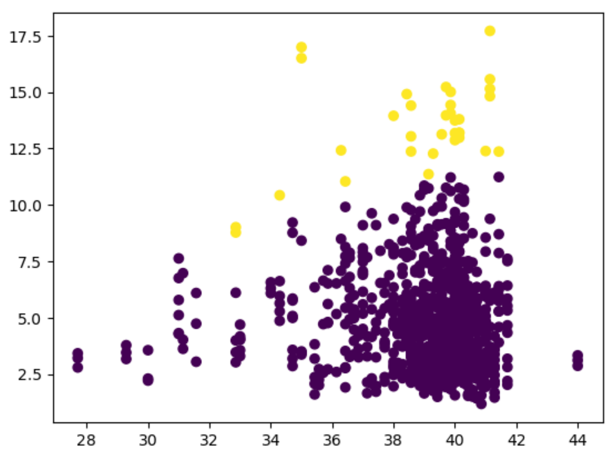


Fig. 9: When the data is classified according to the MeanShift Clustering algorithm, the estimates or cluster assignments obtained are not considered fairly consistent. The outcomes of MeanShift Clustering are unstable and unreliable, resulting in inconsistent results.

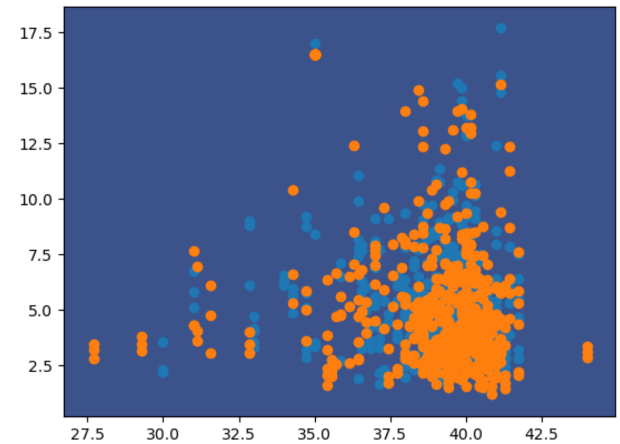


Fig. 10: The estimates or predictions obtained from the SVC (Support Vector Classifier) algorithm exhibit a high level of consistency. This model has a high F-score. Therefore, it means that the SVC classifier is performing well in terms of classification accuracy and the ability to correctly classify positive instances.

performance evaluation. The code provides a comprehensive analysis of the TPEHG DB dataset, extracting valuable insights from the data.

IV. Conclusion

In conclusion, preterm labor, which occurs when a baby is born before the completion of 37 weeks of gestation, remains a significant global health issue. The conventional methods for predicting preterm labor have several limitations; they are often not continuous, real-time, or easily accessible. Moreover, they may have low sensitivity, require invasive procedures, or fail to reliably confirm or rule out the onset of preterm labor. Therefore, developing an innovative approach for monitoring the biosignals of mother and child during pregnancy is critical. Such an approach needs to be continuous, real-time, accurate, and affordable.

Recent advancements in technology have introduced the prospect of employing wearable devices for this purpose. Particularly, analyzing the biosignals such as electromyography (EMG) and uterine electrical signals electrohysterography (EHG) through wearable devices coupled with machine learning algorithms offers a promising avenue. These signals hold valuable insights into uterine activity, which is essential in identifying genuine labor. By using machine learning models to analyze these signals in real-time, healthcare providers can potentially predict preterm labor with higher accuracy and in a timelier manner.

The analysis of various literature reveals that telemonitoring using medical cyber-physical systems, wearable technology, and continuous glucose monitoring are emerging approaches in real-time biosignal monitoring. These approaches have their own strengths, but also suffer from various limitations such as lack of empirical evidence, user acceptance assessment, and ethical considerations. Addressing these limitations through comprehensive clinical trials and user feedback can significantly enhance the efficacy and acceptability of these technologies.

In the proposed approach, the focus is on creating a device that utilizes machine learning algorithms to analyze EMG signals from the abdomen of pregnant women. This approach is centered on the idea that real-time monitoring of these signals can provide critical information on uterine activity, thus enabling the early detection of preterm labor. The device would be designed to offer user-friendly feedback to patients or healthcare providers, potentially facilitating timely interventions to improve neonatal outcomes.

In essence, embracing the integration of wearable technology with machine learning represents an enormous leap forward in obstetric care. Not only does this innovative approach have the potential to significantly reduce the neonatal mortality and morbidity associated with preterm labor through early detection and intervention, but it also exemplifies the transformative impact of technology on healthcare. Future research and development should focus on optimizing the accuracy, accessibility, and affordability of these wearable devices, as well as ensuring ethical considerations and user acceptability are adequately addressed.

REFERENCES

[1] J. D. Iams, "Prediction and early detection of preterm labor," *Obstetrics and Gynecology*, vol. 101, no. 2, pp. 402–412, 2003.

[2] M. Lucovnik, R. J. Kuon, L. R. Chambliss, W. L. Maner, S.-Q. Shi, L. Shi, J. Balducci, and R. E. Garfield, "Use of uterine electromyography to diagnose term and preterm labor," *Acta Obstetricia et Gynecologica Scandinavica*, vol. 90, no. 2, pp. 150–157, 2011.

[3] P. Fergus, P. Cheung, A. Hussain, D. Al-Jumeily, C. Dobbins, and S. Iram, "Prediction of preterm deliveries from ehg signals using machine learning," *PLOS ONE*, vol. 8, no. 10, p. e77154, 2013.

[4] M. Vinken, C. Rabotti, M. Mischi, J. van Laar, and G. Oei, "Nifedipine-induced changes in the electrohysterogram of preterm contractions: Feasibility in clinical practice," *Obstetrics and gynecology international*, vol. 2010, p. 325635, 01 2010.

[5] A. Pawlak, K. Horoba, J. Jezewski, J. Wrobel, and A. Matonia, "Telemonitoring of pregnant women at home — biosignals acquisition and measurement," in *2015 22nd International Conference Mixed Design of Integrated Circuits Systems (MIXDES)*, pp. 83–87, 2015.

[6] C. Wakefield, L. Yao, S. Self, and M. G. Frasch, "Wearable technology for health monitoring during pregnancy: an observational cross-sectional survey study," *Archives of Gynecology and Obstetrics*, vol. 308, pp. 73–78, July 2023. Epub 2022 Jul 14.

[7] J. M. Yamamoto and H. R. Murphy, "Benefits of real-time continuous glucose monitoring in pregnancy," *Diabetes Technology & Therapeutics*, vol. 23, pp. S8–S14, Mar. 2021.

[8] G. Fele-Zorz, G. Kavsek, Z. Novak-Antolic, and F. Jager, "A comparison of various linear and non-linear signal processing techniques to separate uterine emg records of term and pre-term delivery groups," *Med Biol Eng Comput*, vol. 46, no. 9, pp. 911–922, 2008. Epub 2008 Apr 24.

Real-Time Demonstration of Electrohysterography (EHG) Analysis for Predicting Preterm Labor

June 11, 2023

[1]:
```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns
import csv
import os
import matplotlib.pyplot as plt
```

[2]:
```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import confusion_matrix
from sklearn import preprocessing
```

[3]:
```
tpehgdb_filter_03_3Hz = open("/Users/efdalozel/Desktop/23' Bahar Dönemi/Medic/
↪term-preterm-ehg-database-1.0.1/tpehgdb_features__filter_0.3_Hz-3.0_Hz.fvl",␣
↪"r")
print(tpehgdb_filter_03_3Hz.read())
```

```
  Record    | Chann | Gestation | Rec. time |   Group    |   RMS    |  Fmed  |
Fpeak  | Samp. en. | Premature | Early
-----------+-------+-----------+-----------+-----------+---------+--------+-----
---+-----------+-----------+-------
 tpehg1007 |     1 |     35.00 |     31.29 | >=26-PRE   | 16.5069 | 0.3688 |
0.3324 |    0.5180 | t         | f
 tpehg1007 |     2 |     35.00 |     31.29 | >=26-PRE   | 16.9946 | 0.3688 |
0.3159 |    0.3810 | t         | f
 tpehg1027 |     2 |     37.14 |     31.29 | >=26-TERM |  1.6525 | 0.5284 |
0.4090 |    0.8160 | f         | f
 tpehg1027 |     3 |     37.14 |     31.29 | >=26-TERM |  2.1390 | 1.4289 |
1.4602 |    0.8250 | f         | f
 tpehg1029 |     1 |     38.57 |     31.00 | >=26-TERM |  5.5345 | 0.4338 |
0.3340 |    0.8430 | f         | f
 tpehg1029 |     2 |     38.57 |     31.00 | >=26-TERM |  5.7970 | 0.3903 |
0.3269 |    0.5170 | f         | f
 tpehg1029 |     3 |     38.57 |     31.00 | >=26-TERM |  3.5197 | 0.5121 |
0.3234 |    0.8860 | f         | f
 tpehg1031 |     1 |     38.86 |     22.14 | <26-TERM   |  5.2597 | 0.5139 |
```

```
0.3402 |     0.8060 | f           | t
 tpehg1031 |       2 |      38.86 |      22.14 | <26-TERM   |  2.8064 | 0.3760 |
0.3674 |     0.6190 | f           | t

   ...      |    ... |      ...   |      ...   |     ...    |    ...

 tpehg1039 |       1 |      40.29 |      22.86 | <26-TERM   |  5.0666 | 0.4620 |
0.3547 |     0.7520 | f           | t
 tpehg1039 |       2 |      40.29 |      22.86 | <26-TERM   |  7.4758 | 0.3760 |
0.3547 |     0.5620 | f           | t
 tpehg1039 |       3 |      40.29 |      22.86 | <26-TERM   |  3.1620 | 0.4335 |
```

[4]:
```python
file_path = "/Users/efdalozel/Desktop/23' Bahar Dönemi/Medic/
↪term-preterm-ehg-database-1.0.1/tpehgdb_features__filter_0.3_Hz-3.0_Hz.fvl"

df = pd.read_csv(file_path, delimiter='|')
```

[5]:
```python
df
```

[5]:
```
                                               Record      Chann    Gestation   \
0    -----------+-------+-----------+-----------+--...        NaN          NaN
1                                           tpehg1007        1.0        35.00
2                                           tpehg1007        2.0        35.00
3                                           tpehg1007        3.0        35.00
4                                           tpehg1021        1.0        38.57
..                                                ...        ...          ...
896                                         tpehg991         2.0        39.86
897                                         tpehg991         3.0        39.86
898                                         tpehg994         1.0        37.14
899                                         tpehg994         2.0        37.14
900                                         tpehg994         3.0        37.14

      Rec. time      Group         RMS        Fmed       Fpeak      Samp. en.   \
0           NaN          NaN         NaN         NaN         NaN          NaN
1         31.29    >=26-PRE      16.5069     0.3688      0.3324        0.518
2         31.29    >=26-PRE      16.9946     0.3688      0.3159        0.381
3         31.29    >=26-PRE       8.4198     0.3981      0.2938        0.564
4         22.29    <26-TERM      14.4062     0.4170      0.3828        0.446
..          ...          ...         ...         ...         ...          ...
896       23.29    <26-TERM       3.4104     0.5391      0.3203        0.890
897       23.29    <26-TERM       2.8062     0.5547      0.3288        0.904
898       31.00    >=26-TERM      3.7121     0.6887      0.3436        0.858
```

```
899         31.00   >=26-TERM     7.6799   0.9405   0.6377         0.707
900         31.00   >=26-TERM     4.1809   0.5588   0.3156         0.856

      Premature   Early
0           NaN     NaN
1     t               f
2     t               f
3     t               f
4     f               t
..          ...     ...
896   f               t
897   f               t
898   f               f
899   f               f
900   f               f

[901 rows x 11 columns]
```

[6]:
```
df.columns =['record',␣
↪'chann','gestation','rec_time','group','rms','fmed','fpeak','samp_en','premature','early']
```

[7]:
```
df = df.loc[1:]
```

[8]:
```
df
```

[8]:
```
          record  chann  gestation  rec_time       group      rms    fmed  \
1     tpehg1007    1.0      35.00     31.29   >=26-PRE   16.5069  0.3688
2     tpehg1007    2.0      35.00     31.29   >=26-PRE   16.9946  0.3688
3     tpehg1007    3.0      35.00     31.29   >=26-PRE    8.4198  0.3981
4     tpehg1021    1.0      38.57     22.29   <26-TERM   14.4062  0.4170
5     tpehg1021    2.0      38.57     22.29   <26-TERM   12.3715  0.4120
..          ...    ...        ...       ...         ...      ...     ...
896   tpehg991     2.0      39.86     23.29   <26-TERM    3.4104  0.5391
897   tpehg991     3.0      39.86     23.29   <26-TERM    2.8062  0.5547
898   tpehg994     1.0      37.14     31.00   >=26-TERM   3.7121  0.6887
899   tpehg994     2.0      37.14     31.00   >=26-TERM   7.6799  0.9405
900   tpehg994     3.0      37.14     31.00   >=26-TERM   4.1809  0.5588

      fpeak  samp_en   premature early
1    0.3324    0.518   t             f
2    0.3159    0.381   t             f
3    0.2938    0.564   t             f
4    0.3828    0.446   f             t
5    0.3136    0.437   f             t
..      ...      ...         ...   ...
896  0.3203    0.890   f             t
897  0.3288    0.904   f             t
```

```
898  0.3436    0.858   f              f
899  0.6377    0.707   f              f
900  0.3156    0.856   f              f

[900 rows x 11 columns]
```

1 “Group”feature converted to numeric values with label encoder.

[9]:
```python
feature_cols =␣
 ↪['chann','gestation','rec_time','group','rms','fmed','fpeak','samp_en']
from sklearn.preprocessing import LabelEncoder
label_encoder = LabelEncoder()
df.group = label_encoder.fit_transform(df.group)
```

```
/var/folders/hm/x5993gs50ml0mh3_bdqwcnn40000gn/T/ipykernel_45805/3339577967.py:4
: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df.group = label_encoder.fit_transform(df.group)
```

[10]:
```python
x = df[feature_cols]
y = np.array(df['early'] == " t")
```

[11]:
```python
x
```

[11]:

	chann	gestation	rec_time	group	rms	fmed	fpeak	samp_en
1	1.0	35.00	31.29	2	16.5069	0.3688	0.3324	0.518
2	2.0	35.00	31.29	2	16.9946	0.3688	0.3159	0.381
3	3.0	35.00	31.29	2	8.4198	0.3981	0.2938	0.564
4	1.0	38.57	22.29	1	14.4062	0.4170	0.3828	0.446
5	2.0	38.57	22.29	1	12.3715	0.4120	0.3136	0.437
..	...	...	...	...	...	...	...	...
896	2.0	39.86	23.29	1	3.4104	0.5391	0.3203	0.890
897	3.0	39.86	23.29	1	2.8062	0.5547	0.3288	0.904
898	1.0	37.14	31.00	3	3.7121	0.6887	0.3436	0.858
899	2.0	37.14	31.00	3	7.6799	0.9405	0.6377	0.707
900	3.0	37.14	31.00	3	4.1809	0.5588	0.3156	0.856

[900 rows x 8 columns]

[12]:
```python
y
```

[12]:
```
array([False, False, False,  True,  True,  True, False, False, False,
       False, False, False, False, False, False,  True,  True,  True,
```

```
       True,  True,  True, False, False, False,  True,  True,  True,
       True,  True,  True,  True,  True,  True,  True,  True,  True,
       True,  True,  True, False, False, False,  True,  True,  True,
       True,  True,  True, False, False, False,  True,  True,  True,
       True,  True,  True, False, False, False,  True,  True,  True,
       True,  True,  True,  True,  True,  True,  True,  True,  True,...])
```

2 The data set was separated as train and test.

[15]:
```
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=50,␣
 ↪random_state=2)
```

[16]:
```
x_train
```

[16]:
```
     chann  gestation  rec_time  group     rms    fmed   fpeak  samp_en
782    2.0      40.00     22.57      1  7.6811  0.4584  0.3416    0.583
523    1.0      40.71     23.86      1  3.9442  0.4139  0.2819    0.790
682    1.0      38.71     22.43      1  2.2834  1.3638  0.3850    0.832
676    1.0      39.43     30.86      3  6.6296  0.4684  0.3414    0.765
310    1.0      38.86     23.00      1  7.2785  0.5345  0.4214    0.722
..     ...        ...       ...    ...     ...     ...     ...      ...
535    1.0      40.57     22.86      1  4.4103  0.5417  0.4619    0.851
585    3.0      41.00     33.86      3  2.7621  0.5111  0.5771    0.712
494    2.0      38.57     23.14      1  3.7055  0.5084  0.3049    0.932
528    3.0      39.43     24.57      1  4.9464  0.4408  0.3889    0.856
169    1.0      39.71     24.00      1  3.3484  1.4360  1.5497    0.667

[850 rows x 8 columns]
```

[17]:
```
x_test
```

[17]:
```
     chann  gestation  rec_time  group      rms    fmed   fpeak  samp_en
564    3.0      41.00     29.29      3   6.4139  0.6973  0.5028    0.759
553    1.0      40.00     30.71      3   6.5874  0.5889  0.3493    0.648
440    2.0      40.00     22.14      1   7.5132  0.3528  0.3146    0.538
170    2.0      39.71     24.00      1   2.3880  0.5583  0.4167    0.719
153    3.0      39.71     30.29      3   2.9914  0.6721  0.3393    0.852
66     3.0      40.57     22.71      1   2.8820  0.9901  1.3605    0.918
100    1.0      38.57     31.57      3   3.0180  0.5000  0.3314    0.740
276    3.0      39.29     22.57      1   2.8288  0.7153  0.3174    0.890
845    2.0      40.00     31.86      3   9.2159  0.4442  0.3795    0.396
249    3.0      37.29     22.14      1   5.8834  0.3549  0.3319    0.725
331    1.0      39.86     32.14      3  10.6134  0.3595  0.3267    0.552
637    1.0      40.14     22.71      1   4.7310  0.4372  0.3424    0.567
847    1.0      34.29     30.71      2   5.6398  0.4620  0.3305    0.588
341    2.0      40.00     22.29      1   7.5695  0.3597  0.3444    0.523
505    1.0      40.00     23.29      1   6.6357  0.4569  0.3313    0.695
```

16	1.0	38.86	22.14	1	5.2597	0.5139	0.3402	0.806
21	3.0	40.29	22.86	1	3.1620	0.4335	0.3547	0.728
173	2.0	38.00	24.43	1	13.9522	0.4209	0.3794	0.423
790	1.0	40.71	23.14	1	7.0341	0.4872	0.3283	0.656
380	2.0	40.43	22.71	1	1.5344	0.7931	0.3118	0.942
503	2.0	38.00	23.57	1	5.2741	0.3681	0.3146	0.693
198	3.0	40.57	23.57	1	4.8115	0.4729	0.2542	0.650
165	3.0	40.29	23.14	1	5.5983	0.4512	0.4505	0.756
232	1.0	40.43	30.57	3	5.4494	0.5004	0.3099	0.787
841	1.0	41.43	23.29	1	8.7018	0.4171	0.3386	0.902
722	2.0	40.14	23.43	1	13.1979	0.3888	0.3475	0.517
521	2.0	39.71	22.86	1	3.1613	0.4674	0.3681	0.926
143	2.0	41.71	30.71	3	5.5215	0.3707	0.3243	0.482
780	3.0	35.43	22.29	0	2.0713	0.5761	0.3698	0.896
884	2.0	40.29	23.14	1	2.1029	0.5687	0.3791	0.683
102	3.0	38.57	31.57	3	4.2095	0.5164	0.3021	0.893
578	2.0	39.00	35.57	3	2.9295	0.4542	0.4139	0.744
38	2.0	39.57	24.43	1	2.5301	0.5014	0.3471	0.692
725	2.0	38.43	22.29	1	3.9474	0.4790	0.3468	0.680
336	3.0	36.71	22.71	0	5.3726	0.3896	0.3833	0.682
186	3.0	40.00	23.14	1	5.2806	0.3255	0.3048	0.747
729	3.0	38.86	22.43	1	4.8760	0.4275	0.3144	0.872
668	2.0	38.71	31.86	3	2.9359	0.4601	0.3495	0.577
848	2.0	34.29	30.71	2	5.9670	0.4222	0.2999	0.403
8	2.0	38.57	31.00	3	9.0688	0.4479	0.3730	0.447
147	3.0	41.71	22.43	1	4.3376	0.5892	0.3402	0.678
899	2.0	37.14	31.00	3	7.6799	0.9405	0.6377	0.707
144	3.0	41.71	30.71	3	7.5103	0.3614	0.3293	0.527
287	2.0	40.29	30.29	3	9.0799	0.4319	0.4514	0.448
364	1.0	40.00	31.14	3	3.2468	0.4986	0.2972	0.877
890	2.0	39.57	30.57	3	10.3020	0.4747	0.4057	0.472
397	1.0	40.43	22.57	1	6.4144	0.7021	0.3882	0.811
750	3.0	41.14	23.00	1	2.6623	0.7845	0.4680	0.852
323	2.0	39.14	22.29	1	3.7187	0.4132	0.4583	0.770
574	1.0	40.00	31.00	3	5.8347	0.5084	0.4709	0.502

[18]:
```
y_test
```

[18]:
```
array([False, False,  True,  True, False,  True, False,  True, False,
        True, False,  True, False,  True,  True,  True,  True,  True,
        True,  True,  True,  True,  True, False,  True,  True,  True,
       False,  True,  True, False, False,  True,  True,  True,  True,
        True, False, False, False,  True, False, False, False, False,
       False,  True,  True,  True, False])
```

3 Logistic Regression model was applied.

[19]:
```
logreg = LogisticRegression(random_state=2,max_iter=100)
```

[20]:
```
logreg.fit(x_train,y_train)
```

[20]:
```
LogisticRegression(random_state=2)
```

4 Making predictions based on the trained model.

[23]:
```
y_pred = logreg.predict(x_test)
```

5 Confusion Matrix (cnf_matrix)

[24]:
```
from sklearn import metrics
cnf_matrix = metrics.confusion_matrix(y_test,y_pred)
cnf_matrix
```

[24]:
```
array([[20,  0],
       [ 0, 30]])
```

6 #f-score

[25]:
```
from sklearn.metrics import f1_score
f1_score(y_test,y_pred)
```

[25]:
```
1.0
```

7 KMeans Model

[29]:
```
colss = ['gestation','rms','early']
k = df[colss]
k.early = label_encoder.fit_transform(k.early)
```

```
/var/folders/hm/x5993gs50ml0mh3_bdqwcnn40000gn/T/ipykernel_45805/1445324759.py:3
: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  k.early = label_encoder.fit_transform(k.early)
```

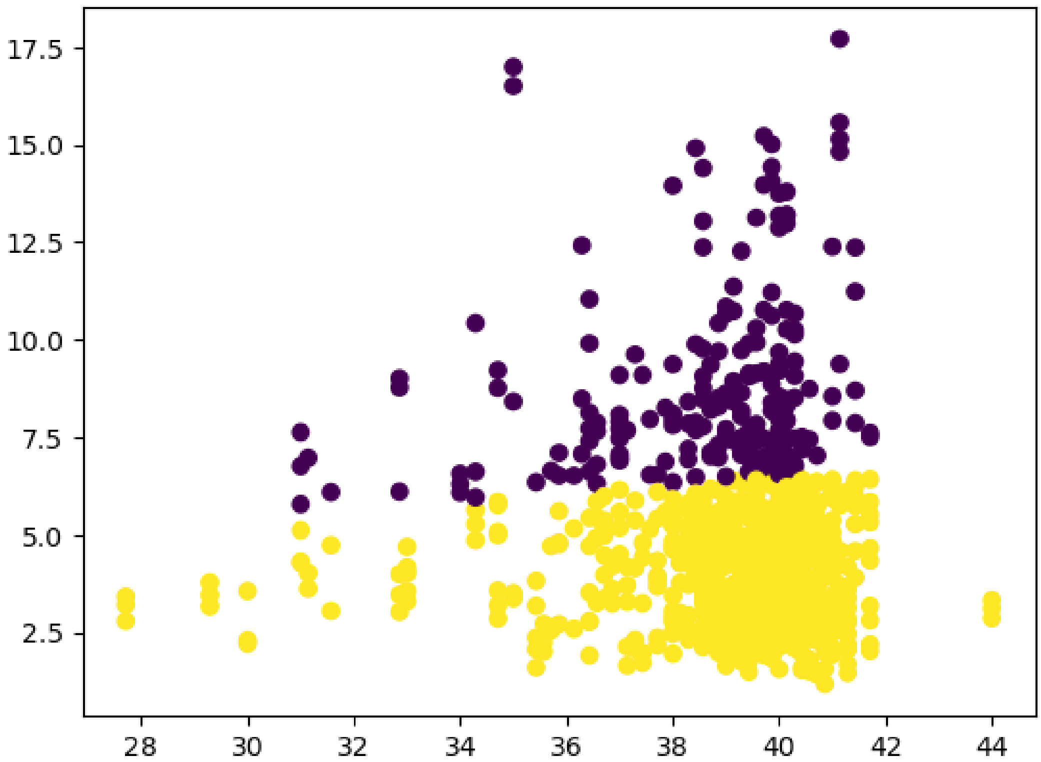
[30]:
```
x_train_km, x_test_km, y_train_km, y_test_km = train_test_split(k, k.early,
 ↪test_size=50, random_state=2)
```

[31]:
```
ks = k
from sklearn.cluster import KMeans
kmeans = KMeans(2)
kmeans.fit(ks)
```

[31]: KMeans(n_clusters=2)

[32]:
```
identified_clusters = kmeans.fit_predict(ks)
identified_clusters
data_with_clusters = k.copy()
data_with_clusters['Clusters'] = identified_clusters
plt.
 ↪scatter(data_with_clusters['gestation'],data_with_clusters['rms'],c=data_with_clusters['Clust
```

[32]: <matplotlib.collections.PathCollection at 0x7fd030224940>



8 KMeans predictions based on the trained model.

[34]:
```
y_pred_km = kmeans.predict(x_test_km)
cnf_matrix_kmeans = metrics.confusion_matrix(y_test_km,y_pred_km)
cnf_matrix_kmeans
```

```
[34]: array([[ 9, 11],
              [ 7, 23]])
```

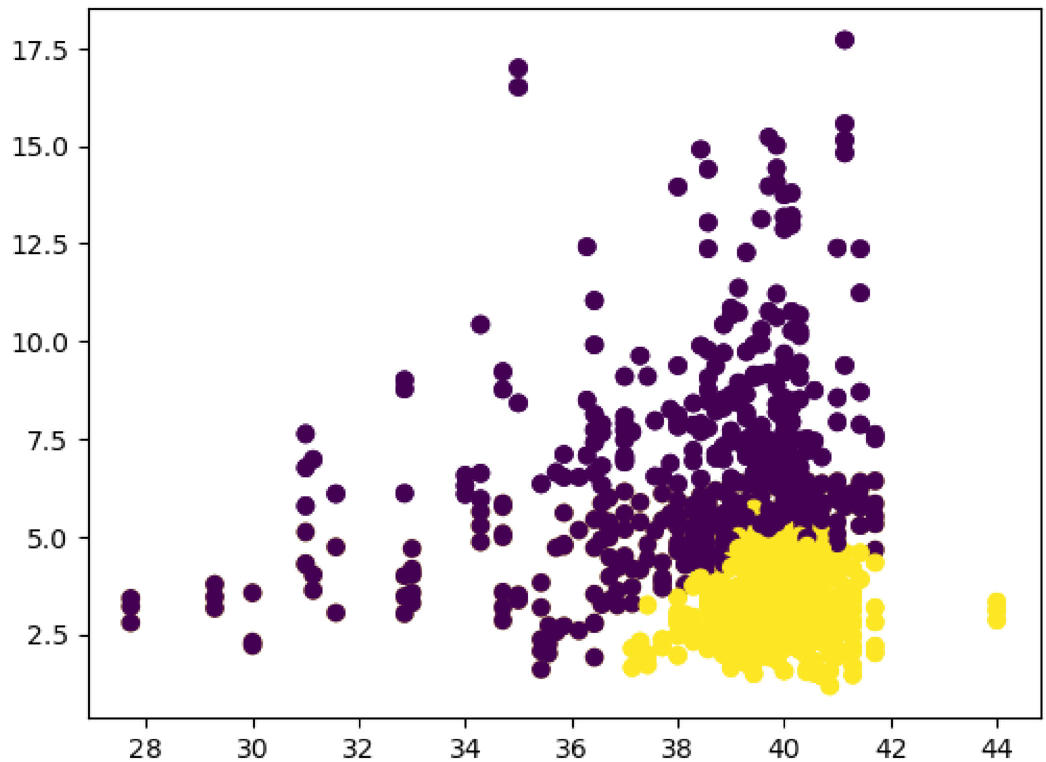
9 f-score

```
[36]: f1_score(y_test_km,y_pred_km)
```

```
[36]: 0.71875
```

10 Agglomerative Clustering

```
[37]: import numpy as np
import matplotlib.pyplot as plt
from sklearn.cluster import AgglomerativeClustering

hierarchical_cluster = AgglomerativeClustering(n_clusters=2,␣
 ↪affinity='euclidean', linkage='ward')
labels = hierarchical_cluster.fit_predict(k)
plt.
 ↪scatter(data_with_clusters['gestation'],data_with_clusters['rms'],c=data_with_clusters['Clust

plt.scatter(data_with_clusters['gestation'], data_with_clusters['rms'], c=labels)
plt.show()
```

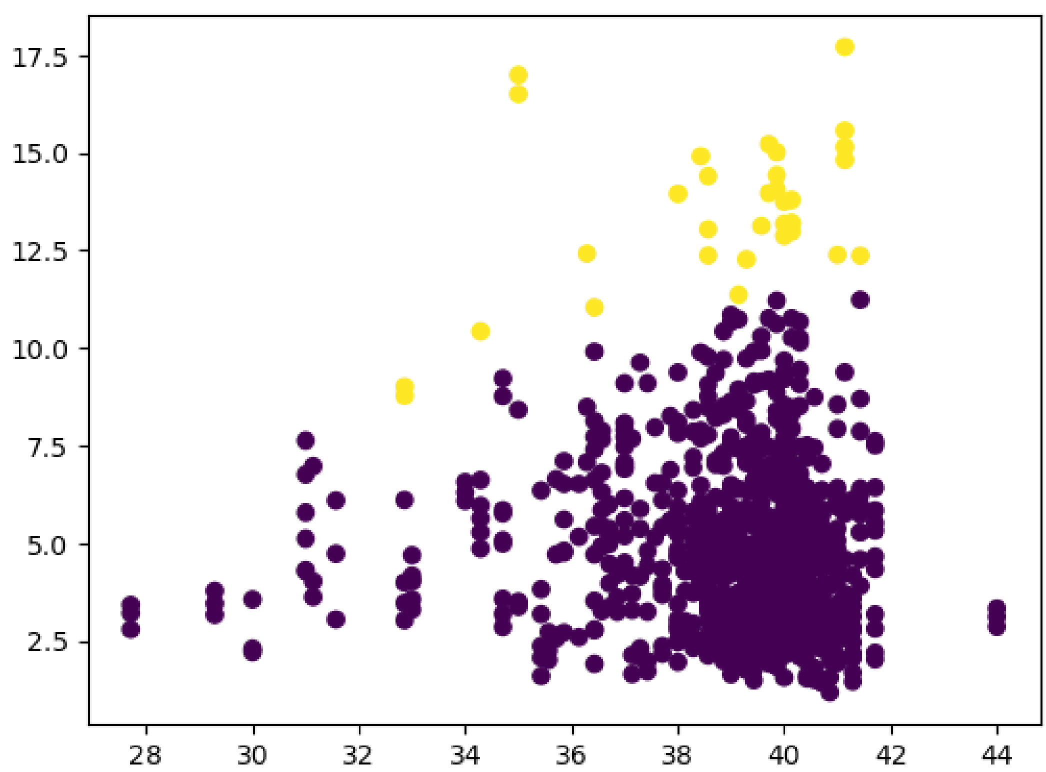


11 MeanShift Clustering

[90]:
```
from sklearn.cluster import MeanShift, estimate_bandwidth
bandwidth = estimate_bandwidth(k, quantile=0.34, n_samples=640)

ms = MeanShift(bandwidth=bandwidth, bin_seeding=True)
ms.fit(k)
mslabels = ms.labels_
cluster_centers = ms.cluster_centers_

labels_unique = np.unique(mslabels)
n_clusters_ = len(labels_unique)

print("number of estimated clusters : %d" % n_clusters_)
```

```
number of estimated clusters : 2
```

[91]:
```
plt.scatter(data_with_clusters['gestation'], data_with_clusters['rms'], c =␣
↪mslabels) # plotting the clusters
plt.show() # showing the plot
```



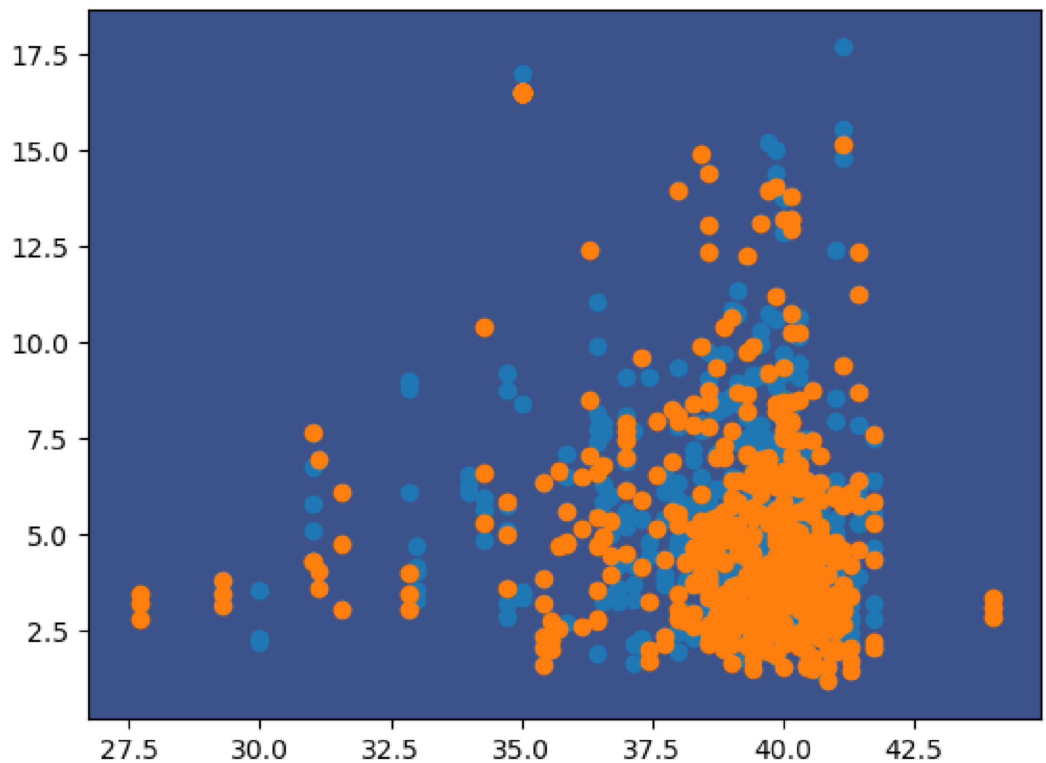
12 SVC Classifier

[97]:
```
from sklearn.svm import SVC
from sklearn.neural_network import MLPClassifier

a=np.array(k)
X=(a[:, :2])
Y=(a[:, 2:3])

model2 = SVC(kernel='rbf',random_state=0)
model2.fit(X,Y)

min1,max1 = X[:,0].min()-1,X[:,0].max()+1
min2,max2 = X[:,1].min()-1,X[:,1].max()+1

x1grid = np.arange(min1,max1,0.1)
x2grid = np.arange(min2,max2,0.1)
```

```
/Users/efdalozel/opt/anaconda3/lib/python3.9/site-
packages/sklearn/utils/validation.py:993: DataConversionWarning: A column-vector
y was passed when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  y = column_or_1d(y, warn=True)
```

[98]:
```
xx,yy=np.meshgrid(x1grid,x2grid)
r1,r2= xx.flatten(),yy.flatten()
r1,r2 = r1.reshape((len(r1),1)),r2.reshape((len(r2),1))
grid = np.hstack((r1,r2))
yhat = model2.predict(grid)
zz=yhat.reshape(xx.shape)
```

[100]:
```
plt.contourf(xx,yy,zz)
for class_value in range(2):
    row_ix = np.where(Y == class_value)
    plt.scatter(X[row_ix,0],X[row_ix,1],cmap='Paired')
    #orange true, blue false
```



[101]:
```
x_train_svc, x_test_svc, y_train_svc, y_test_svc = train_test_split(k, k.early,␣
 ↪test_size=50, random_state=2)
b=np.array(x_test_svc)
X=(b[:, :2])
y_pred_svc = model2.predict(X)
f1_score(y_test_svc,y_pred_svc)
```

[101]: 0.7499999999999999