Data Article

Comprehensive miscarriage dataset for an early miscarriage prediction

Hiba Asri [a],[*],[1], Hajar Mousannif [b], Hassan Al Moatassime [a]

[a] *OSER Laboratory, Cadi Ayyad University, Marrakech, Morocco*
[b] *LISI Laboratory, Cadi Ayyad University, Marrakech, Morocco*

ARTICLE INFO

ABSTRACT

We present risk factors for predicting miscarriage. Our data is created through an android mobile application that collects automatically real-time data about the pregnant woman. This process is done every 60 s while the mobile application is on active mode. We distinguish two types of data: data from mobile phone and data from healthcare sensors. Data generated is real and concerns real pregnant women to test and validate the proposed system and assess its performance and effectiveness.

Specifications Table

Subject area	*Healthcare & Computer Science*
More specific subject area	*Predictive and preventive medicine*
Type of data	*Text file.*
How data was acquired	*Data is acquired from mobile phone and healthcare sensors.*
Data format	*Analyzed*
Experimental factors	

---

* Corresponding author.
   E-mail addresses: asri.hiba@gmail.com (H. Asri), mousannif@uca.ac.ma (H. Mousannif), hassan.al.moatassime@gmail.com (H. Al Moatassime).
   [1] Laboratory of Optimization System Evolution & Network (OSER); Faculty of Sciences & Techniques, Cadi Ayyad University, Marrakech, Blvd Prince My abdellah, B.P. 2390, 40000 Marrakech, Morocco.

Experimental features	*Age, Heart Rate Variability, BMI, History of Previous Miscarriage, Activity, Location, Body Temperature, Body Mass Index (BMI), Stress motion, Blood Pressure, Weight, Height*
Data source location	*Marrakech, Morocco.*
Data accessibility	*The dataset is available on GitHub platform via the following link:* https://github.com/hibaasri/Miscarriage-Prediction

Value of the data

- Data is of value to the researches because it is a real data generated.
- Data can be used in the development of other experiments in healthcare area.
- Data can be used for comparing efficiency and effectiveness of data mining algorithms in predicting outcomes.
- The volume of data is prominent for accurate results.
- Data can be used as a benchmark for other researchers for making real test and validate their results.

1. Data

The data includes all risk factors of miscarriage that the mobile application generates from healthcare sensors and mobile phone. The Dataset contains risk factors of miscarriage, patient's personal information and data's file: Age, Heart Rate Variability (BPM), History of Previous Miscarriage (nmisc), Activity, Location, Body Temperature (Temp), Body Mass Index (BMI), Stress motion (stress), Blood Pressure (BP), Weight, Height, Email address, File's Type, File's Saved Time and File's Identifier. All risk factors data are in numeric type for analytical reasons.

2. Experimental design, materials, and methods

Different sources are used to get the data: Mobile phone and healthcare sensors. Table 1 presents a description of each attribute of our dataset.

- Data from sensors:

– Heart rate variability [1],
– Stress and blood pressure [2],
– Temperature variation [3],
– Physical Activity [4].

- Data from mobile phone:

– BMI [5],
– Weight,
– Height,
– Number of previous miscarriages,
– Maternal age [6],
– Location [7],
– Actual activity [4].

Attributes like Weight, Height, maternal age and number of previous miscarriage are collected via a registration form that the patient fills during his first use of mobile application. Location data is

Table 1
Miscarriage dataset attributes.

|	Attribute	Type	Description
1	ID	Integer	The key of JSON document.
2	Activity	Integer	The level of the activity of the woman during the day.
3	Location	Integer	Location where the woman spends her time.
4	BMI	Double	Body Mass Index: It is an attempt to quantify the amount of tissue mass (muscle, fat, and bone) in an individual, and then categorize him/her.
5	nMisc	Integer	The number of previous miscarriages of the woman during her pregnancies.
6	Age	Double	The maternal age of the woman.
7	Weight	Double	The weight of the woman: The quantity of heaviness or mass. It is used in BMI calculation.
8	Height	Double	The height of the woman. It is used in BMI calculation.
9	Temp	Double	Body Temperature of the woman.
10	BPM	Long	Heart Rate Variability (HRV) per minute.
11	Stress	Long	Stress Emotions.
12	BP	Long	Blood Pressure indicator.
13	Time	String	The time to save the file in the database server.
14	User_email	String	The ID of the woman to whom belongs the current document. It is used to extract the right data about woman.
15	Type	String	The type of document. It is used to differentiate between authentication documents and documents that contain prediction attributes.
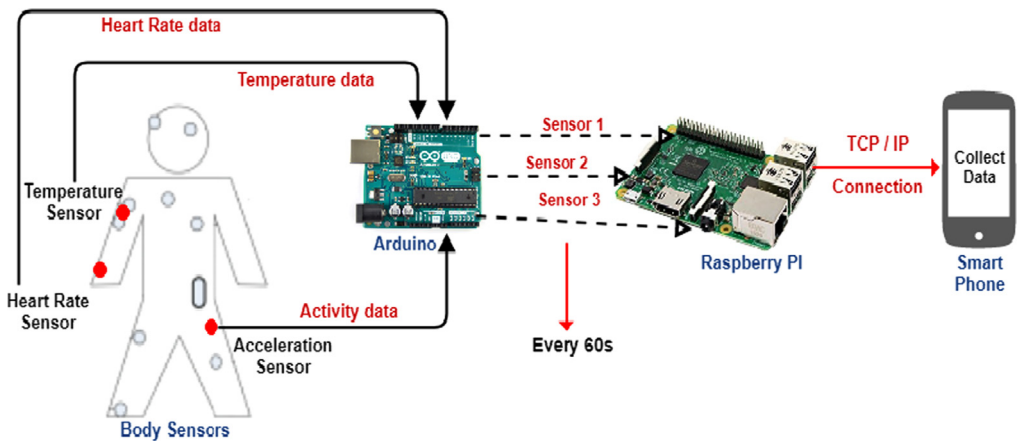
Fig. 1. Gathering sensors data workflow.

collected via GPS mobile tool [8], while actual activity is detected through a predefined machine learning library on android. The BMI is calculated based on height and weight values.

Data from sensors are collected using a microprocessor ARDUINO UNO as it contains many input for linking wires of sensors and sent to Raspberry Pi 3 which is a Nano-computer where process is done [9] (see Fig. 1). It collects data every 60 s and send it to our mobile phone application to be analyzed in a Big Data Platform.

Transparency document. Supplementary material

Transparency data associated with this article can be found in the online version at https://doi.org/10.1016/j.dib.2018.05.012.

References

[1] J.F. Thayer, F. Åhs, M. Fredrikson, J.J. Sollers, T.D. Wager, A meta-analysis of heart rate variability and neuroimaging studies: implications for heart rate variability as a marker of stress and health, Neurosci. Biobehav. Rev. 36 (2) (2012) 747–756. http://dx.doi.org/10.1016/j.neubiorev.2011.11.009.

[2] H. Tanaka, K. Monhah, D. Seals, Age-predicted maximum heart rate revisited, J. Am. Coll. Cardiol. 37 (1) (2001) 153–156.

[3] O. Anselem, D. Floret, V. Tsatsaris, F. Goffinet, O. Launay, Influenza infection and pregnancy [French; English] Grippe au cours de la grossesse, Press Med. 42 (11) (2013) 1453–1460. http://dx.doi.org/10.1016/j.lpm.2013.01.064.

[4] E.Y. Wong, R. Ray, D.L. Gao, et al., Physical activity, physical exertion, and miscarriage risk in women textile workers in Shanghai, China, Am. J. Ind. Med. 53 (2010) 497–505. http://dx.doi.org/10.1002/ajim.20812.

[5] C. Toma, Z. Veleva, A. Tiitinen, et al., High and low BMI increase the risk of miscarriage after IVF/ICSI and FET, Hum. Reprod. 23 (4) (2008) 878–884.

[6] E. Rochebrochard, P. Thonneau, Paternal age and maternal age are risk factors for miscarriage; results of a multicentre European study, Hum. Reprod. 17 (6) (2002) 1649–1656. http://dx.doi.org/10.1093/humrep/17.6.1649.

[7] P. Women, Food safety for pregnant women, Intern. Med. News 38 (2005) 79. http://dx.doi.org/10.1016/S1097-8690(05)72089-8.

[8] A.K. Dubey, Global positioning system, in: Understanding an Orogenic Belt, Springer International Publishing, 2014, pp. 215–230.

[9] Raspberry Pi and Arduino – Raspberry Pi, 2017. ⟨https://www.raspberrypi.org/blog/raspberry-pi-and-arduino/⟩, (Accessed 25 October 2018).