Journal of Big Data

RESEARCH Open Access

Toward a smart health: big data analytics and IoT for real-time miscarriage prediction

Hiba Asri* and Zahi Jarir

*Correspondence:
Hiba.asri@uca.ac.ma

LISI Laboratory, Department of Computer Sciences, Faculty of Sciences Semlalia, Marrakech, Morocco

Abstract

Background: We are living in an age where data is everywhere and grows up in a very speedy way. Thanks to sensors, mobile phones and social networks, we can gather a hug amount of information to understand human behavior as well as his individual life. In healthcare system, big data analytics and machine learning algorithms prove their effectiveness and efficiency in saving lives and predicting new diseases. This triggered the idea of taking advantages of those tools and algorithms to create systems that involve both doctors and patients in the treatment of disease, predict outcomes and use real-time risk factors from sensors and mobile phones.

Methods: We distinguish three types of data: data from sensors, data from mobile phones and data registered or updated by the patient in a mobile app we created. We take advantages from IoT systems such as Raspberry Pi to collect and process data coming from sensors. All data collected is sent to a NoSql Server to be then analyzed and processed in Databricks Spark. K-means centroid clustering algorithms is used to build the predictive model, create partitions and make predictions. To validate results in term of efficiency and effectiveness, we used clustering validations techniques: Random K, Silhouette and Elbow methods.

Results: The main contribution of our work is the implementation of a new system that has the capability to be applied in several prediction disease researches using Big Data Analytics and IoT. Also, comparing to other studies in literature that use only medical or maternal risk factors from echography; our work had the advantage to use real-time risk factors (maternal and medical) gathered from sensors, react in advance and track diseases. As a case study, we create an e-monitoring real-time miscarriage prediction system to save baby's lives and help pregnant women. In fact, doctors receive the results of clustering and track theirs patient through our mobile app to react in term of miscarriage to avoid non-suitable outcomes. While pregnant women receive only advices based on their behaviors. The system uses 15 real-time risk factors and our dataset contains more than 1,000,000 JSON files. Elbow method affirm three as the optimal number of clusters and we reach 0.99 as a value of Silhouette method, which is a good sign that clusters are well separated and matched.

Keywords: Big Data, Miscarriage prediction, Predictive analytics, K-means, Clustering

Introduction

Big data and reality mining improve well from sciences to several fields including healthcare, industry, education, agriculture … among others. Nowadays, sensors, mobile phones and machines are generating a huge amount of data every short time. In this context, Artificial Intelligence comes to help us making decisions, react in advance to avoid non-suitable outcomes and predict things that were impossible before [1].

Several researchers are creating new models, new algorithms and new knowledge to make machines act like humans and make decision in real-time. For instance, Google can propose answers for you in a short time, Amazon can suggest books for you, and our smartphones suggest and add words to their dictionaries based on the context. Another famous example is Google's Flu Trends that collect the flu activity based on Google Search. In fact, Google compares between people who search for the flu and people who have flu symptoms [2].

As its core, AI and Big Data are about making predictions and react in the right moment. Today, big data became a synonymous of data mining and predictive analytics, and change the process from reporting and decision to predicting results. In other aspect, we think that the challenge in using Big Data is not just collecting and storing data, but it is about creating values and make machines act and think like humans to make complex tasks and make people's live easier [3–5].

After all, reality mining and data mining techniques have the capabilities to understand our individual and collective behaviors [6]. In this context and with respect to all studies in background section, we created a new e-monitoring system for real-time diseases prediction [7]. This system can be applicable to any case study of disease prediction by using the appropriate data inputs. The proposed system benefits from the use of big data tools, machine learning algorithms and Internet of Things, in order to offer a performant model with accurate results. As a case study, we applied the proposed system on a real-time miscarriage prediction where we used K-means centroid base algorithm for clustering data and get accurate results (Miscarriage / No Miscarriage / Probable Miscarriage) [8]. The proposed model used a dataset of 15 attributes where only 11 are real-time risk factors of miscarriage [9]. The following study enhance the previous one by:

- Adding more risk factors in the training process. In fact, the present dataset added four features and contains 15 attributes that are all real time risk factors of miscarriage.
- Using only the raspberry pi instead of both Arduino UNO and Raspberry for collecting and processing the model of prediction to minimize response time for doctors and patients.
- Clustering data into three clusters: Miscarriage (M = 0), Probable Miscarriage (PM = 1) and No Miscarriage (NM = 2).
- Reaching a good accuracy of 99% using Silhouette method, which is a good metric to validate experiment using clustering algorithms.

Data represents the key-value of all researchers since having good and accurate results depends on inputs quality. In healthcare, data is used in several studies and researches to propose new systems, new algorithms and new methodologies; in order

to make predictions, help patients and make people's lives healthier. In addition, machine learning algorithms, IoT tools and reality mining show their power in making prediction and decision-making [10–13].

The rest of this paper is organized as follows:

- Background section highlights studies of predictive analytics in several areas such as education, agriculture and healthcare; and discusses important works of miscarriage prediction.
- Method section discusses the architecture of the proposed model, methodology and implementation.
- Experiment section presents the experiment environment, dataset as well as the use of the predictive model K-means in Databricks Spark. In the same section, we discuss results in term of efficiency, performance and clustering accuracy.
- Evaluation section, presents how we evaluate the model using clustering methods including Random K, Relative Clustering Validation RCV, Internal Clustering Validation ICV and External Clustering Validation ECV.
- Conclusion and future work section concludes the paper and highlights future work.

Background: data mining opportunities and disease prediction

Authors in [14] present a literature review about the use of predictive analytics using data mining algorithms, for medical industry and business. Authors assume that predictive analytics using datamining methods are a powerful tool to make predictions about future outcomes. They also mention that the process of prediction start from historic data analysis and evaluation to get prediction results.

In another study [15], authors discusses the IT-based methods for disease prevention. In fact, they present a state-of-the-art data analytics models used for classification and clustering of diseases, anomalies detection, as well as their advantages, and guidelines for choosing the accurate model in each situation.

In [16], authors made a comparison of four machine learning classification algorithms: Support vector Machine (SVM), Decision Tree (C4.5), Naïve Bayes (NB) and K-Nearest Neighbor (K-NN). In this study, they use the cited algorithms to predict breast cancer, using the Wisconsin Breast Cancer (WBC) dataset to train models. As a result, SVM shows its power in term of efficiency with an accuracy of 97.13%. Authors enhance this study by proposing a hybrid data mining classifier for breast cancer prediction using the same dataset WBC. Experimental results show that the classification using fusion of SVM, NB and C4.5 reached a highest accuracy of 97.31% [17].

In [18], authors propose an interesting view of the role of datamining and predictive analytics in the healthcare services. They affirm that data mining are useful in some areas of healthcare system such as determination of clinical pathways and quality of care, while it stills an emerging axe of research in other situations. First, they examine existing works associated with healthcare delivery. Second, they provides

a multi-layered framework for analyzing data mining algorithms in healthcare service management. Third, they used two approaches of deductive and inductive fitting to classify the research. Furthermore, they reveal that data mining successes in three main healthcare applications namely: disease pathways, healthcare capacity and enhancing the quality of care.

In another study [19], authors propose a new algorithm that predict the next intercell behavior of a mobile user covered by a personal communication systems network. In the prediction process, the new system includes three main phases: information of mobile user trajectories as user mobility patterns, deducing rules from the gathered patterns and lastly predicting outcomes. Experiments show that the new algorithm shows its power in term of precision and recall comparing to other methods.

Authors in [20] develop a process model for actionable mining "Actionable Mining and Predictive analysis" for public safety and security. This process includes multiple steps namely: question or challenge, fusion of data and collection, operationally preprocessing (recoding and selection), identification and modeling, evaluation of public safety and security, and finally actionable output.

In [21], authors assert that Body Mass Index (BMI) represents an interesting risk factor of miscarriage during pregnancy. In fact, women who suffer from obesity or underweight have more chance to have miscarriage than women with a normal weight. Thus, pregnant women with high or low BMI have a higher risk of having miscarriage than females with a normal BMI.

Another study [22] demonstrates that vaginal blood loss, uterine adenomyosis, C-reactive protein and premature rupture of membranes are significant and independent risk factors during the second-trimester of pregnancy. In this study, they use data collected from patients under miscarriage treatment in hospital, to be then analyzed using logistic regression algorithm to determine the most significant risk factors.

In [23], authors conduct a study to estimate the burden of miscarriage in the Norwegian population and to evaluate the dependency between miscarriage, maternal age and miscarriage recurrence. 421 201 pregnant women participated in this study. As a result, women aged 25–29 (10%) have a lowest chance to have miscarriage compared to pregnant women aged 30 and over. In addition, they notice that the recurrence of miscarriage expand with women aged 45 and over. Lastly, they conclude that maternal age and history of miscarriage are greatly good risk factors that produce unwanted pregnancy outcomes.

Authors in [24] propose a system that make medical data available, especially of patients who travel in different countries. The proposed approach reduces the time of requests by a noticeable margin. The proposed system takes around three seconds to upload and 7.5 s to download 250 Mb of data. The system has a latency of 413.76 ms when retrieving 100 records, compared to other systems. Security and performance of the system are insured using a blockchain.

Table 1 presents a summary of several studies that have been conducted for making predictions. Through this analysis, we analyze and compare works in term of type of prediction, algorithms and methods used for collecting data and making predictive models, advantages and disadvantages of the study.

Table 1 Empirical Analysis of diseases prediction in healthcare system

References	Outcomes	Methods	Real-time Dataset	Advantages	Disadvantages
[25]	Heart disease	j48, Naïve Bayes, REPTREE and CART	No	-Result shows that prediction accuracy is 99% -j48, REPTREE, and CART gave a prediction model of 89 cases with a risk factor positive for heart attacks	Tasks like heart attack still complex to predict
[26]	Chronic disease	Fog Computing	Yes Sensors	-Monitor patients suffering from chronic diseases -Big number of data -Sort out context-sensitive data	not readily available software and complexity of the architecture
[27]	Prognosis of coma	ARIMA Autoregressive integrated moving average	No electroencephalogram dataset	-Early prognosis -The use of biological signals and context data -The use of the response loop	Not suitable for real-time dataset
[16]	Breast Cancer Prediction	C4.5, Naïve Bayes,SVM and K-NN	No Wisconsin Breast Cancer Dataset	-A comparison of four machine learning algorithms -The highest accuracy a xchieved is 97,13% using SVM	Small dataset
[17]	Breast Cancer Prediction	C4.5, Naïve Bayes,SVM and K-NN	No Wisconsin Breast Cancer Dataset	-Hybrid Model using the fusion of SVM, NB and C4.5 -The highest accuracy Achieved is 97,31%	Small dataset
[7]	Miscarriage Prediction	K-means centroid clustering	Yes Healthcare sensors Mobile phone	-An E-monitoring system -The use of IoT and Predictive analytics -Doctors and patients are involved in the treatment of the disease	More factors need to be included
[23]	Miscarriage Prediction	Prospective register based study	No Data from Registration	-Evaluate dependency between miscarriage, maternal age and miscarriage recurrence	A focused number of Pregnant women
[19]	Location prediction	Proposed model	No History of trajectories	-The use of mobile user covered by a personal communication systems network	Not applicable on new cases

Table 1 (continued)

References	Outcomes	Methods	Real-time Dataset	Advantages	Disadvantages
[28]	Medical Emergency Response	Proposed Model	Yes IoT sensors	-The capability of processing WBAN sensory data from multiple users -Real–time responses in case of emergencies	Security and privacy must be highlighted to avoid any ethical issues

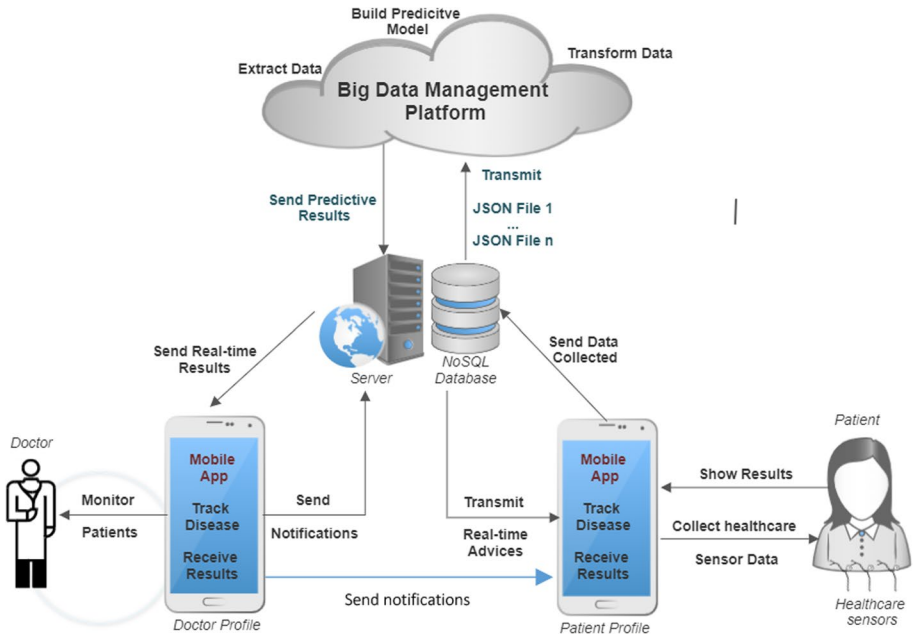

Fig. 1 E-monitoring system for real-time disease prediction architecture

Methods

Real-Time disease prediction: a proposed system

Nowadays, people suffer from several diseases that they could avoid if they take specific treatment earlier. In fact, individuals go to make a checkup just in case of severe symptoms; but it is often too late to deal with a disease at late level. In addition, people do not have an overview about their health until they make a consultation or a radiography. Therefore, patients are unaware about their health problems and are not involved in the treatment of diseases.

This triggered the idea of creating an e-monitoring system for real-time disease prediction (see Fig. 1). The system can take decision in real-time to make people lives healthier and safer. Thanks to this system, we can react in advance, because it is not just about predicting diseases but it is about preventing also. The main contributions of this system are:

- Applying the system in many case study: heart-attack prediction, obesity prediction, miscarriage prediction, skin cancer prediction….

- Involving doctors and patients in treatment of the disease.
- Using of healthcare sensors to collect real-time data: activity, heart rate, body temperature, alcohol consumption …
- Using of mobile phone to collect real-time data from the profile of the user: age, previous diseases, BMI, weight …
- Using IoT tools to collect and process data gathered.
- Using Big Data platform to train, analyze and create the predictive model.
- Making a mobile phone application with multiple profiles (doctor's profile and pregnant woman's profile) to communicate and react in advance in critical cases.
- Making smart bracelet to incorporate the system for an easy use.

The proposed system can be divided into four main parts:

1. Front system
   - A profile for doctors in a mobile application to track and monitor their patients, as well as make decisions in real-time.
   - A profile for patients to track the status of their disease and to receive recommendations to avoid unsuitable outcomes.
   - A background services to collect data in real-time from mobile phones.
2. Big Data Management system
   - Collect data received from mobile phones and healthcare sensors.
   - Analyze and transform data.
   - Build the predictive model.
   - Evaluate the model.
   - Transmit results to doctors.
3. IoT technologies
   - Healthcare sensors
   - Tools of processing and retrieving data from sensors.
4. Data management
   - A performant server that store and manage data in files to be then extracted by the Big Data Management tools.

Case study: real-time miscarriage prediction

Miscarriage represents one of the most outcome that hurt family's lives, especially the pregnant women [29]. In fact, during the pregnancy, the woman must do regular consultations to see how her baby is growing and how healthy he is. In some cases, the pregnant woman is scandalize by the death of his baby, and in some cases, she could save his life by reacting in advance. So here, we notice that saving lives depends on having the accurate information in a specific moment. That pushed us to use the previous system for a miscarriage prediction on real-time.

To be able to use the proposed model for miscarriage prediction, pregnant women are equipped by several healthcare sensors to collect real-time health data. Additional data are gathered from mobile phone through the mobile application that we created. All those data; in JSON format; are then sent to a database server to be recovered by

the Big Data Platform on real-time. Data collected is trained, transformed and then covered by the predictive model to get results of miscarriage. Not to mention that before sending results, the predictive model must be evaluated using accurate evaluation method of clustering.

Miscarriage outcome remains a critical information. Thus, we choose to send results of miscarriage (M, PM or NM) to the doctor that will ask for an urgent consultation in some critical cases to save the baby's life, as well as the life of the mother. While, pregnant women will receive recommendations based on her real behavior during her pregnancy. Recommendations are related to her well-being, food quality, activity …

Reality mining for gathering data

Reality mining and Big Data are powerful ways that help researchers and developers understand human's individual behaviors. Thanks to the advancement of technologies and tools, we can now collect data that were impossible to gather earlier. In fact, from sensors and mobile phones, we can attend a huge amount of data every second about a person. Thus, our system takes benefit from those tools to collect all possible information about pregnant women. We can distinguish multiple source of data (see Fig. 2):

- Data from healthcare sensors.
- Data from mobile phone sensors.
- Data from the patient's profile.

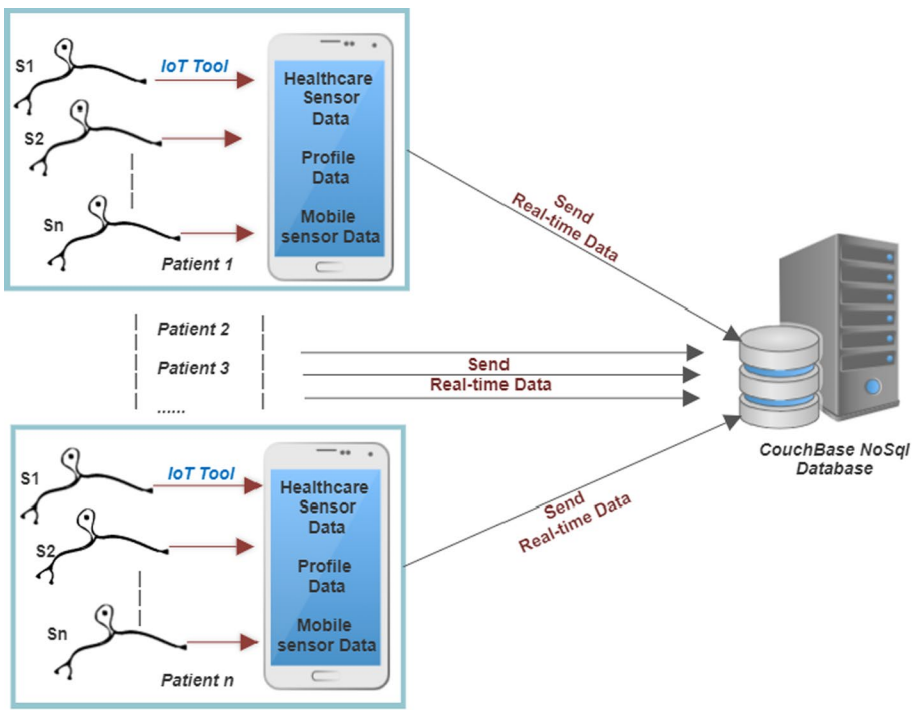

Fig. 2 Reality mining for collecting real-time healthcare data in the proposed system

All the above data; that contains risk factors related to disease prediction; are filled in JSON files every minute to be then sent to a server to be analyzed and processed. We used Couchbase Server to store Files in clusters [30].

IoT implementation for healthcare data

Sensors are the source of data of IoT and become so advanced that can collect as much data you want. In healthcare system, having accurate information is crucial and influential because we deal with life and death of the person. In our case, we used multiple healthcare sensors: temperature sensor, pulse sensor, acceleration sensor and alcohol/drunk sensor (see Table 2).

To manage and collect data from sensors, multiple IoT tools exist such as:

- Arduino Uno [31]: a small electronic board equipped with a microcontroller and an open source platform for gathering data.
- Raspberry Pi 3 Model B [32]: a single-board computer for programming and processing.

In this study, we just implement raspberry pi for collecting and processing data. Before we opt to use Arduino Uno for collecting data and Raspberry pi for processing; but we notice that we consume much time, while we have to reduce the time response instead. Lastly, we transfer all data collected to the mobile phone application "e-preg monitoring" as shown in Fig. 3.

Predictive model

Prediction process

To build our predictive model, we used Databricks Big Data management tool based on Spark [33, 34]. The choice of working with Spark instead of other EcoSystems such as Hadoop is due to the need of processing streaming data. Using Hadoop, we can only do batch processing which not reply to our demand since we need results in real-time to save lives [35]. For creating the model and use K-means clustering algorithm to group our data, we used Machine Learning Library of Spark (MLlib) [36] by following the below four steps:

Table 2 List of Healthcare sensors

Sensor	Feature collected
Pulse sensor[a]	Heart rate variability, blood pressure and emotion
Temperature sensor[b]	Temperature variation
Acceleration sensor[c]	Activity degree
Alcohol / Drunk sensor[d]	Alcohol consumption

[a] Heart rate sensor pulse sensor for Arduino Raspberry Pi

[b] I2 C infrared temperature sensor mlx90614 for Arduino Raspberry Pi/0.1 °C accuracy

[c] kwmobile acceleration sensor with digital output, 3-axis gyroscope for Arduino, Genuino and Raspberry Pi

[d] Alcohol Sensor: MQ-3, Alcohol Gas Sensor

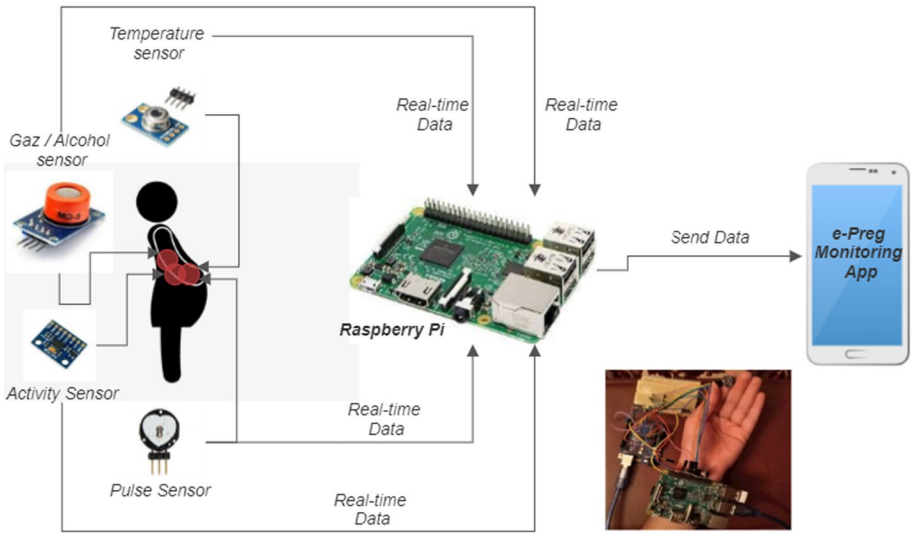

Fig. 3 IoT implementation for gathering healthcare sensor data

1. Collect and store data: Data are collected from different sources: healthcare sensors, mobile phone and patient's profile. Data collected is then stored in database server on real-time in JSON files.
2. Extract data from database: In Databricks platform, we extract data from Couchbase server using a listener that order to do extraction when new data arrived.
3. Make transformation

a. In Spark, data must be numeric for analysis.
b. In Spark, we use RDD (Resilient Data Distributed) to pass data to Kmeans clustering algorithm.

4. Train dataset: function "train" is used to train the model, by passing the following parameters: Number of clusters, Data and Number of iterations.
5. Make predictions: define the appropriate cluster of each sample in the dataset.
6. Evaluate the model: Evaluation is an essential step during prediction since having accurate model depends on whether or not we reach good accuracy of predicting outcomes.

Kmeans clustering algorithm

Machine learning algorithms; which are a part of AI; remain performant models to learn and predict outcomes with high accuracy. Several families of machine learning exist: supervised learning, unsupervised learning and reinforcement learning. Many applications in several domains such as education, healthcare and agriculture use data mining techniques in the process of prediction [37].

Among unsupervised machine learning algorithms, we distinguish clustering algorithms like K-means, Hierarchical clustering HCA, Expectation Maximization EM … among others [38–40]. The challenge here is how to choose the right algorithm to work with. According to [41], multiple parameters must be considered: size of the data, number of clusters, type of dataset …. For large dataset, K-means and EM algorithms show

their performance and efficiency. While HCA and SOM become very good when dataset is small.

In our study, we opt for K-means algorithm based centroid clustering since the size of our data is large, its simplicity to understand and its popularity. When using K-means algorithm, we calculate the probability of the most relevant function. Then functions are grouped using Euclidian distance. K is the number of clusters that will group your data into K clusters. The main goal of K-means is to minimize the squared error objective function given by:

$$J(V) = \sum_{i=1}^{C} \sum_{j=1}^{c_i} (||x_i - v_i||)^2$$

where.

- "$||x_i - \text{vj}||$": Euclidean distance between $x_i and v_i$
- $\text{X} = \{\text{x}_1, \text{x}_2, \text{x}_3 \ldots \text{x}_\text{n}\}$ is the set of data points and $\text{V} = \{\text{v}_1, \text{v}_2 \ldots \text{v}_\text{c}\}$ is the set of centers.

Experiment

Experiment environment

Table 3 presents the experiment environment of our study. We used Databricks Spark as a big data management tool for processing and building the predictive model, Couchbase for storing data, IoT tools and healthcare sensors for collecting and processing real-time data. Android programming language is used for creating the mobile application.

Experiment dataset

In this study, we used sensors and mobile phones to collect data about pregnant women. The current system use a dataset that contain more than 1 000 000 JSON files of 50 pregnant women that accept to contribute in this study [52]. It contains 15 features that are all miscarriage risk factors; gathered from mobile phones and sensors during the period of 2019 and 2021 (see Table 4).

Prediction process steps in Databricks Spark

As shown in Fig. 2 and Fig. 3**,** we send all data gathered to a database server for storage and then retrieved by Big Data Platform on real time. Once data is available in Databricks Spark Cloud, we have to build our predictive model to predict whether a pregnant

Table 3 Experiment environment

Big Data platform	Databricks 10.4, Spark 3.2.1 Programming language: Scala
Database server	Couchbase Server 7.1 hosted with a public IP address
IoT tools	Raspberry pi 3 Model B (Linux OS) Programming Language: Python
Sensors	Pulse sensor, temperature sensor, alcohol sensor and acceleration sensor
Mobile tools	Arduino Studio for creating our mobile application "e-preg monitoring"

Table 4 Miscarriage Dataset description

	References	Risk factor	Description	Sources
1	[42]	Heart Rate variability (HR)	A marker of stress. A high value of HR is a sign of a hypertension and elevated blood pressure	Healthcare Sensors
2	[43]	Stress and Blood Pressure (BP)	Values define based on the value of HRmax	
3	[44]	Temperature variation (TP)	A spontaneous miscarriage or premature delivery is associated with any viral infection. A higher Body temperature and flu increase the risk of having a miscarriage	
4	[44]	Physical Activity	An extreme activity of the body is associated with an elevated risk of miscarriage	
5	[45]	Alcohol Consumption	The level of alcohol consumption Ethyl alcohol value in the human body	
6	[45]	Drunk	State of the drunk level	
7	[46]	BMI	A sign of obesity and underweight BMI = w(kg) / H(m) Where W and *H* are respectively the weight and the height of the pregnant woman	Mobile Phone
8	[47]	Number of previous miscarriages	miscarriage is categorized by previous number previous pregnancy losses	
9	[23]	Maternal age	Increased maternal is associated with an increased chance of having miscarriage	
10	[48]	Location	Food safety and eating well play well during pregnancy to avoid getting an illness	
11 to 15	[44]	Current activity (running, walking, biking,)	Activities like running, walking, biking, among others can encourage having miscarriage	

```
Cmd 11

val rowsRDD = allDF.rdd.map(r => (r.getString(0), r.getInt(1), r.getInt(2),
                                  r.getInt(3), r.getInt(4), r.getInt(5), r.getInt(6),
                                  r.getInt(7), r.getInt(8), r.getInt(9), r.getInt(10),
                                  r.getInt(11), r.getInt(12), r.getInt(13), r.getInt(14), r.getIn(15) ))
rowsRDD.cache()

rowsRDD: org.apache.spark.rdd.RDD[(String, Int, Int, Int, Int, Int, Int, Int, Int, Int, Int, Int, Int, Int, :
3844494367788461:1
res208: rowsRDD.type = MapPartitionsRDD[3290] at map at command-3844494367788461:1

Command took 0.58 seconds -- by asri.hiba@gmail.com at 27/04/2022, 12:34:12 on My Cluster
```

Fig. 4 RDD creation for Kmeans

woman is having a miscarriage or not. To build K-means model using Spark, we follow the steps below:

Transformation

We have to convert data to RDD to be passed to Kmeans algorithm as shown in *Fig. 4*.

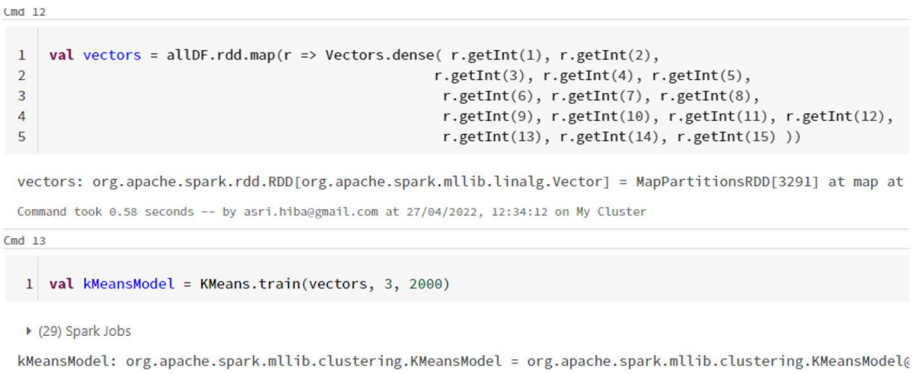

Fig. 5 Kmeans training model

Fig. 6 Prediction of clusters using Kmeans model

- *AllDF* represents the dataset
- *r.getInt(1) … r.getInt(15)* represents the following attributes: *Age: Int, BMI: Int, Nmisc:Int, Activity: Int, Biking: Int, Walking: Int, Driving: Int, Sitting: Int, Location: Int, temp: Int, bpm: Int, stress: Int, bp: Int, alcohol: Int*

Train the model

To train K-means algorithm (see Fig. 5), we pass several parameters: Data: "Vectors", Numbers of clusters: "3"Number of iteration: "400".

Make predictions

We Predict result using Predict function that dense vectors of attributes as described in Fig. 6. r._1 to r._15 represent the following attributes: Age: Int, BMI: Int, *Nmisc:Int, Activity:*

Table 5 Performance parameters of k-means

Parameters	K = 1	K = 2	K = 3	K = 4
Time to build the model (s)	1.28	1.69	2.85	3.80
Time for cluster Distribution (s)	0.69	0.96	0.95	1.08
Time for center Definition (s)	0.35	0.62	0.40	0.49
Time for Model evaluation (s)	47.65	48.52	49.89	48.28

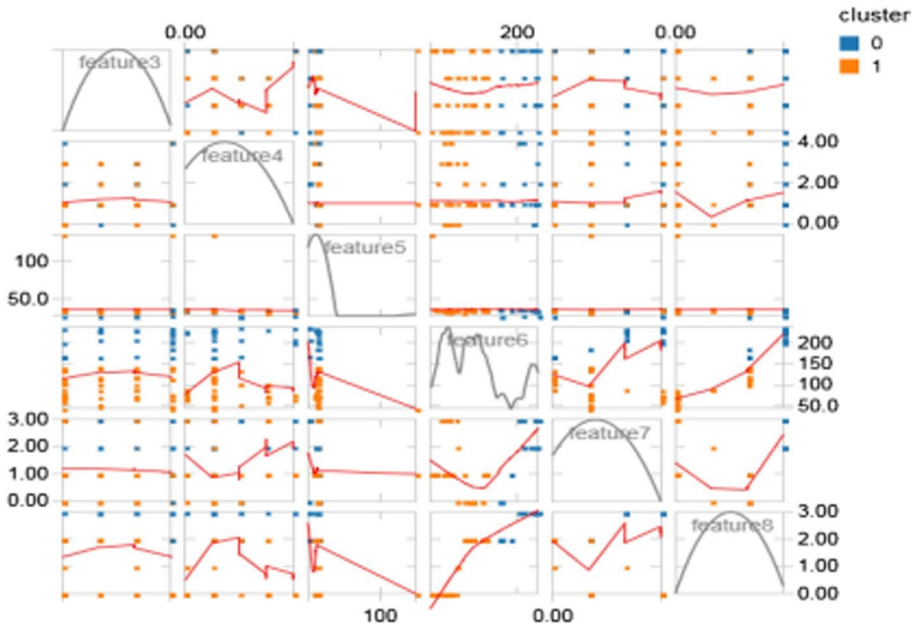

Fig. 7 Scatter Plot for 2 clusters and 11 risk factors

Int, Biking: Int, Walking: Int, Driving: Int, Sitting: Int, Location: Int, temp: Int, bpm: Int, stress: Int, bp: Int, alcohol: Int.

Send results

Spark send results of predictions to doctors through the mobile application "e-preg monitoring" to act on the right time in case of a probable miscarriage. While we send recommendations to patients to avoid psychological problems.

Experiment results and discussion

Model efficiency

In healthcare system, time response is a crucial metric because we are dealing with the life or the death of the person. In our study, we train our model with different values of K to compare the model efficiency in term of time to train the model, time to create clusters, time to define centroids and finally time to evaluate the model (see Table 5).

Table 6 Features meaning

Feature	Meaning	Feature	Meaning
Feature 0	Age	Feature 7	Sitting
Feature 1	BMI	Feature 8	Location
Feature 2	Nmisc	Feature 9	Temp
Feature 3	Activity	Feature 10	Bpm
Feature 4	Biking	Feature 11	Stress
Feature 5	Walking	Feature 12	Bp
Feature 6	Driving	Feature 13	Alcohol
		Feature 14	Drunk

Table 7 Summary of features

Summary	Mean	Stddev	min	25%	50%	75%	max
Features							
Age	22.000	0.000	22	22	22	22	22
BMI	17.000	0.000	17	17	17	17	17
Nmisc	1.558	1.134	0	1	2	3	3
Activity	2.000	1.500	0	1	2	3	4
Biking	0.510	0.500	0	0	1	1	1
Walking	0.485	0.500	0	0	0	1	1
Driving	0.005	0.070	0	0	0	0	1
Sitting	0.512	0.500	0	0	1	1	1
Location	2.005	1.409	0	1	2	3	4
Temp	37.782	1.788	35	36	38	39	41
Bpm	129.157	56.093	43	79	119	167	238
Stress	1.134	1.114	0	0	1	2	3
Bp	1.522	1.099	0	0	2	2	3
Alcohol	476.539	216.925	100	295	475	659	854
Drunk	1.891	0.413	0	2	2	2	2

K-means scatter plot

Scatter plot represents an interesting graph that represents the distribution of clusters and the correlation between features and foresee trends. Because of the size of dataset and the big number of features, the scatter plot cannot be visualize. Databricks Community Platform only suggest a scatter plot for the first 1000 rows. Thus, clusters are not all viewed in the whole plot because of the small variation in the first rows. Figure 7 illustrates a previous scatter plot of a previous study for two clusters. We can notice variation of clusters because the dataset is large and we can have variation of clusters in the first 1000 rows. Table 6 and Table 7 represent respectively the meaning of features and the summary of the features.

Results transmission for doctors

Once the predictive model determines clusters, results are then transmitted to doctors in a mobile application we created. In his profile, the doctor lists his patients, tracks each pregnant woman and receives the results of having miscarriage on real-time (see Fig. 8).

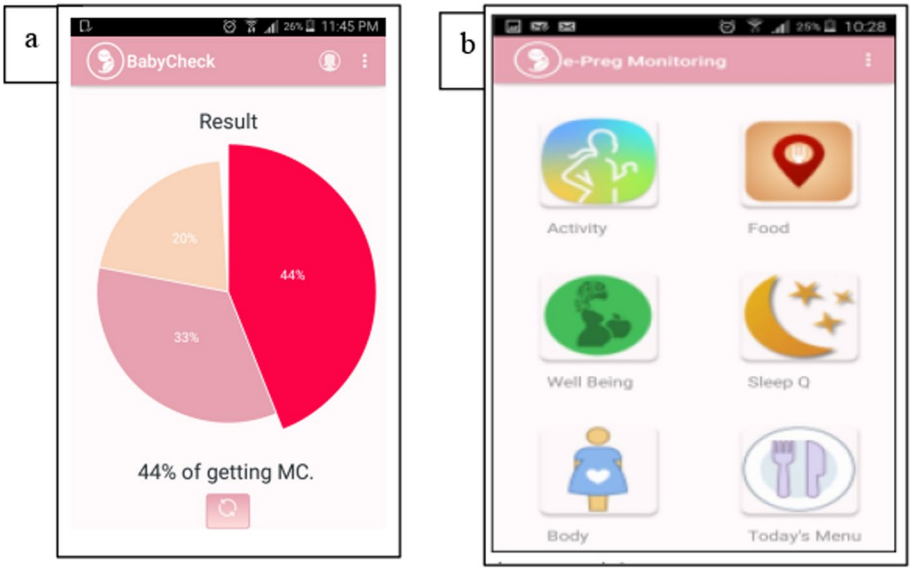

Fig. 8 **a** Results in Doctor Profile **b** List of recommendations for pregnant women

Thus, in case of a high percentage of miscarriage, doctors can notify the woman to come for a consultation to avoid unwanted outcomes.

Recommendations for patients

During Pregnancy, the woman experiences an emotional imbalance due to changes in hormones and body. Sending bad news is not appropriate thing since it can cause some psychological problems. Therefore, we preferred to send only recommendations based on her behavior during pregnancy as shown in Fig. 8. For example, if we notice that the pregnant woman is spending most of her time in snacks and restaurants, the system will notify her by a message: "Please eat healthy foods! You are pregnant and you should take care of your future baby nutrition". Pregnant woman must eat healthy foods, and preferably, those made at home.

Model evaluation

The main goal of K-means algorithm is to minimize the sum of squares of distance (Euclidian Distance for example) among the points of each cluster. It remains important to evaluate the obtained model and check whether it represents data correctly. In this study, we evaluate and validate the model through a number of techniques:

- Within Set Sum of Squared Errors (WSSSE): Evaluation with a random K
- Clustering validation techniques:
    - Relative Clustering Validation (RCV)
    - Internal Clustering Validation (ICV)
    - External Clustering Validation (ECV)

Table 8 Values of WSSSE for different values of K

Value of K	Value of WSSSE (Thousand)
1	1,01E+24
2	3,89E+23
3	1,01E+23
4	1,24E+23

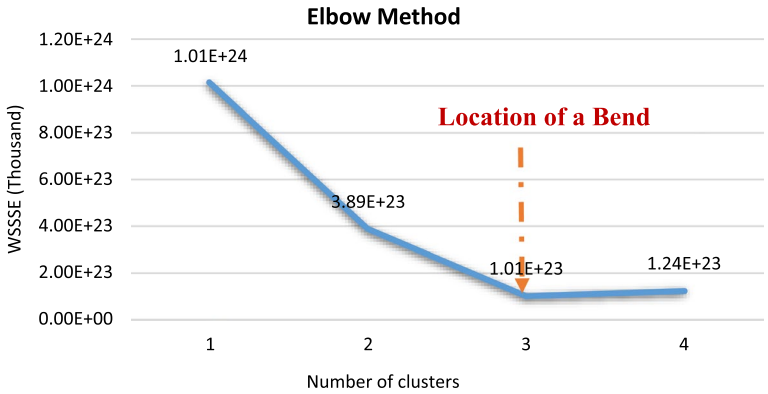

Fig. 9 Elbow method results

Within Set Sum of Squared Errors (WSSSE): evaluation with a random K

When using K-means, we define three as a value of "K". However, we have to validate this value and check if it is the appropriate value of K in our case or not. Within Set Sum of Squared Error represents a good metric to evaluate models, since it is the sum of the distances of each point in each "K" cluster.

Normally, we get better results when the number of cluster is important. In spite of that, having a high value of K is not always a good indicator of accurate outcomes. The context of the case study and meaning of clusters is considered. To use evaluation by a random "K", we calculate WSSSE of k-means for four values of "K" as presented in Table 8. We reach the highest value of WSSSE when k = 1 and k = 2 while we attend lowest WSSSE when k = 3 and k = 4. Therefore, we need to validate either grouping our data into three or four clusters through the following sections.

Relative Clustering Validation (RCV)

RCV evaluates the general structure of clustering method and determines the optimal number of K through its "*Elbow*" method. In fact, we represent a graph of the correlation between the value of "K" and its WSSSE value. The optimal number of K is when we notice a bend (knee) in the graph.

In Spark, to obtain the result of Elbow method, we called the function *"fviz_nbclust"*:

$$fvic\_nbclust, kmeans, method = \text{wss} + geomylinexintercept = 4, linetype = 2 + labssubtitle = "Elbow\ method"$$

Table 9 Different values of Silhouette width in clustering models

Value of Silhouette width S_i	Meaning
Almost 1	Observations are well grouped
Around 0	Too few or too much clusters Points are not well clustered in their own cluster
Negative value	Observations are not in the appropriate partition

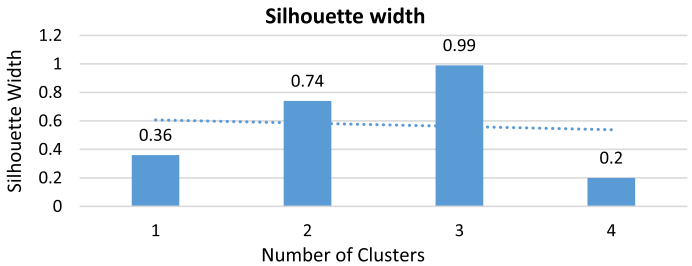

Fig. 10 Results of Silhouette method

From the plot in *Fig. 9*, we notice the location of two bends. The first is when k = 2 with a high value of WSSSE (3,89E + 23). This case is to avoid since the error rate is important. The second one is located when K = 3 with a lowest error rate. At this age, elbow method mention that clusters are well structured in the model when we group data into three partitions. For us, it is a good sign since our goal is to group our data into three outcomes: M, PM and NM. Choosing k = 1 (Highest WSSSE) is not appropriate, while choosing k = 4 (lowest WSSSE) do not satisfy our aims in our case study.

Internal Clustering Validation (ICV)

In contrary of RCV, Internal Clustering Validation ICV evaluates the density and the segregation between clusters. We focus on how a collection is separated from other partitions and how points are distributed in the same cluster. The majority of ICV methods include both density (compactness) and separation to calculate indexes (α and β are weights):

$$Index = \frac{\alpha * Separation}{\beta * \text{Density}}$$

Silhouette method is one of the most valuable evaluation method in ICV because of its robustness. We can deduce how well a point is grouped by calculating the average distance between clusters. In other words, this method helps to rate the distance from an observation in cluster A to all points in cluster B. For each observation *i*, the silhouette width S_i is calculated as follows

1. For each point *i,* we calculate the average variation a_i between *i* and all other observations in the same cluster of *i*.

2. We calculate the average variation d(i,C) of i to all points of C (other clusters to which i does not belong). We call the lowest value of this variation bi = $\min_C$ d(i,C) that represents the dissimilarity between *i* and its nearest cluster.
3. At last, we get the silhouette width (a number between -1 and 1) of the observation *i* through the formula below. Meanings of values of silhouette width are described in Table 9.

$$Index = \frac{b_i - a_i}{\max(a_i, b_i)}$$

From the plot in Fig. 10, we notice that we attend almost 1 with three clusters. Silhouette method validate the elbow method and prove the compactness of groups when k = 3. We used clustering evaluator package of MLlib Spark to compute Silhouette width (see Fig. 11).

External Clustering Validation (ECV)

External clustering validation remains on comparing results of experiences with other studies working on the same case study prediction. As we mentioned earlier, miscarriage prediction researches in literature use either maternal factors or echography to predict

```
// Evaluate clustering by computing Silhouette score
val evaluator = new ClusteringEvaluator()

val silhouette = evaluator.evaluate(predictionsBKM)
println(s"Silhouette with squared euclidean distance = $silhouette")
```

```
▸ (7) Spark Jobs
Silhouette with squared euclidean distance = 0.999999999947614
evaluator: org.apache.spark.ml.evaluation.ClusteringEvaluator = ClusteringEvaluat
Euclidean
silhouette: Double = 0.999999999947614
```

Fig. 11 Silhouette method value for k = 3

Table 10 External Clustering Validation: Comparison with previous studies

References	Real-time risk factors	Maternal risk factors	Medical risk factors
[49]		x	x
[47]			x
[50]		x	x
[51]			x
Experiment	X	x	x

Table 11 Comparison with a previous study

Reference	No of risk factors	Silhouette method	Elbow Method	Value of WSSSE
Old study	11	0.95	K = 2	978.24
Current study	**15**	**0.99**	K = 3	**1,01E + 23**

an eventual miscarriage. However, it is often too late to react and save baby's life. Comparing to previous studies, we attend good results. Table 10 and Table 11 present a comparison with other researches studies on miscarriage prediction.

Conclusion and future work

Nowadays, sensors and mobile phones remains important sources of data about human being behavior. Taking advantages of those technologies and benefit from reality mining and big data analytics present a challenge in predicting diseases. The present paper propose an e-monitoring system for real-time disease prediction that can be applied in several case studies. As a proof of concept, we propose a real-time miscarriage prediction system that benefit from the use of IoT (healthcare sensors, mobile phone and IoT tools), data mining algorithms and big data predictive analytics. The proposed system involves both pregnant women and doctors in the treatment of the disease through a mobile phone that we created.

K-means centroid clustering algorithm shows its performance by grouping data into three clusters: Miscarriage, Probable Miscarriage and Non-Miscarriage. To validate results, we used several methods including WSSSE, ICV, ECV and RCV. We achieve the lower value of WSSE when $k=3$ and Elbow method of RCV assert this value of K through its graph. Thus, the optimal number of K is 3 and we can no more regroup dataset to more partitions. In addition, the silhouette width value is 0.99, which is almost 1. So, compactness and separation of clusters are well structured in K-means model.

Future works consist on enhancing the proposed system by including more healthcare sensors to collect healthcare data about a person, adding risk factors collected from social networks, texts, calls, images … among others, developing the predictive model under a faster real-time framework like Flink. In addition, the proposed system represents a part of an ongoing project of Humanoid healthcare robots. In fact, future system include humanoids, the proposed prediction system and assisting system. HIYAM represents the first part created which is a new Moroccan humanoid robot that will serve as a nurse robot in hospitals. Several functions are developed and we are improving HIYAM in term of decision-making and prediction.

Abbreviations

AI	Artificial Intelligence
IoT	Internet of Things
RCV	Relative Clustering Validation
ICV	Internal Clustering Validation
ECV	External Clustering Validation
SVM	Support Vector Machine
NB	Naive Bayes
K-NN	K-Nearest Neighbor
C4.5	Decision Tree
WBC	Wisconsin Breast Cancer
BMI	Body Mass Index
MLlib	Machine Learning Library
RDD	Resilient Data Distributed
HCA	Hierarchical clustering
EM	Expectation Maximization
HR	Heart Rate
WSSSE	Within Set Sum of Squared Errors

Acknowledgements
Not applicable.

Author contributions
The author HA defined the study methodology, developed the e-monitoring miscarriage prediction system and wrote the main manuscript. Author ZJ brought his expertise in the workflow of the system. All authors reviewed, corrected the manuscript. All authors agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. Both authors read and approved the final manuscript.

Authors' information
Pr Hiba Asri: is an assistant professor at the computer science department of the Faculty of Sciences Semlalia at Cadi Ayyad University. She is member of LISI Laboratory at Cadi Ayyad University She holds a doctorate in computer science at Cadi Ayyad University, for her work on Big Data, Predictive / Preventive Analysis, Machine Learning and Reality Mining applied to the field of health. She obtained an engineering degree in Computer Networks and Information System in 2014. Her main research interests include Big Data, Machine Learning, IoT, Robotics, Data Mining algorithms, human–machine interaction and new generation internet technologies; in various fields of application such as health, education and e- learning. In addition to her academic experience, she chaired the program committee of many international conferences. She is a Keynote Speaker / Session Chair in various conferences and international workshops. She is Co-researcher in several international projects such as: Challenge AI-BioDiv Project and Climate Change Project. Hiba Asri is the creator of a bracelet Smart for real-time miscarriage prediction that is submitted for patent of invention. Also, she chairs the HIYAM project, A Moroccan Artificial Intelligence Humanoid Robot. **Pr Zahi Jarir**: received his postgraduate degree in computer science in 1997 on Natural Language Processing at Faculty of Sciences in Rabat, Morocco. From 1997 to 2006, he was assistant professor at Faculty of sciences, Cadi Ayyad University in Marrakech, Morocco. In 2006, he received academic accreditation from Cadi Ayyad University in the field of Personalization of Telecommunication Services and Web applications. Currently, he is a full professor of Computer Science at Faculty of Sciences of Cadi Ayyad University. He has participated actively in several research projects (RNTL, Volubilis, CSPT, PMARS, etc.). His research interests include distributed systems, adaptive and reflective middleware, ubiquitous computing, service-oriented computing, cloud computing, Information Security, M2M and IoT coordination, artificial Intelligence techniques and blockchain. He is a member of the editorial boards and member for several international journals, and a program committee member for multiple international conferences. He has published several publications in international conferences and journals, and chaired and organized several international scientific events.

Funding
Not applicable.

Availability of data and materials
The dataset is published and available through the link: http://dx.doi.org/10.17632/5sbmhh6t3r.1, under the Mendeley Repository name: HIBA ASRI_ Miscarriage Prediction Risk Factors.

Declarations

Ethical approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Received: 4 May 2022 Accepted: 23 February 2023
Published online: 14 March 2023

References

1. Asri H, Mousannif H, Al Moatassime H, Noel T. 'Big data in healthcare: Challenges and opportunities', in 2015 International Conference on Cloud Technologies and Applications (CloudTech). pp. 1–7. doi: https://doi.org/10.1109/CloudTech.2015.7337020. 2015.
2. Lazer D, Kennedy R, King G, Vespignani A. The parable of google Flu: traps in big data analysis. Science. 2014;343(6176):1203–5. https://doi.org/10.1126/science.1248506.
3. Tsuji K, et al. Book recommendation based on library loan records and bibliographic information. Procedia Soc Behav Sci. 2014;147:478–86. https://doi.org/10.1016/j.sbspro.2014.07.142.
4. 'Internet Research Methods - Claire Hewson, Carl Vogel, Dianna Laurent - Google Livres'. https://books.google.co.ma/books?hl=fr&lr=&id=w8mlCwAAQBAJ&oi=fnd&pg=PP1&dq=Internet+research+methods&ots=olZQy1Of_n&sig=0lTha5hwlU-t-BKtxYNokqyFso0&redir_esc=y#v=onepage&q=Internet%20research%20methods&f=false. Accessed Apr. 28, 2022.
5. Asri H. IoT and Reality Mining for Real-Time Disease Prediction. In: Azrour M, Irshad A, Chaganti R, editors. In IoT and Smart Devices for Sustainable EnviroCent. Cham: Springer; 2022.
6. Paul A, Ahmad A, Rathore MM, Jabbar S. Smartbuddy: defining human behaviors using big data analytics in social internet of things. IEEE Wirel Commun. 2016;23(5):68–74.
7. Asri H, Mousannif H, Al Moatassime H. Reality mining and predictive analytics for building smart applications. J Big Data. 2019;6:66. https://doi.org/10.1186/s40537-019-0227-y.

8. Asri H, Mousannif H, Moatassime HA. Big data analytics in healthcare: case study - miscarriage prediction. IJDST. 2019;10(4):45–58. https://doi.org/10.4018/IJDST.2019100104.
9. Asri H, Mousannif H, Al Moatassime H. Comprehensive miscarriage dataset for an early miscarriage prediction. Data Brief. 2018;19:240–3. https://doi.org/10.1016/j.dib.2018.05.012.
10. 'Comparing different supervised machine learning algorithms for disease prediction | SpringerLink'. https://link.springer.com/article/https://doi.org/10.1186/s12911-019-1004-8. Accessed Apr. 28, 2022.
11. 'Automated machine learning: Review of the state-of-the-art and opportunities for healthcare - ScienceDirect'. https://www.sciencedirect.com/science/article/pii/S0933365719310437. Accessed Apr. 28, 2022.
12. Harerimana G, Jang B, Kim JW, Park HK. Health big data analytics: a technology survey. IEEE Access. 2018;6:65661–78.
13. Bahri S, Zoghlami N, Abed M, Tavares JMR. Big data for healthcare: a survey. IEEE access. 2018;7:7397–408.
14. Poornima S, Pushpalatha M. A survey of predictive analytics using big data with data mining. Int J Bioinform Res Appl. 2018;14(3):269–82. https://doi.org/10.1504/IJBRA.2018.092697.
15. Razzak MI, Imran M, Xu G. Big data analytics for preventive medicine. Neural Comput Appl. 2020;32(9):4417–51.
16. Asri H, Mousannif H, Moatassime HA, Noel T. Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis. Procedia Computer Sci. 2016;83:1064–9. https://doi.org/10.1016/j.procs.2016.04.224.
17. Asri H, Mousannif H, Al Moatassim H. A Hybrid Data Mining Classifier for Breast Cancer Prediction. In: Ezziyyani M, editor. Advanced Intelligent Systems for Sustainable Development (AI2SD'2019). Cham: Springer; 2020.
18. Malik MM, Abdallah S, Ala'raj M. Data mining and predictive analytics applications for the delivery of healthcare services: a systematic literature review. Ann Oper Res. 2018;270:287–312. https://doi.org/10.1007/s10479-016-2393-z.
19. Yavaş G, Katsaros D, Ulusoy Ö, Manolopoulos Y. A data mining approach for location prediction in mobile environments. Data Knowl Eng. 2005;54(2):121–46. https://doi.org/10.1016/j.datak.2004.09.004.
20. McCue C. Data mining and predictive analytics in public safety and security. IT Professional. 2006;8(4):12–8. https://doi.org/10.1109/MITP.2006.84.
21. Pasquali R. Obesity, fat distribution and infertility. Maturitas. 2006;54(4):363–71. https://doi.org/10.1016/j.maturitas.2006.04.018.
22. Li Z, He Y-D, Chen Q et al. A risk-prediction nomogram for patients with second-trimester threatened miscarriage associated with adverse outcomes, 20 November 2020, PREPRINT (Version 1) available at Research Square. https://doi.org/10.21203/rs.3.rs-111117/v1.
23. Magnus MC, Wilcox AJ, Morken N-H, Weinberg CR, Håberg SE. Role of maternal age and pregnancy history in risk of miscarriage: prospective register based study'. BMJ. 2019;364:l869. https://doi.org/10.1136/bmj.l869.
24. Butt GQ, Sayed TA, Riaz R, Rizvi SS, Paul A. Secure Healthcare Record Sharing Mechanism with Blockchain. Appl Sci. 2022;12(5):2307.
25. Masethe HD, Masethe MA. Prediction of heart disease using classification algorithms. Proc World Congr Eng Comp Sci. 2014;2:25–9.
26. Paul A, Pinjari H, Hong WH, Seo HC, Rho S. Fog computing-based IoT for health monitoring system. J Sensors. 2018. https://doi.org/10.1155/2018/1386470.
27. Ganesan A, Paul A, Nagabushnam G, Gul MJJ. Human-in-the-Loop predictive analytics using statistical learning. J Healthcare Eng. 2021. https://doi.org/10.1155/2021/9955635.
28. Rathore MM, Ahmad A, Paul A, Wan J, Zhang D. Real-time medical emergency response system: exploiting IoT and big data for public health. J Med Syst. 2016;40(12):1–10.
29. Kong G, Chung T, Lai B, Lok I. Gender comparison of psychological reaction after miscarriage—a 1-year longitudinal study. BJOG: An IntJ Obstet Gynaecol. 2010;117(10):1211–9. https://doi.org/10.1111/j.1471-0528.2010.02653.x.
30. Hubail MA, et al. Couchbase analytics: NoETL for scalable NoSQL data analysis. Proc VLDB Endow. 2019;12(12):2275–86. https://doi.org/10.14778/3352063.3352143.
31. Badamasi YA. 'The working principle of an Arduino', in 2014 11th International Conference on Electronics, Computer and Computation (ICECCO). pp. 1–4. Doi: https://doi.org/10.1109/ICECCO.2014.6997578. 2014.
32. Upton E, Halfacree G. Raspberry Pi User Guide. Hoboken: Wiley; 2014.
33. L'Esteve RC. Machine Learning in Databricks. In: L'Esteve RC, editor. The Definitive Guide to Azure Data Engineering: Modern ELT, DevOps, and Analytics on the Azure Cloud Platform. Berkeley: Apress; 2021. p. 543–59.
34. Salloum S, Dautov R, Chen X, Peng PX, Huang JZ. Big data analytics on Apache Spark. Int J Data Sci Anal. 2016;1(3):145–64. https://doi.org/10.1007/s41060-016-0027-9.
35. Aziz K, Zaidouni D, Bellafkih M. 'Real-time data analysis using Spark and Hadoop', in 2018 4th International Conference on Optimization and Applications (ICOA). pp. 1–6. doi: https://doi.org/10.1109/ICOA.2018.8370593. 2018.
36. Meng X, et al. 'MLlib: Machine Learning in Apache Spark', p. 7.
37. Bonaccorso, G. Machine Learning Algorithms; Packt Publishing Ltd.: Birmingham, UK, 2017. https://scholar.google.com/scholar_lookup?title=Machine+Learning+Algorithms&author=Bonaccorso,+G.&publication_year=2017.
38. Sinaga KP, Yang M-S. Unsupervised K-means clustering algorithm. IEEE Access. 2020;8:80716–27. https://doi.org/10.1109/ACCESS.2020.2988796.
39. Fan J. OPE-HCA: an optimal probabilistic estimation approach for hierarchical clustering algorithm. Neural Comput Applic. 2019;31(7):2095–105. https://doi.org/10.1007/s00521-015-1998-5.
40. Hamidi M, Sheikhalishahi M, Martinelli F. 'Privacy Preserving Expectation Maximization (EM) Clustering Construction', in Distributed Computing and Artificial Intelligence, 15th International Conference, Cham. pp. 255–263. doi: https://doi.org/10.1007/978-3-319-94649-8_31. 2019.
41. Gilmore CJ, Barr G, Dong W. 'Choice of clustering method', urn:isbn:978–1–118–41628–0, 2019. https://onlinelibrary.wiley.com/iucr/itc/Ha/ch3o8v0001/sec3o8o3o5/. Accessed Apr. 29, 2022.
42. Thayer JF, Åhs F, Fredrikson M, Sollers JJ, Wager TD. 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. 2012;36(2):747–56. https://doi.org/10.1016/j.neubiorev.2011.11.009.
43. Anselem O, Floret D, Tsatsaris V, Goffinet F, Launay O. Influenza infection and pregnancy. Presse Med. 2013;42(11):1453–60. https://doi.org/10.1016/j.lpm.2013.01.064.

44. Wong EY, et al. Physical activity, physical exertion, and miscarriage risk in women textile workers in Shanghai China. Am J Ind Med. 2010;53(5):497–505. https://doi.org/10.1002/ajim.20812.
45. Nizard J, et al. Pathologies maternelles chroniques et pertes de grossesse. Recommandations françaises. J Gynecol Obstet Biol Reprod. 2014;43(10):865–82. https://doi.org/10.1016/j.jgyn.2014.09.017.
46. Veleva Z, et al. High and low BMI increase the risk of miscarriage after IVF/ICSI and FET. Hum Reprod. 2008;23(4):878–84. https://doi.org/10.1093/humrep/den017.
47. Stamatopoulos N, et al. Prediction of subsequent miscarriage risk in women who present with a viable pregnancy at the first early pregnancy scan. Aust N Z J Obstet Gynaecol. 2015;55(5):464–72. https://doi.org/10.1111/ajo.12395.
48. Chakrabarti S, Chakrabarti A. Food taboos in pregnancy and early lactation among women living in a rural area of West Bengal. J Family Med Prim Care. 2019;8(1):86–90. https://doi.org/10.4103/jfmpc.jfmpc_53_17.
49. Bottomley C, Bourne T. Diagnosing miscarriage. Best Pract Res Clin Obstet Gynaecol. 2009;23(4):463–77. https://doi.org/10.1016/j.bpobgyn.2009.02.004.
50. Mastrodima S, Akolekar R, Yerlikaya G, Tzelepis T, Nicolaides KH. Prediction of stillbirth from biochemical and biophysical markers at 11–13 weeks. Ultrasound Obstet Gynecol. 2016;48(5):613–7. https://doi.org/10.1002/uog.17289.
51. Kumari S, Roychowdhury J, Biswas S. Prediction of early pregnancy failure by use of first trimester ultrasound screening. Int J Reprod Contracept Obstet Gynecol. 2016;5(7):2135–41.
52. Asri H. HIBA ASRI_ Miscarriage Prediction Risk Factors. 2021. https://doi.org/10.17632/5sbmhh6t3r.1.

Publisher's Note