Noname manuscript No.
(will be inserted by the editor)

Machine learning as a tool to study the influence of chronodisruption in preterm births

Elena Díaz · Catalina Fernández-Plaza · Inés Abad · Ana Alonso · Celestino González · Irene Díaz

Received: date / Accepted: date

Abstract It is well known that there are some maternal and fetal issues that directly influence preterm births. However, all the variables provoking it are not completely determined. On the other hand, chronodisruption alters maternal circadian rhythms, with negative consequences for the maturation of the fetus. Thus, the objective of this work is to add other factors related to maternal chronodisruption factors and to check if all together can improve preterm birth prevention. The methodology followed to reach this objective is based on machine learning approach. The data are composed by a cohort of 380 births labelled as preterm or term births. Variables defining each individual are related to maternal habits, night exposure to light or sleep duration during gestation. In addition, maternal variables related to the gestation were obtained as well as fetal characteristics. Preliminary statistical tests confirm that cervix dilatation, fetus estimated weight and weight at birth were significantly lower ($p < 0.05$) in preterm group than in term group as expected. A deeper study based on machine learning highlights some interesting and non obvious relations between some factors related to night exposure to light and sleeping habits. In fact, the decision tree obtained as predictive model

I. Díaz would like to thank for the support of Spanish Ministry of Science and Technology project TIN-2017-87600-P and FICYT Project IDI/2018/000176. A. Alonso, E. Díaz and C. Fernández-Plaza would like to thank for the support of University of Oviedo project PAPI-17-PUENT-06.

E. Díaz, I. Abad, A. Alonso, C. González
Area of Physiology, Department of Functional Biology, University of Oviedo, Oviedo, Spain

C. Fernández-Plaza
Department of Gynecology and Obstetrics, Central University Hospital of Asturias (HUCA), Oviedo, Spain

I. Díaz
Area of Computing and Artificial Intelligence, University of Oviedo, Oviedo, Spain
E-mail: sirene@uniovi.es

indicates that light coming in through the window or lightness level of the bedroom during the night are key features in predicting preterm delivery.

Keywords chronodisruption · light at night · sleep habits · machine learning · predictive model · preterm birth

1 Introduction

Preterm birth is defined as the birth of a baby at fewer than 37th week of gestation, more exactly, before 259 days counted from the last menstruation. It is therefore accepted that prematurity is related to the time of pregnancy. Although there is no regular view for all authors, it is considered that a fetus is viable from the 22 week of gestation (Lumley, 2003). A threat of preterm is defined as the beginning of regular uterine contractions, dilation and cervical effacement between 23 and 36 gestation weeks. In the United States, the preterm birth rate is 12-13% whereas in Europe and other developed countries it is close to 5-9% (Goldenberg et al., 2008). Although the factors that determine the beginning of the birth are not yet known exactly, it is often postulated that the inflammatory response is of vital importance in the triggering of both term and preterm delivery. Other factors are uterine infection, premature rupture of membranes (RPM), uterine distention, or insufficient maternal-fetal immune recognition. There are also some risk factors associated with prematurity such as maternal age, obesity, anaemia, folic acid ingestion, gestational diabetes, tobacco, alcohol, pollution, stress or multiple gestation (Llurna et al., 2015), whose control or disappearance could reduce the prevalence of preterm births.

It is well known that as consequence of night exposure at artificial light, the prevalence of disturbances in circadian rhythms and disturbed sleep are increasing in modern society (Pallesen et al., 2014). Chronodisruption is defined as the relevant alteration of the internal temporal order of physiological and behavioral circadian rhythms (Garaulet and Ordovás, 2013). There are many works focused on studying issues related to this topic. For example, De Arriba-Pérez et al. (2018) uses off the shelf wrist wearables to estimate sleep quality, sleepiness level, chronotype and sleep regularity indicators. In relation to premature delivery, the intensity of light during the night causes an endogenous suppression of the melatonin levels produced (Reiter et al., 2014). In this way, chronodisruption alters maternal circadian rhythms, with negative consequences for the maturation of the fetus, which can lead to psychological and behavioral problems in the newborn (Ferreira et al., 2012). Thus, most of these factors could be used as a prognostic of premature delivery.

On the other hand, machine learning (ML) is nowadays an efficient tool in decision support system design for many broad research areas (Montañés et al., 2009; Gil-Pita et al., 2015) and in particular in Biomedicine (Fernández-Navarro et al., 2019). Preterm birth prediction and related issues are a long-standing problem where ML has been frequently applied. In (Rawashdeh et al.,

2020) it is developed a model that acts as a decision support system for pregnant women at high risk of delivering prematurely before having cervical cerclage. They used data from 274 pregnancies managed with cervical cerclage and build different ML models using decision trees, random forest, K-nearest neighbors and neural networks. Weber *et al.* in (Weber et al., 2018) predicts early spontaneous preterm birth among non-Hispanic black and white women by applying ML to multilevel data from a large birth cohort. Gao *et al.* in (Gao et al., 2019) model extreme preterm birth using deep learning models that consider temporal relations documented in electronic health records. Fergus *et al.* (Fergus et al., 2016) uses electrohysterography signals and different artificial neural networks to identify relevant features for predicting preterm births. In (Rezaeian et al., 2020) neural networks and logistic regression are presented as a tool for prediction of mortality in premature neonates upon admission to neonatal intensive care units. Related to this topic, Li et al. (2018) identifies the risk factors associated to small-for-gestational-age condition using machine learning and feature selection techniques.

The relation between pollutants and some chronobiology markers with regard to preterm birth has also attracted attention. Regarding pollutants, In (Ren et al., 2018) some maternal exposure to some ambient air pollutants are identified as the primary risk factor for congenital heart defects. In addition they show that a ML model has better predictive performance than traditional logistic regression models. In the same line, it is suggested that maternal exposure to outdoor gaseous air pollutants increases risk of preterm birth in (Guo et al., 2019). (Mustieles et al., 2020) states that maternal preconception urinary Bisphenol A and Bisphenol S concentrations, as well as paternal preconception urinary parabens concentrations are associated with a higher risk of preterm birth.

Regarding chronobiology markers, (Loy et al., 2020) examined the associations of maternal night-time eating and sleep duration during pregnancy with gestation length and preterm birth. In addition, sleep habits are also considered in (Facco et al., 2019) where they concluded that self-reported late sleep midpoint in both early and late pregnancy, but not short sleep duration, is associated with an increased rate of preterm birth. Other factors as temperature are studied in (Sun et al., 2019), concluding that days of extreme heat, but not extreme cold, are associated with higher risk of preterm birth.

However, the factors affecting preterm births are not fully described yet but the rate of preterm births is increasing and thus it causes health, development and economic problems.

The aim of this study is to design a predictive model to forecast the risk of premature delivery based on several potential risk factors. The effect of some of these risk factors (age, or diabetes for example) on preterm births is already known. However, the impact of other risk factors related to chronodisruption is rather studied. Thus, the model will help in preterm birth prevention considering factors related to night exposure to light.

2 Methodology

2.1 Study population

An observational, retrospective and descriptive study was performed with 380 births that occurred between January 1st, 2015 and May 10th, 2016 in the Universitary Central Hospital of Asturias (HUCA) (Spain). The study was approved by Research Ethics Committee of the Principado of Asturias (Spain). Those births in which the baby was born dead or the term was induced were eliminated from the study. The control group births were selected by choosing a term delivery occurred on the same day, just before each preterm delivery. 380 births were collected, 157 were preterm (between 24.4 and 36.9 weeks of pregnancy) and 223 were term (between 37 and 42.3 weeks).

2.1.1 Measurement

Issues related to maternal, fetal and maternal sleep habits are considered to define each patient. The main maternal characteristics are Age (years), Body mass index (BMI) (Kg/m2), Obstetric history (Primiparity and primigravity), Weight at the beginning of pregnancy (Kg), Weight gain thought pregnancy (Kg), Multiple gestation (Yes or not), Toxic habits (Tobacco or alcohol), Group B Streptococcus agalactiae carriage (Yes or not). The considered fetal variables are Sex (Male or female), Nuchal translucency analysis in the first trimester (NT) (mm), Estimated weight in the third trimester (g), Weight at birth (g), Folic acid administration (Yes or not), Test O'Sullivan (Normal or altered), Hemoglobin concentration at first trimester of pregnancy (g/dL), Cervix dilatation (cm) , Premature rupture of membranes (PRM) (Yes or not).

With regard to maternal sleep habits a questionnaire (Hersh et al., 2015) was used. In addition, the sleep habits related to night exposure to light is coded according to the numerical scale established by (Davis et al., 2001). The answers to both questionnaires were obtained by telephone survey. 387 patients answered to survey, 223 of which experienced preterm deliveries whereas the other 157 were term. The questionnaire items are the following:

- Time at which she turned off the light to go to sleep.
- Time at which she woke up on each day of a usual week.
- Number of hours usually spent looking at an electronic device after lights are turned off.
- Whether a light or television was turned on in (or near) the bedroom while sleeping.
- Average number of times per night (if any) that their sleep was interrupted.
- If sleep was interrupted, whether she turned a light on, and intensity and duration of light exposure.
- Average sleep duration (in hours) for weekdays and weekend nights, separately.
- Level of light in the bedroom during sleep on working days and weekends.

Based on these data the variables detailed in Table 1 are considered.

VARIABLE	MEANING
DeviceMF	Hours of usage of electronic devices before sleeping with room lights off on working days (Monday to Friday)
DeviceSS	Hours of usage of electronic devices before sleeping with room lights off on wekkends
Interruption	Awakenings at night (with light turned on)
Midnight light	If any light or television is on close to midnight
Sleeping hoursWD	Sleeping hours (in average) during working days
Sleeping hoursWE	Sleeping hours (in average) during weekend days
Lightness	Level of room lightness at night. Observed levels: LEVEL 1: Room completely dark LEVEL 2: You can see your hands LEVEL 3: You can see to the end of the bed LEVEL 4: You can see through the room
BedtimeWD	Bedtime on working days. 3 categories: BEFORE 23: She goes to bed before 11 at night. 23-00: She goes to bed between 11 and 12 at night. AFTER 00: She goes to bed after midnight
BedtimeWE	Bedtime on weekends. 3 categories: BEFORE 23: She goes to bed before 11 at night 23-00: She goes to bed between 11 and 12 at night AFTER 00: He goes to bed after midnight.
Occupation	Occupation during pregnancy: No: Does not work during pregnancy No shift work: Work with a non-shift schedule Shift work: Work with a shift schedule
Change of habits	Number of hours, depth of sleep or lighting of the room

Table 1 Variables related to sleep habits

2.1.2 Statistical Analysis of variables

First of all a statistical analysis was carried out to study the behaviour of both qualitative and quantitative variables depending on preterm or term delivery. Quantitative variables are: age, BMI, initial weight, weight gain, hemoglobin, cervix dilatation, nuchal translucency, estimated weight, weight at birth. Qualitative variables are primigravity, primiparity, multiparity, tobacco, alcohol, group B streptococcus, folic acid, O'Sullivan test, PRM, fetal sex. Variables related to sleep habit are all qualitative. Note that there are some variables that are not useful for predicting preterm births. For example, although there are statistical differences between preterm and term groups in *weight at birth* variable, it is obviously excluded as it is a data we can only obtain after delivery. Table 2 and Table 3 lists the basic statistics for maternal and fetal variables while Table 4 and Table 5 show the same information for variables related to maternal habits. Note that for these last set of variables, the $76,3\%$ of mothers associated to preterm delivery have answered to questionaries while in case of term deliveries this percentage decreases to $67,2\%$. Chi-square independence test was performed for these qualitative variables while quantitative variables were analysed with Student's t-test, previously checking the normality of the variables using the Kolmogorov-Smirnov test ($p < 0.05$ was considered significant for both tests). R-studio was used to perform this analysis. In case of qualitative variables, it is not possible to reject the null hypothesis that each qualitative variable in Table 4 and Table 5 is independent of preterm birth variable. On the other hand, variables with an asterisk in Table 2 and Table 3 show significant differences between preterm and term classes. However, note that all of them are related to delivery and they are not useful to predict preterm births in advance. Thus, they are not considered in this study. Note that multiparity is strongly correlated to preterm births and consequently is also removed from this study.

	Preterm (n=157)		
	Minimun	Mean	Maximun
Age (years)	15	$32,5 \pm 0,5$	46
BMI (Kg/m2)	19,3	$29,1 \pm 1,3$	203,7
Initial weight (g)	42	$62,2 \pm 1{,}4$	100
Weight gain(g)	-0,9	$3,5 \pm 0,8$	13
Hemoglobin (g/dL)	7,7	$12,8 \pm 0,1$	15,2
Cervix dilatation (cm)*	0	$2,5 \pm 0,2$	9
Nuchal Translucency (mm)	0,8	$1,5 \pm 0,04$	2,5
Estimated Weight (g)*	656	$1914 \pm 48,3$	3535
Weight at birth (g)*	600	$2192,1 \pm 52$	3600
	Term (n=223)		
	Minimun	Mean	Maximun
Age (years)	17	$32,8 \pm 0,4$	43
BMI (Kg/m2)	18,9	$28,9 \pm 0,4$	68,4
Initial weight (g)	45,5	$65,2 \pm 0,9$	114
Weight gain(g)	-3	$3,8 \pm 0,6$	26
Hemoglobin (g/dL)	9,8	$12,8 \pm 0,1$	14,9
Cervix dilatation (cm)*	0	$3,3 \pm 0,1$	10
Nuchal Translucency (mm)	0,7	$1,6 \pm 0,04$	6,7
Estimated Weight (g)*	1887	$3042,2 \pm 38$	4410
Weight at birth (g)*	1040	$2655,7 \pm 40,1$	4325

Table 2 Basic statistics for continuous variables related to maternal and fetal issues

2.2 Machine learning

As it was introduced before, the goal of this work is to study if night exposure to light and habits related to sleep can influence in preterm delivery. The approach followed in this work is based on ML which has been proved to be successful solving biomedicine problems as it was pointed out in the introduction section. In this section we describe the particular approach we follow that is sketched in Figure 1 and detailed in the following subsections.

2.2.1 Preprocessing

As it was described in section 2.1.1, the patient set is characterized by both continuous and categorical variables. It is well known that some ML methods

	Preterm (n=157)	
	%YES	%NO
Primigravity	44,6	54,8
Primiparity	72,6	26,8
Multiparity*	100	0
Tobacco	16,6	76,4
Alcohol	2,5	90,4
Group B streptococcus	9,6	38,2
Folic acid	61,8	0,6
O'Sullivan test	39,5	7,6
PRM*	55,4	44,6
	%MALE	%FEMALE
Fetal sex	45.8	54.2
	Term (n=223)	
	%YES	%NO
Primigravity	45,3	53,8
Primiparity	66,8	32,3
Multiparity*	0,9	99,1
Tobacco	12,1	83,9
Alcohol	1,3	94,2
Group B streptococcus	12,1	57,4
Folic acid	75,3	0,5
O'Sullivan test	42,2	6,7
PRM*	20,2	51,6
	%MALE	%FEMALE
Fetal sex	52.1	42.9

Table 3 Basic statistics for categorical variables related to maternal and fetal issues

are quite sensitive to variable scale. Thus, continuous variables are normalized to avoid this circumstance. In addition, missing values are treated using K-nearest neighbor imputation (Zhang, 2012).

In addition to the statistical analysis, predictive factors for preterm births are obtained via ML methods and noise reduction techniques using R package version 3.4.3. In particular, caret package version 6.0.78 has been used to build classifiers and NoiseFiltersR package (version 0.1.0) was used for noise reduction.

2.2.2 Noise elimination

ML methods strongly depend on the input data. Thus, it is extremely important to study if the input data are sound. To that extent, the following techniques to identify noisy examples are employed:

- RobustFilter (Verbaeten, 2002) is a noise filter that builds a C4.5 decision tree from the training data and then removes those instances misclassfied by this tree. The process is repeated until no instances are removed.
- IteratedVoting method (Verbaeten, 2002) splits the dataset into n folds and then builds and tests a C4.5 tree on every combination of $n-1$ folds. Thus, $n-1$ votes are gathered for each instance. Removal is carried out

Preterm (n=157)				
	%YES		%NO	
Midnight light	50.3		26	
Change of habits	27.5		48.8	
	%< 1	%> 1	%NO	
DeviceMF	33.3	6	37	
DeviceSS	30.7	7.8	37.8	
	%with light	%without light		%NO
Interruption	35.5	20.4		20.4
	%< 6	%$[6, 9]$		%> 9
Sleeping hours-WD	20.4	39.4		16.5
Sleeping hours-WE	11.8	40.2		8.3
Lightness	%LEVEL 1	%LEVEL 2	%LEVEL 3	%LEVEL 4
	41.7	11.8	11	11.8
	%BEFORE 23	%23-00	%AFTER 00	
BedtimeMF	20.5	39.3	16.5	
BedtimeWE	11.8	40.3	24.4	
	%NO	%No shift work	%Shift work	
Occupation	22.4	42.4	11.5	

Table 4 Basic statistics for variables related to maternal habits (for preterm births)

Term (n=223)				
	%YES		%NO	
Midnight light	15.5		51.7	
Change of habits	36.9		30.3	
	%< 1	%> 1	%YES	
DeviceMF	31.5	9.5	26.2	
DeviceSS	31.5	9.5	26.2	
	%with light	%without light		%NO
Interruption	25	23.8		18.4
	%< 6	%$[6, 9]$		%> 9
Sleeping hoursWD	13.1	39.3		14.8
Sleeping hoursWE	10.1	30.3		26.8
Lightness	%LEVEL 1	%LEVEL 2	%LEVEL 3	% LEVEL 4
	45.2	8.3	5.4	8.3
	%BEFORE 23	%23-00	%AFTER 00	
BedtimeMF	20.2	36.3	10.7	
BedtimeWE	10.7	30.9	25.6	
	%NO	%No shift work	%Shift work	
Occupation	18.4	40.5	8.3	

Table 5 Basic statistics for variables related to maternal habits (for term births)

by majority or consensus voting schemes. The process is repeated until no more noisy instances are removed.

- EdgeBoostfilter (Wheway, 2001) combines an AdaBoost scheme with a default C4.5 tree as weak classifier. After m iterations, those instances with edge values are considered noisy and thus removed.
- HARF (Sluban et al., 2010) is a classifier-based filter that uses a Random-Forest classifier. It considers the rate of disagreement in the predictions made by the individual trees in the forest to detect the noisy examples: if

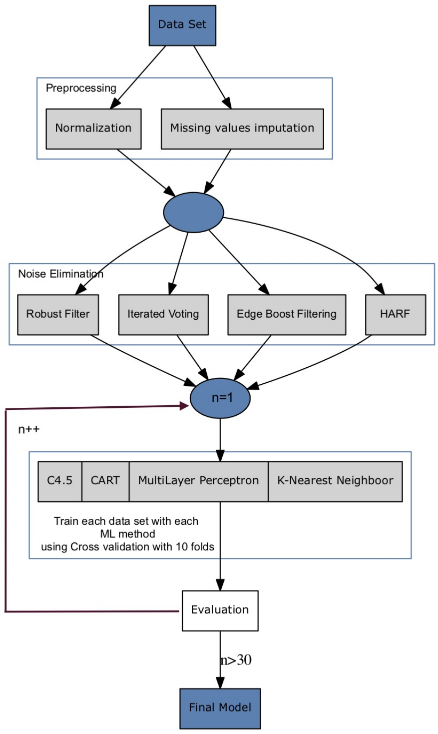

Fig. 1 Machine learning procedure

this rate is high, the example is probably noisy; otherwise, it is considered to be clean.

Each one of these noise elimination methods leads to a different data set. Thus from the initial dataset we obtain for different ones, each one of them will be the input to the different classifiers.

2.2.3 Classifiers

A classification system is a systematic approach to building classification models from an input data set. Examples include decision tree classifiers, neural networks or instance based learners among other classifiers. Each technique employs a learning algorithm to identify the model that best fits the relationship between the attribute set and class label of the input data. The model generated by a learning algorithm should both fit the input data well and correctly predict the class labels of examples it has never seen before. Therefore, a key objective of any learning algorithm is to build models with good generalization capability, which is equivalent to look for models that accurately predict the class labels of previously unknown examples. Thus, the classification procedure is a cornerstone in any predictive problem. In addition, there is no a standard classification method so far. Thus, several different methods were tested to select the one performing the best for this task, taking into account the trade-off between performance and interpretability. The methods consider in this work are described below.

– **Tree based methods**: Decision trees are one of the more extensively applied classifiers in many different domains. A decision tree (generally defined) is a tree whose internal nodes are tests on the variables that define the inputs and whose leaf nodes are categories. A decision tree has three different kind of nodes: A root node that has no incoming edges and zero or more outgoing edges, internal nodes with exactly one incoming edge and two or more outgoing edges and leafs or terminal nodes with exactly one incoming edge and no outgoing edges. In a decision tree, each leaf node is assigned a class label. The nonterminal nodes contain attribute test conditions to separate examples that have different characteristics. Each path from the root of the tree to a leaf determines a region, that is, a more homogeneous group subset of the input data. Initially, the whole training set is associated with a leaf. These classifiers usually adopt a greedy approach in which decision trees are constructed in a top-down recursive divide-and-conquer manner. Most algorithms for decision tree induction also follow a top-down approach, which starts with a training set of tuples and their associated class labels. The training set is recursively partitioned into smaller subsets as the tree is being built. Applying a test in a recursive procedure, it is decided if the set associated to a leaf is split into smaller subsets associated to new leaves. When a subset is homogeneous (in some sense) the procedure halts and the node is labelled as a leaf (terminal node). Note that these greedy strategy makes decision tree grow by making a series of locally optimum about which attribute to use for partitioning the data. Every construction method of a tree revolves around the selection of the splits, the decisions when to declare a node terminal or to continue splitting it and the assignment of each terminal node to a class.
  Different combinations of splitting selection measures, stopping criteria and pruning techniques and class assignments to a leaf lead to different methods. C5.0, C4.5 (Kuhn et al., 2018; Quinlan, 1994, 2000) and CART (Breiman, 2001) are examples of this kind of classifiers. CART uses Gini index as criteria to select the splits while C4.5 and C5.0 use Gain ratio. CART uses cost complexity pruning to evaluate whether the tree should be simplified while while C4.5 employs pessimistic pruning. C5.0 is an extension of C4.5 that is faster and in general more accurate than C4.5 and has several basic improvements that are likely to generate smaller trees. Although in recent years several new ML methods have been developed, decision tree methods are still useful when the problem has a lot of nominal data (in addition to numeric data), unknown attributes or when the dataset is not very large and with a lot of attributes. Unlike neural networks or lazy learners, decision trees are human readable and easy to comprehend (Kuhn and Johnson, 2013).

Some other methods are also considered in this work due mainly to their performance. On of them is the random forest method. It is supposed that the prediction provided by a set of trees combined in some way improves predictive performance over a single tree by reducing variance of the prediction. Random forest is based on building trees using a random subset of the top k predictors at each split in the tree. Each tree in the ensemble is then used to generate a prediction for a new sample and these m predictions are averaged to give the forests prediction.

– **Lazy learners**: Instance based learner such as k Nearest Neighbors (k-NN) (Peterson, 2009) are characterized for being one of the simplest ML methods and for their reduced training time. k-NN classifiers are based on learning by comparing a given test example with each example of the training set. Each example represents a point in an n-dimensional space. Then, to classify an example, a k-NN classifier searches the pattern space for the k training tuples that are closest to the example to classify. These k training examples are the k nearest neighbors of the example to classify. Closeness is defined in terms of a distance metric, such as Euclidean distance. k-NN is quite sensitive to variable scale as it is based on distances, does not work with missing values. On the other hand, the parameter k is not known in advance and it is necessary to fix it experimentally. k-NN can be useful when interpretability is not a requirement to model the solution to a predictive problem.
– **Neural Networks**: Artificial Neural Networks (NN) (Cook, 2020) are inspired in biological neural networks and have been applied in a wide variety of classification and regression problems. In particular, multilayer perceptrons (MLP) are loop-free networks whose neurons are arranged in layers, with each neuron providing input only to neurons in the next layer of the sequence. The first layer contains input neurons and thus the number of neurons in this layer is determined by the dimension of the examples in the training set. The last layer contains the output neurons. The number of hidden layers is not determined beforehand and it is usually experimentally determined. MLP training involves adjusting the weights (the connections between two neurons each of them lying in consecutive layers) of the model in order to minimize the prediction error. This weight adjusting is iteratively performed in two steps: First, the input example moves from the input layer through the hidden layers to the output layer. Thus the output layer is measured against the ground truth labels. Then the weights are updated using back-propagation. This process is then repeated until convergence. The learning of this kind of methods are the weight values. Thus, the interpretability of these models is low although they are accurate in general.

2.2.4 Training

Training a ML method is as complex as necessary to avoid overfitting and to correctly optimize the different hyperparameters associated to each method.

In this case we have applied repeated cross-validation with 10 folds. The repetition parameter was 30.

2.2.5 Evaluation

To evaluate the performance of the presented method several metrics are selected from the ***confusion matrix*** shown in Table 6. The confusion matrix is a contingency table that shows how the objects of a binary classification problem are classified according to their real value and the one predicted by the classification method.

		Predicted	
		Preterm	Term
Real	Preterm	True Positive (TP)	False Negative (FN)
	Term	False Positive (FP)	True Negative (TN)

Table 6 Confusion matrix.

From the afore defined confusion matrix the $F_1 = \frac{2*TP}{2*TP+FP+FN}$, *Sensitivity*= $\frac{TP}{TP+FN}$ and *Specificity*= $\frac{TN}{TN+FP}$ are computed.

3 Results

We consider the original data set plus other 4 different data sets obtained from the original one, after applying the noise removal techniques described in subsection 2.2.2. Thus, 24 different learning procedures were tested as result of combining 6 different classifiers (see subsection 2.2.3) with 4 different noise reduction techniques. Thus, it is necessary to select the best combination of learner plus noise filter to construct the predictive model. In addition, each method is optimized during the training step according to its particular parameter setting.

Table 7 shows the performance in test of each classification method plus noise reduction combination in terms of Sensitivity, Specificity, Precision and F1. As it can be seen, Robust noise reduction technique together with C5.0 classification method is the combination obtaining the best performance in terms of F, $Sensitivity$ and $Specificity$. Thus, this combination is the one selected to build the model for predicting pre-term births.

T-tests were applied to validate the statistical significance for each comparison between C5.0 and the other ML methods when noise filtering methods is fixed with p value 0.05 (see Table 8). Data in table 8 show that when Robust noise filtering method is applied there is no significant differences between the values of F, Sensitivity and Specificity obtained by C5.0 and J48. On the other hand, the values of F, Sensitivity and Specificity obtained by C5.0 are

Robust			
	F	Sensitivity	Specificity
C5.0	0.85	0.86	0.78
J48	0.83	0.85	0.73
rpart	0.82	0.80	0.80
MLP	0.78	0.76	0.75
k-NN	0.69	0.76	0.45
rf	0.83	0.82	0.81
IterativeVoting			
	F	Sensitivity	Specificity
C5.0	0.74	0.73	0.76
J48	0.79	0.80	0.76
rpart	0.78	0.77	0.78
MLP	0.78	0.77	0.79
k-NN	0.64	0.69	0.51
rf	0.76	0.73	0.81
edgeBoost			
	F	Sensitivity	Specificity
C5.0	0.53	0.53	0.53
J48	0.57	0.62	0.49
rpart	0.57	0.61	0.50
MLP	0.59	0.61	0.55
k-NN	0.56	0.60	0.46
rf	0.56	0.57	0.55
HARF			
	F	Sensitivity	Specificity
C5.0	0.64	0.65	0.56
J48	0.66	0.66	0.60
rpart	0.65	0.72	0.47
MLP	0.69	0.68	0.67
k-NN	0.60	0.64	0.45
rf	0.66	0.68	0.57

Table 7 Performance (in average) of the different combinations of machine learners and noise reducers in terms of F, Sensitivity and Specificity.

statistically better than those obtained by Rpart and MLP. In addition, the values of F, Sensitivity obtained by C5.0 are statistically better than those k-NN and Specificity and Sensitivity for C5.0 are also statistically higher than the obtained by rf.

When Iterative Voting method is applied as noise filter, we find significant differences only for Specificity in case of comparing C5.0 and J48 performances and for F, Sensitivity and Specificity when C5.0 and k-NN are compared. If EdgeBoost is considered as noise filter, we found almost no significant differences among the different ML methods (only between F-measure obtained by C5.0 and k-NN). In case of using HARF noise filter, only the F measure obtained by C5.0 is significantly higher than that obtained by rpart and k-NN and the Sensitivity obtained by C5.0 also outperforms the one obtained by k-NN. Thus, it seems clear that the combination Robust filtering together with C5.0 as ML method is the best. The underlying model is depicted in Figure 4. From this model the most important variables are also extracted (see Figure 3).

	Robust	Iterative Voting	EdgeBoost	HARF
C5.0-J48	-	Specificity	-	-
C5.0-rpart	F	-	-	
	Specificity	-	-	F
	Sensitivity	-	-	
C5.0-MLP	F	-	-	-
	Specificity	-	-	-
	Sensitivity	-	-	-
C5.0-k-NN	F	F	F	F
		Specificity	F	
	Sensitivity	Sensitivity	F	Sensitivity
C5.0-rf	Specificity	-	-	-
	Sensitivity	-	-	-

Table 8 Statistical differences among different ML methods when the noise filter is fixed

We have also compared the results obtained by C5.0 across the different noise reduction methods, obtaining in this case statistical significance between the performance obtained by C5.0 together with Robust filtering and the obtained between C5.0 and one of the other three noise filters.

Figure 2 deepens in the performance of the selected combination across the 30 Cross-validation iterations.

4 Discussion

Our results from statistical analysis allow us to confirm that several well known obstetrical factors as cervix dilatation, multiple gestation or estimated birth weight or premature rupture of membranes increase preterm risk as previously other authors pointed out (Goldenberg et al., 2008; Denbow and Lyon, 2005).

Preterm birth is considered a public health problem but the etiology of prematurity is not completely understood due to complex interactions of different factors as genetics, environmental or mother‘s own causes. Previous authors have been identified various environmental exposures as potential risk factors for preterm birth (Stieb et al., 2012; Ferguson et al., 2013). In addition Giorgis-Allemand *et al.* (Giorgis-Allemand et al., 2017) have proposed that meteorological conditions and environmental pollutants may be considered as risk factors. Artificial light at night is also considered a form of pollution which involves disturbances with circadian rhythms affecting many physiological parameters such as sleep-wake cycle or temperature rhythm (Touitou et al., 2017). In addition, night exposure to artificial light increases the risk of miscarriage. In this way shift work has been associated with risk of preterm birth (Zhu et al., 2004). Given that in today’s society future mothers, like all of us, are exposed to the deleterious effect of artificial light at night, it seems interesting to be able to predict the risk of preterm delivery based on light and other classic risk factors. So we have designed a model to predict preterm birth risks that takes into account for the first time the current lifestyle related to sleep habits and night exposure to light. In fact, Figure 3 shows the

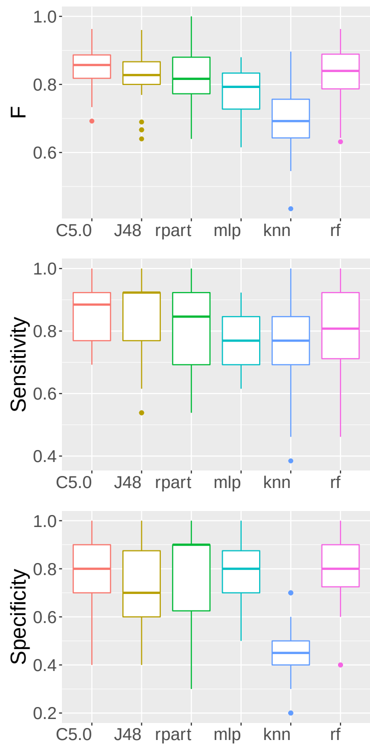

Fig. 2 Boxplot of F, Sensitivity and Specificity of C5.0 plus Robust filtering method

importance of each variable in this process. It shows the overall importance (ranging from 0 to 100) of the factors considered in this work. Note that there are some traditional well known factors as Mother's age, NT or BMI. However, the most important variables are those related to maternal light exposure.

The resulting predictive model can be seen in Figure 4. As it can be checked, the model includes previously confirmed risk factors (BMI) but also those related to maternal habits linked to hours of sleep and night exposure to light. In addition, it is shown for the first time the great importance of the artificial light pollutant as a critical risk factor for preterm labor. In fact, our model shows that other classical preterm risk factors like BMI become less important as predictive risk factors. The model reveals that the main risk factor for preterm labor is the maternal exposure to artificial light close to midnight. Furthermore according to our model other maternal habits like the use of electronics devices before sleeping is also an important risk factor even if they are not are exposed to light at midnight. This factor, together with high levels of lightness at night influences preterm births.

Shift work is also a critical factor to induce preterm labor in those pregnant women usually exposed to artificial light during midnight. In view of our results

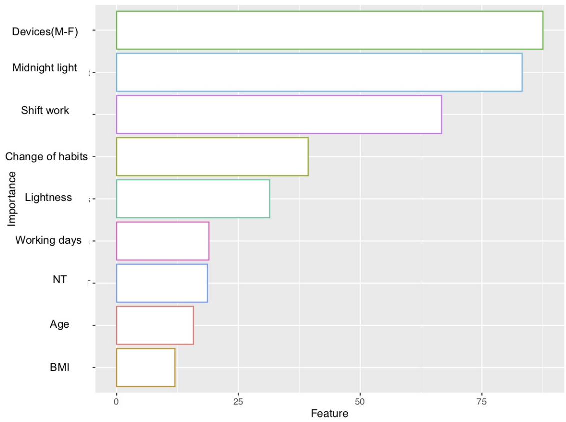

Fig. 3 Mean importance of each variable during the training process

maternal exposure to artificial light is an important risk factor to cause preterm labor, these habits are clearly avoidable factors. Our predictive model may be therefore very useful in the obstetrics clinic to prevent preterm birth.

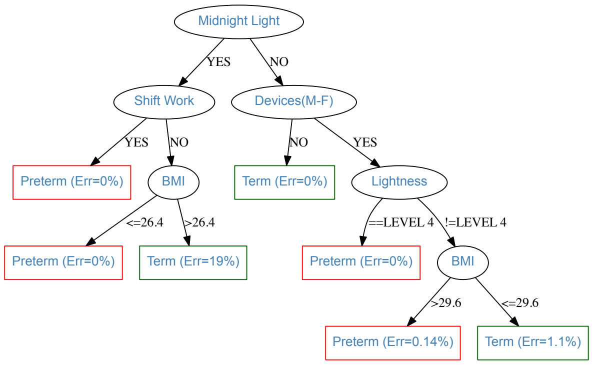

Fig. 4 Model for predicting preterm births

There are an increasing number of scientific articles that suggests the kind of work (shift work, work in fixed night shifts, longer hours) as a factor that increases the risk of adverse pregnant outcomes (Croteau et al., 2007),(Davari et al., 2018),(Cai et al., 2019). In this way, in (Nehme et al., 2019) found that pregnant women working at night exposed to continuous light, decreased nocturnal melatonin levels and increased the risk of bleeding and miscarriage.

The results of this study suggest that we should change our lifestyle mainly during gestation. In present society we are overexposed to artificial light at night specially during weekends, generating social jet lag (Loy et al., 2020).

In addition we are daily exposed to electronic devices like computers, mobiles or ebooks, all of them emitting intense light. It is known that light exposure during night results in chronodisruption, thus altering the main biological rhythms among them the hormone melatonin circadian rhythm (Reiter et al., 2014). In this way, it has been observed that melatonin concentration increases during pregnancy (Nakamura et al., 2001); and it seems to be involved in gestation normal evolution, since it is of vital importance for the placenta development and function, and for the induction of the rhythmicity of the fetal organs (Voiculescu et al., 2014). So, the suppression of this hormone during pregnancy affects the maturation process in the fetus (Ferreira et al., 2012), such as retardation in growth (Nakamura et al., 2001). It is possible to argue that chronodisruption caused by night exposure to light causes uterine growth retardation, and finally premature delivery. Note that the influence of the use of electronic devices, or the abuse of artificial light until late at night on pregnant has not been previously studied in depth. Our results are therefore pioneering in this area.

ML has been extensively applied in biomedicine as an efficient tool to help in decision making processes. In particular, it has been previously used for studying risk factors associated to delivery (Gao et al., 2019; Rawashdeh et al., 2020; Weber et al., 2018). In this work we have tested different ML approaches, showing decision trees as an accurate tool to predict preterm delivery from variables based on maternal habits, especially related to night exposure to light.

In conclusion, we have developed a model to predict the risk of preterm birth in pregnant women considering classical factors as well as other factors based on night exposure to light. Thus, it is possible to recommend preventive actions and to reduce the incidence of prematurity.

References

Breiman, L. (2001). Statistical modeling: The two cultures. *Statistical Science*.

Cai, C., Vandermeer, B., Khurana, R., Nerenberg, K., Featherstone, R., Sebastianski, M., and Davenport, M. H. (2019). The impact of occupational shift work and working hours during pregnancy on health outcomes: a systematic review and meta-analysis.

Cook, T. R. (2020). Neural Networks. In *Advanced Studies in Theoretical and Applied Econometrics*.

Croteau, A., Marcoux, S., and Brisson, C. (2007). Work activity in pregnancy, preventive measures, and the risk of preterm delivery. *American Journal of Epidemiology*.

Davari, M. H., Naghshineh, E., Mostaghaci, M., Mirmohammadi, S. J., Bahaloo, M., Jafari, A., and Mehrparvar, A. H. (2018). Shift Work Effects and Pregnancy Outcome: A Historical Cohort Study. *Journal of family & reproductive health*.

Davis, S., Mirick, D. K., and Stevens, R. G. (2001). Night shift work, light at night, and risk of breast cancer. *Journal of the National Cancer Institute.*

De Arriba-Pérez, F., Caeiro-Rodríguez, M., and Santos-Gago, J. M. (2018). How do you sleep? Using off the shelf wrist wearables to estimate sleep quality, sleepiness level, chronotype and sleep regularity indicators. *Journal of Ambient Intelligence and Humanized Computing.*

Denbow, M. and Lyon, A. (2005). Obstetrics Illustrated. *The Obstetrician & Gynaecologist.*

Facco, F. L., Parker, C. B., Hunter, S., Reid, K. J., Zee, P. P., Silver, R. M., and et al. (2019). Later sleep timing is associated with an increased risk of preterm birth in nulliparous women. *American Journal of Obstetrics & Gynecology MFM*, 1(4):100040.

Fergus, P., Idowu, I., Hussain, A., and Dobbins, C. (2016). Advanced artificial neural network classification for detecting preterm births using ehg records. *Neurocomputing*, 188:42 – 49. Advanced Intelligent Computing Methodologies and Applications.

Ferguson, K. K., O'Neill, M. S., and Meeker, J. D. (2013). Environmental contaminant exposures and preterm birth: A comprehensive review.

Fernández-Navarro, T., Díaz, I., Gutiérrez-Díaz, I., Rodríguez-Carrio, J., Suárez, A., de Los Reyes-Gavilán, C. G., Gueimonde, M., Salazar, N., and González, S. (2019). Exploring the interactions between serum free fatty acids and fecal microbiota in obesity through a machine learning algorithm. *Food research international (Ottawa, Ont.)*, 121:533–541.

Ferreira, D. S., Amaral, F. G., Mesquita, C. C., Barbosa, A. P. L., Lellis-Santos, C., Turati, A. O., Santos, L. R., Sollon, C. S., Gomes, P. R., Faria, J. A., Cipolla-Neto, J., Bordin, S., and Anhê, G. F. (2012). Maternal melatonin programs the daily pattern of energy metabolism in adult offspring. *PLoS ONE.*

Gao, C., Osmundson, S., Edwards], D. R. V., Jackson, G. P., Malin, B. A., and Chen, Y. (2019). Deep learning predicts extreme preterm birth from electronic health records. *Journal of Biomedical Informatics*, 100:103334.

Garaulet, M. and Ordovás, J. M. (2013). *Chronobiology and obesity.*

Gil-Pita, R., Ayllón, D., Ranilla, J., Llerena-Aguilar, C., and Díaz, I. (2015). A computationally efficient sound environment classifier for hearing AIDS. *IEEE Transactions on Biomedical Engineering.*

Giorgis-Allemand, L., Pedersen, M., Bernard, C., Aguilera, I., Beelen, R. M., and et al. (2017). The influence of meteorological factors and atmospheric pollutants on the risk of preterm birth. *American Journal of Epidemiology.*

Goldenberg, R. L., Culhane, J. F., Iams, J. D., and Romero, R. (2008). Epidemiology and causes of preterm birth.

Guo, P., Miao, H., Chen, Y., Fu, Y., Wu, Y., Zhao, Q., and Li, B. (2019). Maternal exposure to gaseous ambient air pollutants increases the risk of preterm birth in the pearl river delta, china 2014–2017. *Science of The Total Environment*, 671:959 – 970.

Hersh, C., Sisti, J., Richiutti, V., and Schernhammer, E. (2015). The effects of sleep and light at night on melatonin in adolescents. *Hormones.*

Kuhn, M. and Johnson, K. (2013). *Applied predictive modeling.*

Kuhn, M., Weston, S., Culp, M., Coulter, N., and Quinlan, R. (2018). C50: C5.0 Decision Trees and Rule-Based Models.

Li, J., Liu, L., Zhou, M. C., Yang, J. J., Chen, S., Liu, H. T., Wang, Q., Pan, H., Sun, Z. H., and Tan, F. (2018). Feature selection and prediction of small-for-gestational-age infants. *Journal of Ambient Intelligence and Humanized Computing.*

Llurna, O., Fabre, E., Goya, M., Ruiz, C., and Valls, A. (2015). Prevención y manejo del parto pretérmino. *Documentos de Consenso S.E.G.O.*

Loy, S. L., Cheung, Y. B., Cai, S., Colega, M. T., Godfrey, K. M., Chong, Y.-S., Shek, L. P.-C., Tan, K. H., Chong, M. F.-F., Yap, F., and Chan, J. K. Y. (2020). Maternal night-time eating and sleep duration in relation to length of gestation and preterm birth. *Clinical Nutrition*, 39(6):1935 – 1942.

Lumley, J. (2003). Defining the problem: The epidemiology of preterm birth. In *BJOG: An International Journal of Obstetrics and Gynaecology.*

Montañés, E., Quevedo, J. R., Díaz, I., and Ranilla, J. (2009). Collaborative tag recommendation system based on logistic regression. In *CEUR Workshop Proceedings.*

Mustieles, V., Zhang, Y., Yland, J., Braun, J. M., Williams, P. L., Wylie, B. J., Attaman, J. A., Ford, J. B., Azevedo, A., Calafat, A. M., Hauser, R., and Messerlian, C. (2020). Maternal and paternal preconception exposure to phenols and preterm birth. *Environment International*, 137:105523.

Nakamura, Y., Tamura, H., Kashida, S., Takayama, H., Yamagata, Y., Karube, A., Sugino, N., and Kato, H. (2001). Changes of serum melatonin level and its relationship to feto-placental unit during pregnancy. *Journal of Pineal Research.*

Nehme, P. A., Amaral, F. G., Middleton, B., Lowden, A., Marqueze, E., França-Junior, I., Antunes, J. L., Cipolla-Neto, J., Skene, D. J., and Moreno, C. R. (2019). Melatonin profiles during the third trimester of pregnancy and health status in the offspring among day and night workers: A case series. *Neurobiology of Sleep and Circadian Rhythms.*

Pallesen, S., Sivertsen, B., Nordhus, I. H., and Bjorvatn, B. (2014). A 10-year trend of insomnia prevalence in the adult Norwegian population. *Sleep Medicine.*

Peterson, L. (2009). K-nearest neighbor. *Scholarpedia.*

Quinlan, R. J. (1994). C4.5: Programs for machine learning. *Machine Learning*, 16(3):235–240.

Quinlan, R. J. (2000). Data Mining Tools See5 and C5.0.

Rawashdeh, H., Awawdeh, S., Shannag, F., Henawi, E., Faris, H., Obeid, N., and Hyett, J. (2020). Intelligent system based on data mining techniques for prediction of preterm birth for women with cervical cerclage. *Computational Biology and Chemistry*, 85:107233.

Reiter, R. J., Tan, D. X., Korkmaz, A., and Rosales-Corral, S. A. (2014). Melatonin and stable circadian rhythms optimize maternal, placental and fetal physiology. *Human Reproduction Update.*

Ren, Z., Zhu, J., Gao, Y., Yin, Q., Hu, M., Dai, L., and et al. (2018). Maternal exposure to ambient PM10 during pregnancy increases the risk of congenital heart defects: Evidence from machine learning models. *Science of the Total Environment*.

Rezaeian, A., Rezaeian, M., Khatami, S. F., Khorashadizadeh, F., and Moghaddam, F. P. (2020). Prediction of mortality of premature neonates using neural network and logistic regression. *Journal of Ambient Intelligence and Humanized Computing*.

Sluban, B., Gamberger, D., and Lavra, N. (2010). Advances in class noise detection. In *Frontiers in Artificial Intelligence and Applications*.

Stieb, D. M., Chen, L., Eshoul, M., and Judek, S. (2012). Ambient air pollution, birth weight and preterm birth: A systematic review and meta-analysis.

Sun, S., Weinberger, K. R., Spangler, K. R., Eliot, M. N., Braun, J. M., and Wellenius, G. A. (2019). Ambient temperature and preterm birth: A retrospective study of 32 million us singleton births. *Environment International*, 126:7 – 13.

Touitou, Y., Reinberg, A., and Touitou, D. (2017). Association between light at night, melatonin secretion, sleep deprivation, and the internal clock: Health impacts and mechanisms of circadian disruption.

Verbaeten, S. (2002). Identifying mislabeled training examples in ILP classification problems. In *Proceedings of twelfth Belgian-Dutch conference on machine learning*.

Voiculescu, S. E., Zygouropoulos, N., Zahiu, C. D., and Zagrean, A. M. (2014). Role of melatonin in embryo fetal development.

Weber, A., Darmstadt, G. L., Gruber, S., Foeller, M. E., Carmichael, S. L., Stevenson, D. K., and Shaw, G. M. (2018). Application of machine-learning to predict early spontaneous preterm birth among nulliparous non-hispanic black and white women. *Annals of Epidemiology*, 28(11):783 – 789.e1.

Wheway, V. (2001). Using boosting to detect noisy data. In *Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science)*.

Zhang, S. (2012). Nearest neighbor selection for iteratively knn imputation. *Journal of Systems and Software*, 85(11):2541–2552.

Zhu, J. L., Hjollund, N. H., and Olsen, J. (2004). Shift work, duration of pregnancy, and birth weight: The National Birth Cohort in Denmark. *American Journal of Obstetrics and Gynecology*.