Computer Methods and Programs in Biomedicine 219 (2022) 106740

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

Prediction of the mode of delivery using artificial intelligence algorithms

Alberto De Ramón Fernández [a], Daniel Ruiz Fernández [a,*], María Teresa Prieto Sánchez [b]

[a] Department of Computer Technology (DTIC), University of Alicante, Carretera San Vicente s/n, Alicante 03690, Spain
[b] Service of Gynecology and Obstetrics, "Virgen de la Arrixaca" University Clinical Hospital, Institute for Biomedical Research of Murcia (IMIB-Arrixaca), Ctra. Madrid-Cartagena, s/n, El Palmar, Murcia 30120, Spain

A R T I C L E I N F O

Article history:
Received 12 June 2021
Revised 27 February 2022
Accepted 8 March 2022

Keywords:
Gynaecology
Mode of delivery prediction
CDSS
Artificial intelligence
Machine learning

A B S T R A C T

Background and objective: Mode of delivery is one of the issues that most concerns obstetricians. The caesarean section rate has increased progressively in recent years, exceeding the limit recommended by health institutions. Obstetricians generally lack the necessary technology to help them decide whether a caesarean delivery is appropriate based on antepartum and intrapartum conditions.
Methods: In this study, we have tested the suitability of using three popular artificial intelligence algorithms, Support Vector Machines, Multilayer Perceptron and, Random Forest, to develop a clinical decision support system for the prediction of the mode of delivery according to three categories: caesarean section, euthocic vaginal delivery and, instrumental vaginal delivery. For this purpose, we used a comprehensive clinical database consisting of 25,038 records with 48 attributes of women who attended to give birth at the Service of Obstetrics and Gynaecology of the University Clinical Hospital "Virgen de la Arrixaca" in the Murcia Region (Spain) from January of 2016 to January 2019. Women involved were patients with singleton pregnancies who attended to the emergency room on active labour or undergoing a planned induction of labour for medical reasons.
Results: The three implemented algorithms showed a similar performance, all of them reaching an accuracy equal to or above 90% in the classification between caesarean and vaginal deliveries and somewhat lower, around 87% between instrumental and euthocic.
Conclusions: The results validate the use of these algorithms to build a clinical decision system to help gynaecologists to predict the mode of delivery.

1. Introduction

Mode of delivery is nowadays one of the most relevant concerns that worry obstetricians, sanitary authorities and mothers. During the last decades, caesarean rates have been progressively increasing worldwide despite the World Health Organization recommendations that suggest global caesarean rates of less than 15% [1]. However, defining an acceptable caesarean delivery rate and what rate achieves optimal maternal and infant outcomes is also controversial according to others obstetrics scientific societies [2].

Progressive medicalization of childbirth is one of the reasons for which caesarean rates have raised. The increase in the number of caesarean sections also leads to a secondary increase in their rate as a consequence of the iterative ones, since they constitute a risk for having another one in a subsequent pregnancy [3]. Moreover, women fairly want to be aware and participate in the decisions about their pregnancies and births [4]. In this context, the medicalization of birth is highly criticized by some social sectors while others strongly seek to avoid natural experiences at delivery and demand closer medical assistance which lead to higher caesarean section rates. Many countries have designed strategies to control some medical factors that could contribute to reducing this tendency especially in non-elective caesarean sections and in low-risk pregnancies in which the end of delivery and the caesarean indications highly depends on medical criteria and interventions [5,32]. Many clinical conditions have been described as related to a successful vaginal birth. However, obstetricians usually lack modern tools that help them to make a decision on which is the best way

* Corresponding author.
E-mail address: druiz@dtic.ua.es (D. Ruiz Fernández).

https://doi.org/10.1016/j.cmpb.2022.106740

to finish each woman labour and if making a caesarean section is justified according to antepartum and intrapartum conditions. For other purposes, the use of artificial intelligence (AI) algorithms as a tool to support decision making and diagnostic tasks have been increasing in the last decades in the field of obstetrics.

For example, in the study of Mas-Cabo et al. [6] a type of artificial neural network (ANN), the Multilayer Perceptron (MLP) was used to predict imminent labour in women with evidence of preterm labour through the characterization of uterine electrohysterogram (EHG) signals. ANN and EHC signals were also used in [7] for the prediction of labour induction success. Other authors have applied ANN for different purposes such as placenta cell classification [8], prediction of pregnancy disorders development in the early first trimester [9] and evaluation of the risk for caesarean delivery in term nulliparous [10].

Support Vector Machines (SVM) has also been an algorithm frequently used for its ability to solve nonlinear problems through the kernel trick and to maximize classification margins thus obtaining a better prediction. As an example, Ocak [11] developed a computer decision support system (CDSS) based on SVM and genetic algorithms for the evaluation of foetal well-being. Other studies support the use of this algorithm for intrapartum foetal heart rate classification [12], for prediction of the labour onset type (spontaneous or induced) [13] for the intrauterine growth restriction risk evaluation and for foetal weight estimation [14]. Decision trees (DT) were used in [15,16] for the classification of cardiotocograms and for predicting hypertensive disorders in high-risk pregnancies [17]. Finally, in [18] different AI algorithms were compared to suggest the most suitable treatment in ectopic pregnancies.

These works are just some examples of how AI has recently contributed to develop models which can improve diagnostic precision in many areas of the obstetrics field. However, to the best of our knowledge, no previous studies that only use clinical data have been carried out on developing a CDSS that allows obstetricians to decide with greater reliability when to perform a caesarean section or if a vaginal delivery (instrumented or euthocic) will be feasible according to antepartum and intrapartum conditions. It is in this context where our proposal can make sense: situations in which there is one gynaecologist, maybe in a small or rural hospital or in holidays (with a reduced staff service), and the clinical professional has doubts about whether to perform a caesarean section or not; Our proposal can provide another opinion (which does not replace in any case that of the healthcare professional) to help with the final decision. In addition, if the delivery is vaginal, the system can provide information on the type of delivery to be expected (to prepare the necessary instruments).

Depending on time constraints, the use of a clinical decision support system can be problematic (e.g., in an urgent situation where time is very limited). In the framework that concerns us, the decision times allow the analysis of the situation, variables associated with the patient and also the use of the system proposed in this work. Keep in mind that, once the system is trained, the response of the software is immediate. In a real-time clinical environment, the system would have been pre-trained for clinical use.

Significant contributions of this research work are the following:

- A system is presented to help health staff decide if a delivery should end with a caesarean section or not.
- In case of vaginal delivery, our proposal also analyses if you may need instruments or not.
- An analysis, using a clinical dataset, of several AI methods (SVM, MLP and Random Forest) to help to decide the type of delivery is presented.
- The results obtained in the analysis showed a good performance with very high accuracy.

2. Methods

In this paper, we have tested the suitability of using different AI algorithms to develop a CDSS for the prediction of the mode of delivery according to three categories: caesarean section, euthocic vaginal delivery and instrumental vaginal delivery (vacuum, forceps and spatulas). The system is composed of two classifiers that work independently. The first one classifies the cases between vaginal birth and caesarean section. For vaginal deliveries cases, the second classifier classifies between those that required the use of instrumental and those that did not.

The database used consisted of cases of women who attended to give birth at the Department of Obstetrics and Gynaecology of the University Clinical Hospital "Virgen de la Arrixaca" in the Murcia Region (Spain) from January of 2016 to January 2019. Women involved were patients with singleton pregnancies who attended to the emergency room on active labour or undergoing a planned induction of labour for medical reasons. Forty-eight personal and medical variables were obtained from each patient, some of these variables were obtained antepartum and others intrapartum.

End of labour or mode of delivery was recorded from each patient as the main outcome and classified as: euthocic vaginal delivery, instrumental vaginal delivery or intrapartum caesarean section.

Medical assistance for these women during the delivery process was performed by trained midwifes and gynaecologists according to the protocols from the Spanish Gynaecology and Obstetrics Society (SEGO) and the Spanish National Health System recommendations [19]. The ending or mode of delivery (vaginal or caesarean) were decided according to medical criteria to safeguard maternal and foetal wellbeing.

2.1. Dataset preprocessing

Before developing the AI algorithms, it was necessary to preprocess the data using reduction, preparation, and balancing techniques. The original database had 25,038 records with 48 attributes. Data reduction is a preprocessing technique that pursues a reduced representation of the original data, discarding attributes that are not important in decision making, are duplicated, or have some linear relationship with other attributes. This reduces the dimensionality of the data and allows the algorithms to operate faster and efficiently.

For dimensionality reduction we decided to discard those attributes with very few or no values (e.g., pain relief, date and time of anaesthesia start...), duplicate or redundant attributes, attributes with direct relationship (e.g., obstetric formula and G with parity, abortions, caesarean sections and ectopic), attributes with a single value (e.g., type of monitoring, posture in labour), attributes whose value is known after delivery (e.g., sex, episiotomy type, day and month of delivery) and attributes that do not add value or of a private data (e.g., medical record number).

From the remaining attributes, a first selection was made by Dr Sánchez Prieto. Finally, the “Backward Elimination” technique was used applying a linear regression model (ordinary least squares) to exclude those variables that have the least partial correlation with the dependent variable. This will be the first candidate to be considered for elimination. The model is then re-evaluated to verify that precision has not been lost. In this case, a new variable is proposed for elimination. The process is repeated until the elimination of the variable represents a significant loss of precision.

For data reduction, we discarded those records with missing values, incorrect format, inconsistent data or outliers. Regarding the target variable, non-women who underwent elective or urgent caesarean section neither multiple pregnancies were included Fig. 1. shows a flowchart selection of the study cohort.

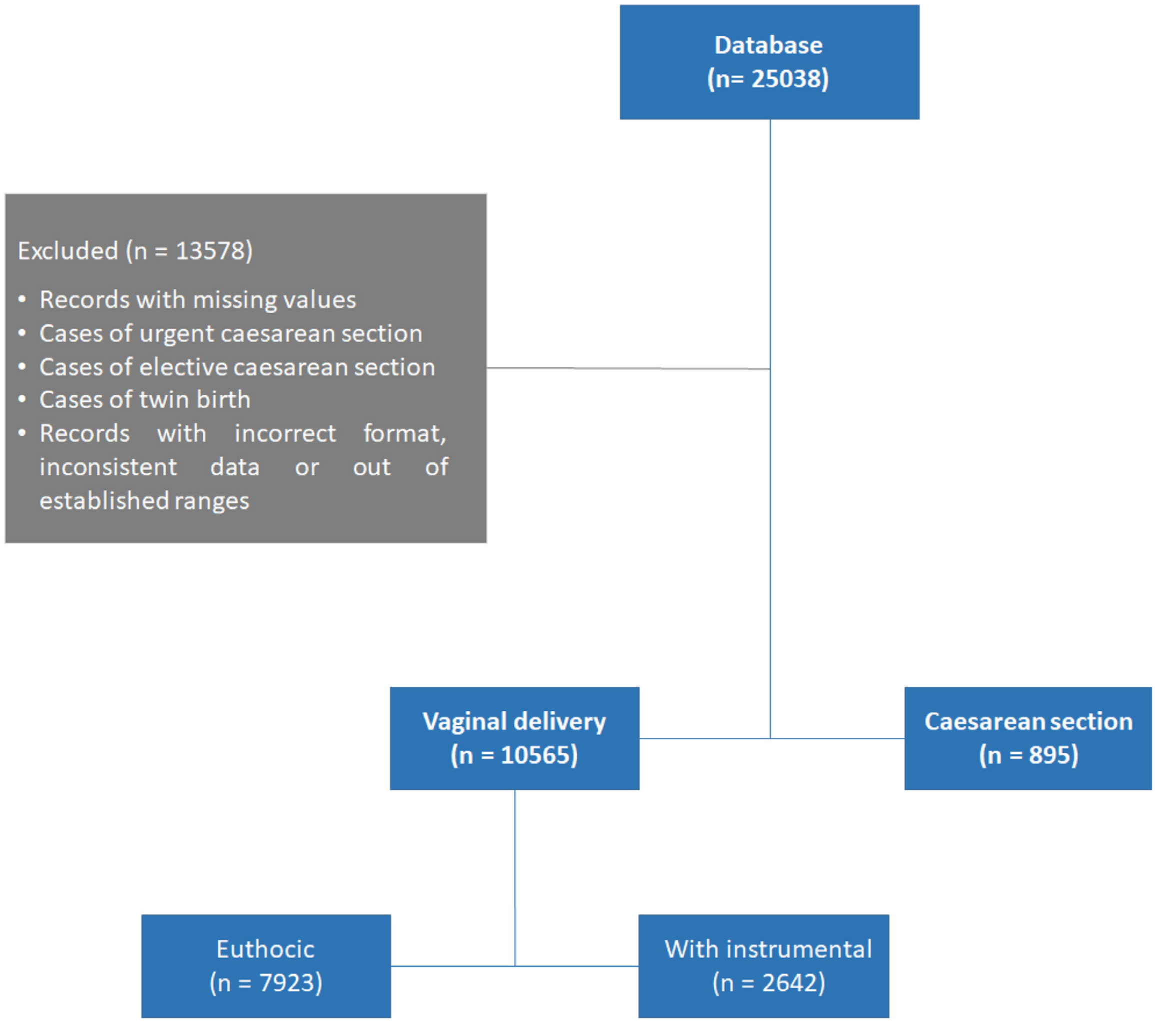

Fig. 1. Flowchart showing selection of the study cohort.

The attributes presented heterogeneous types of data: continuous, discrete and categorical. The preparation of the data is aimed at correctly initializing the data that will serve as input for the algorithms. Continuous data was previously normalized within a uniform range, using the following formula:

X_scaled = X_std * (max - min) + min

where min and max are the limits of the range of values of each attribute and X_std is given by:

X_std = (X - X.min) / (X.max - X.min)

Then, we discretized categorical and continuous data to favour the learning process and allow for simpler models, avoiding overfitting. After preprocessing, 12 attributes were selected and categorized (see Table 1):

For the training of the first classifier, all the cases of vaginal deliveries and caesarean sections were selected. After the reduction of the dataset, there was a strong imbalance of the majority class (vaginal) compared to the minority class (caesarean) with a ratio of 11.8:1. To avoid overfitting and that the model tended to predict the majority class with the highest accuracy and frequency than the minority, the majority class was reduced by random downsampling until to a ratio of 1:1. To develop the second classifier, cases of vaginal deliveries were classified between those that were instrumental (instrumental vaginal deliveries) and those that did not (euthocic vaginal deliveries). In this case, there was also a strong imbalance of the majority class (euthocic) compared to the minority class (instrumental) with a 3 to 1 proportion, so the majority class was reduced until a ratio of 1:1. Thus, a consistent source of data was obtained for the development of AI algorithms.

2.2. Artificial intelligence algorithms

In the experimentation, we have tested the suitability of three of the most implemented AI algorithms in supervised learning for the design of our two-stage classifier: Support Vector Machines, Multilayer Perceptron and Random Forest.

2.2.1. Support vector machines

SVM are a set of supervised learning methods developed by Sain and Vapnik [20]. It is used in classification and regression problems that use machine learning theory to maximize predictive precision and automatically avoid overfitting of data. SVM seek to solve an optimization problem.

2.2.2. Multilayer perceptron

ANN are a set of simple processing units called nodes or neurons connected to each other and organized in layers. ANN offer

Table 1
Dataset attributes and the target variable.

Database attributes	
Name	Description
1. Mother's age (MA) (years*) *Normalized from 1 to 10	Mother's age at delivery.
2. Gestational age (GA) (weeks*) *Normalized from 1 to 10	Total duration of gestation in weeks. This variable is calculated according to the womenś last menstrual period and adjusted (or recalculated) by using the crown-rump length (CRL) of the foetus in the ultrasound performed in the first trimester of pregnancy (from 11 to 14 weeks of gestation).
3. Parity (P) (number) [0 to 8]	Number of previous deliveries of a foetus of more than 20 weeks of gestation.
4. Caesarean sections (C) (number) [0 to 3]	Number of previous caesarean sections.
5. Abortions (A) (number) [0 to 7]	Number of previous abortions or miscarriages defined as the pregnancy losses that occur before 20 weeks of gestation.
6. Ectopic pregnancies (E) (number) [0 to 3]	Number of previous ectopic pregnancies (pregnancies located outside the uterine cavity).
7. Time from rupture of membranes to the onset of labour (TRM) (hours*) *Normalized from 1 to 10	Time between the rupture of membranes to the beginning of the active stage of labour, dilatation period or first stage.
8. Membranes status at admission (MS) (type) [1 to 2]	State of the amniotic sac membrane at admission: ruptured or intact.
9. Intrapartum fever (IF) (yes/no) [0,1]	Maternal temperature higher than 38°C during labour.
10. Beginning of labour (BL) (type) [1 to 2]	Mode of beginning of the active uterine contractions: spontaneous labour or medical induction).
11. Amniotic fluid (AF) (type) [1 to 5]	Amniotic fluid characteristics: normal, increased, diminished, dyed or haemorrhagic.
12. Anaesthesia (yes/no) (AN) [0,1]	Epidural anaesthesia.
Target variable	
Mode of delivery (MD) (caesarean section, vaginal delivery) [0,1]	
Mode of delivery (MD) (euthocic vaginal delivery, instrumental vaginal delivery) [0,1]	

characteristics such as adaptability, fault tolerance, robustness and parallel processing, among others [21]. They can be classified according to different criteria such as their topology (single or multilayer), the type of learning (supervised or unsupervised) or the type of connection between layers (feedforward or feedback). MLP was described by Rosenblatt [22] and it is one of the most implemented network topologies. It consists of an input layer, an output layer, and one or more hidden layers. The neurons in the input layer are responsible for receiving the input parameters and propagating them to the output layer through the hidden layers. One of the commonly implemented training methods for supervised learning is the Backpropagation method, a gradient descent algorithm introduced by Rumelhart et al [23].. The algorithm compares the network output with the desired one and the error made is propagated backwards through the hidden layers. The process is repeated until the correct output from the network is achieved.

2.2.3. Random Forest

Random Forest (RF) algorithm applies the model ensemble technique known as Bagging to improve the overall performance [24]. This technique seeks to obtain a generalized prediction combining the results of multiple models. The original training dataset is divided evenly into as many subsets as models to be combined. The models are run in parallel and are independent of each other. This technique allows reducing the overfitting of models since none of them has all the training data. In classification problems, the predicted class will be the mode of the set of predictions.

2.3. Performance evaluation of algorithms

To obtain a winning algorithm at each stage, they were trained with 70% of the data and the remaining 30% was used to validate the predictions made by using the following metrics:

- Confusion matrix: It allows visualizing the performance of a supervised learning model. The columns of the matrix represent the number of predictions of each class and the rows the number of real cases of each class. Thus, the main diagonal represents the cases correctly predicted and the secondary diagonal the wrong ones.
- Precision: It quantifies the number of predictions of a class that actually belong to that class.
- Recall: It quantifies the number of predictions of a class made from among all the cases of that class existing in the dataset.
- F1-Score: It can be interpreted as a weighted average of the precision and recall, where an F1 score reaches its best value at 1 and worst score at 0.
- Accuracy: It represents the proportion of correct predictions among all the predictions made.

3. Experimentation and results

Models were designed, trained, and validated using Spyder, an open-source development environment for scientific programming written in Python 3.8 and Scikit-learn 0.22.

3.1. Support vector machines

Several configurations of the SVM algorithm were used to find the one which better separates the two classes by modifying the main configuration hyperparameters. In general, there is no rule that determines the optimal value of these parameters, and their value will depend mainly on the characteristics of the dataset and must be found empirically. We used a grid search as a tuning technique to compute the optimum values of the following hyperparameters:

- **C**: It is the regularization parameter that controls the trade-off between achieving a low training error and a low testing error. The strength of the regularization is inversely proportional to C. The larger C is, the less the final training error will be. However, too large a value of C can cause overfitting, and lose the generalization properties of the classifier.

Table 2
SVM performance metrics.

SVM	Class	Precision	Recall	F1-Score	Accuracy	Confusion matrix	
1st classifier	Caesarean	1.00	0.82	0.90	0.91	True caesarean 224	False vaginal 50
	Vaginal	0.84	1.00	0.91		False caesarean 1	True vaginal 262
2nd classifier	Euthocic	0.87	0.86	0.86	0.86	True euthocic 699	False instrumental 113
	Instrumental	0.85	0.86	0.86		False euthocic 109	True instrumental 664

- **Kernel:** It specifies the type of kernel function (linear, RBF, sigmoidal, Gaussian, polynomial…).
- **Degree:** Degree of the polynomial kernel function.
- **Gamma:** Kernel coefficient for RBF, polynomial and sigmoidal kernels.

For the first and second classifier best parameters set found by grid search were {'C': 100, 'gamma': 0.001, 'kernel': 'RBF'} and {'C': 100, 'kernel': 'linear'}, respectively Table 2. shows the obtained results for the winning algorithms. In the first stage, the classifier was able to correctly classify 91% of cases. The performance for the second classifier was 86%.

3.2. Multilayer perceptron

Different topologies of the MLP were trained using the backpropagation method. For this, some of the configuration parameters taken into account were:

- **Hidden layers:** Number of intermediate layers of the MLP. A larger number of hidden layers requires more computing time.
- **Hidden neurons:** Number of neurons in the hidden layers. A larger number of neurons requires more computing time.
- **Activation function:** It defines the output of a neuron based on an input or set of inputs.
- **Epochs:** Number of training cycles used for training the neural network.
- **Batch size:** This parameter controls how often the network weights are updated.
- **Optimizer:** It specifies the algorithm used to update configuration parameters of the network.
- **Loss function:** It is a quantitative measure of deviation between the predicted output and the true output after each epoch.
- **Criterion for early stopping:** It establishes the convergence criteria and avoids overfitting during training.

For the first and second classifier best parameters set found by grid search were: {'batch_size': 10, 'epochs': 400, 'optimizer': adam, 'neurons': 10, activation: relu} and {'batch_size': 10, 'epochs':400, 'optimizer': adam, 'neurons':8, activation: logistic}, respectively. The Cross-Entropy Loss was used as a cost function and it was established as a stopping criterion that the cost function did not improve for 30 consecutive epochs Table 3. shows the results obtained by the MLP at each stage. The first classifier was able to correctly classify 90% of the cases, while the second correctly classified 86% of the cases.

3.3. Random Forest

For the design of this algorithm, different combinations were tested modifying the main configuration parameters:

- **Number of estimators:** The number of trees that will be combined for the final prediction.
- **Criterion:** The function to measure the quality of a split (*Gini index* or *entropy*).
- **Maximum depth:** This parameter sets the maximum of the nodes in the tree.
- **Minimum samples split:** It specifies the minimum number of samples required to split an internal node.
- **Minimum samples leaf:** The minimum number of samples per leaf in a node. A split point at any depth will be only considered if it leaves at least the specified value of training samples in each of the left and right branches of the tree.

To avoid overfitting, a maximum tree depth of 8 and a minimum of 4 samples per node and 2 samples for each leaf of the node were imposed as design conditions. For both, the first and second classifier best parameters set found were: {'n_estimators': 30, 'max_features': ['auto'], 'max_depth':8, min_samples_split=4, min_samples_leaf = 2 'criterion':'gini'} Table 4. shows the results provided by the algorithm. The first classifier was able to correctly classify between vaginal birth and caesarean section in 91% of cases, while the second correctly predicted 87% of cases.

For ensemble models, it is possible, and even common to perform no better than the best-performing member of the ensemble. For this reason, it is necessary to carefully evaluate the results offered by the ensemble model and by each of the models that compose it. For instance, in the study performed by Siddiqui et al [25]. two decision forest classifiers were applied to try to reduce the seizure detection while maintaining 100% accuracy. To evaluate the accuracy of the ensemble model, the average individual accuracy (AIA) and the ensemble accuracy (EA) were studied. AIA is the average of the accuracy of all individual trees in a forest and EA is the combined accuracy of all trees through the majority voting. For each of the datasets tested, EA was equal to or greater than AIA and the best individual accuracy. In our case, we obtain the results showed in Table 5:

Another important aspect is that the use of RF with an imbalanced data set can result in a loss of the predictive capability of the minority class. One of the possible solutions to correct this behaviour is to use a cost-sensitive learning method that relies on the use of a cost matrix describing the misclassification costs of an instance [26]. In our case, we applied random undersampling to manage imbalanced data.

4. Discussion

The results provided by the three AI algorithms have shown a good predictive ability to classify the mode of delivery, with an accuracy of around 90% for the classification between vaginal or cesarean delivery and around 86% for the classification between euthocic and instrumental. Of these, RF showed slightly higher accuracy (only 1%).

These results are also reflected in the AUC – ROC curve (Area under the curve - Receiver Operating Characteristics) shown in Fig. 2. It is a performance measurement for classification problems.

Table 3
MLP performance metrics.

MLP	Class	Precision	Recall	F1-Score	Accuracy	Confusion matrix	
1st classifier	Caesarean	0.97	0.83	0.89	0.90	True caesarean 227	False vaginal 47
	Vaginal	0.84	0.97	0.90		False caesarean 7	True vaginal 256
2nd classifier	Euthocic	0.86	0.87	0.87	0.86	True euthocic 705	False instrumental 107
	Instrumental	0.86	0.85	0.86		False euthocic 113	True instrumental 660

Table 4
RF performance metrics.

RF	Class	Precision	Recall	F1-Score	Accuracy	Confusion matrix	
1st classifier	Caesarean	0.99	0.82	0.90	0.91	True caesarean 224	False vaginal 50
	Vaginal	0.84	0.99	0.91		False caesarean 1	True vaginal 262
2nd classifier	Euthocic	0.85	0.90	0.87	0.87	True euthocic 731	False instrumental 81
	Instrumental	0.89	0.83	0.86		False euthocic 130	True instrumental 643

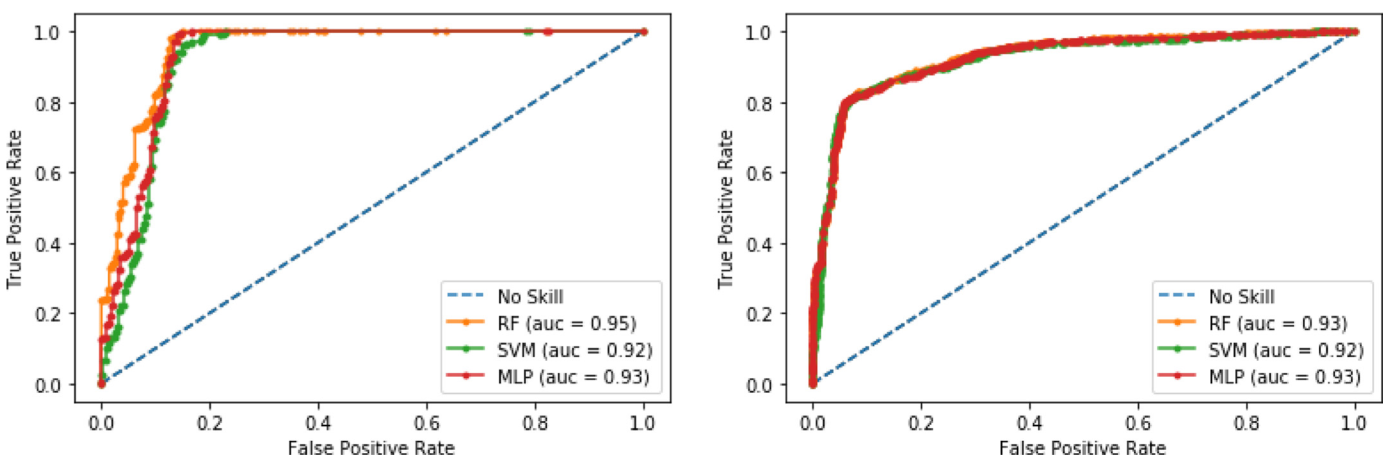

Fig. 2. AUC-ROC curve: (a) 1st classifier, (b) 2nd classifier.

ROC is a probability curve that represents the True Positive Rate (here, recall for vaginal cases) versus the True False Ratio (here, 1-Recall for caesarean cases). For its part, AUC shows the degree of separability. Higher the AUC, the better the model is at predicting between classes.

Between the classes of the target variable, the classifiers of the first stage had very high precision in the diagnosis of caesarean section cases at the cost of a lower recall. The opposite effect happened for vaginal delivery cases. In the second classification, the relation precision-recall was more balanced. At each stage, F1-Score was similar for all algorithms demonstrating its reliability in predicting both classes (see Fig. 3). The development of decision support systems is a complex process in which it is desirable to maximize performance measures such as precision and recall; in this way, systems obtained reduce false positives and negatives. These performance measures depend on many factors, including the nature of the decision itself. Although it is desirable to achieve high and balanced values, sometimes this is not possible and it must be clinical criteria (which can even change between hospitals) that should decide which measure to prioritize. Therefore, in our opinion, a system deployed in a hospital should be flexible enough so that clinicians can adapt it to their protocols (by choosing the appropriate algorithms) and, in this way, prioritize some performance measures over others (for example, in the first model, to prioritize caesarean decision over vaginal or the opposite).

The results obtained in this study improve those obtained with other works that are also dedicated to the prediction of the mode of delivery or the risk of suffering a caesarean section based on antepartum and intrapartum conditions. In [27], a cohort of 9,888 parturients with one previous caesarean delivery was used to evaluate the feasibility of using machine-learning methods to assign a personalized risk score for a successful vaginal birth after caesarean delivery. The gradient boosting model allows for individualized risk assessment with an AUC of 0.745 (95% CI = 0.728–0.762) for the first-trimester model and increases to an AUC of 0.793 (95% CI = 0.778–0.808) when applying the pre-labour model. In the study performed by Burke et.al [28], it was prospectively assessed the use of prenatally determined, maternal and foetal, anthropomorphic, clinical and ultrasound features to develop a predictive tool for unplanned caesarean de-

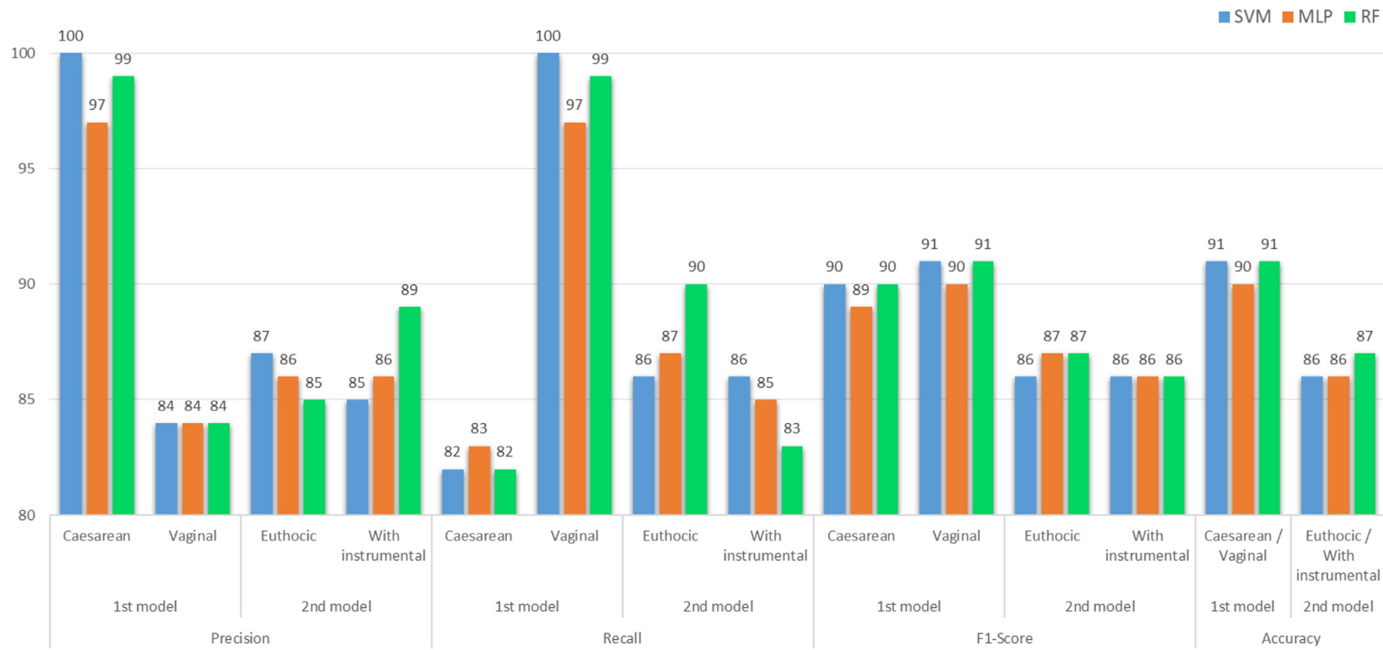

Fig. 3. Comparison of the results obtained by each model.

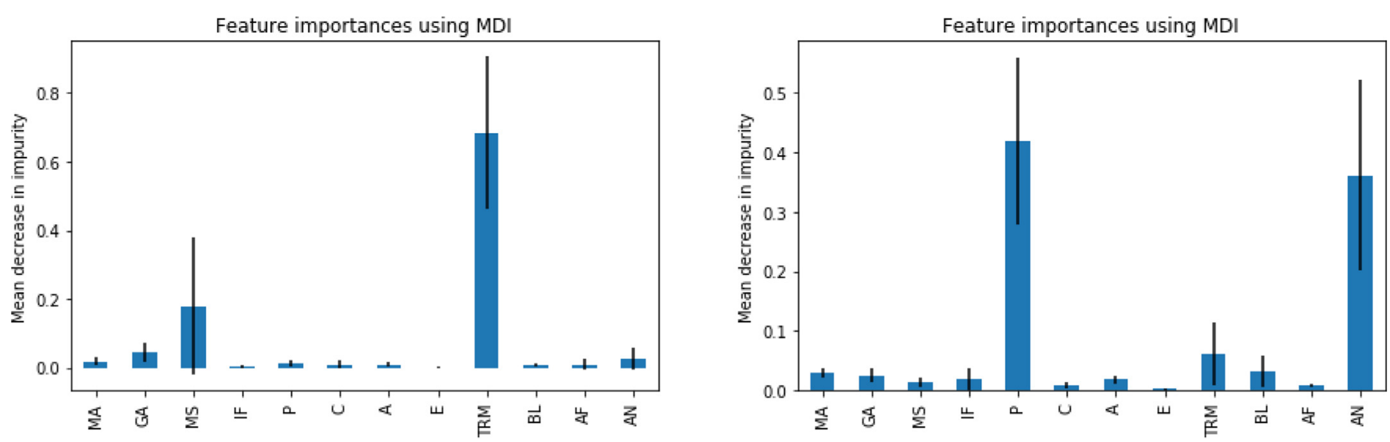

Fig. 4. Attribute importance: (a) 1st classifier, (b) 2nd classifier.

Table 5
AIA and EA for Random Forest classifier.

Classifier	AIA (%)	Best tree	Best individual accuracy	EA (%)
1st classifier	89,9	T3, T6-T8	90	91
2nd classifier	86,4	T15-T17	87	87

liveries from a total enrolled cohort of 2,336 nulliparous participants with a vertex presentation. For this, multiple logistic regression analysis and mathematical modelling were used to develop a risk evaluation tool with excellent calibration and discriminative ability (KS, d-statistic 0.29 95% CI 0.28–0.30) with a misclassification rate of 0.21 (95% CI 0.19–0.25). The same objective was raised in [29]. In this case, a neural network (NN) was trained with 225 patients and its results were compared with two logistic regression models, developed from a cohort with 225 and 600 patients (LR225 and LR600, respectively). The NN, LR225, and LR600 correctly predicted 53%, 26%, and 32% of the patients with caesarean deliveries and 88%, 95%, and 95% of the patients with vaginal delivery, respectively. Thus, compared with LRs, the NN was slightly better in predicting caesarean delivery and was similar for predicting vaginal delivery in nulliparous with term singletons (Table 5).

On the other hand, unlike SVM and MLP which act as black-box classifiers, RF provides logic rules that allow helping the understanding of the model and the extraction of additional information thanks to a knowledge discovery process. In our case, for the first classification and according to the mean decrease impurity (MDI), variables related to the state of the membranes seem to be the most relevant ones to predict vaginal delivery respect to caesarean delivery (see Fig. 4a). This is in agreement with medical evidence since the artificial rupture of membranes is a frequent method to induce labour in some women or to accelerate its course in cases of no progress of labour. Rupture of membranes facilitate the contact between foetal presentation and maternal tissues (especially the cervix) and stimulates dilatation and progression of the labour by Fergusonś reflex [30].

For the second classifier, to predict the need for instrumental deliveries vs. non-instrumental ones, the most relevant variables have been previous parity and the use of epidural anaesthesia (see Fig. 4b). Both results make sense with gynaecology practice. Womenś previous parity is a protective factor for the risk of undergoing an urgent caesarean section during current delivery. Regarding the use of anaesthesia in obstetrics, it is widely known that epidural anaesthesia influences the dilatation course and, also the risk of an instrumental delivery [31].

5. Conclusions

In recent years, the number of caesarean sections has increased to exceed the limit recommended by the main health institutions. Frequently and mainly in non-elective caesarean sections and in low-risk pregnancies, the choice between vaginal delivery and caesarean section highly depends on medical criteria and interventions. However, obstetricians usually lack novel technological solutions to help them predict the best option. In this paper, we study the feasibility of using the SVM, RNA, and RF algorithms to build a CDSS that helps to predict the mode of delivery.

The first classifier discriminated between caesarean and vaginal deliveries and the second one between euthocic vaginal and instrumental vaginal deliveries. The three algorithms showed a similar performance, although all of them were more accurate in classifying between vaginal and caesarean deliveries (accuracy equal or higher than 90%) than between instrumental and euthocic (accuracy equal or higher than 86%).

The fact that good results were obtained for the two classifiers and the three algorithms allow us to validate their use to build a CDSS that helps to predict the mode of delivery and suggests that both the number of records and the selection of attributes used for the training of the models were appropriate for solving the classification problem posed.

Declaration of Competing Interest

None.

References

[1] World Health Organization. WHO statement on caesarean section rates. (No. WHO/RHR/15.02). 2015. https://www.who.int/reproductivehealth/publications/maternal_perinatal_health/cs-statement/en/ (accessed 10 Aug 2020).
[2] J. Gupta, G. Smith, R. Chondankar, Royal college of obstetrician and gynaecologists (RCOG) green-top guideline no. 45: birth after previous caesarean birth, Green Top Guid. 45 (2015) 31 https://www.rcog.org.uk/globalassets/documents/guidelines/gtg_45.pdf.
[3] B.P. Sachs, C. Kobelin, M.A. Castro, et al., The risks of lowering the cesarean-delivery rate, N. Engl. J. Med. (1999), doi:10.1056/NEJM199901073400112.
[4] A. Shorten, A. Fagerlin, J. Illuzzi, et al., Developing an internet-based decision aid for women choosing between vaginal birth after cesarean and planned repeat cesarean, J. Midwifery Women's Health (2015) Published Online First, doi:10.1111/jmwh.12298.
[5] T. Bloomfield, Caesarean section, NICE guidelines and management of labour, J. Obstet. Gynaecol. (2004) (Lahore), doi:10.1080/0144361042331271052.
[6] J. Mas-Cabo, G. Prats-Boluda, J. Garcia-Casado, et al., Electrohysterogram for ann-based prediction of imminent labor in women with threatened preterm labor undergoing tocolytic therapy, Sensors 20 (2020) (Switzerland), doi:10.3390/s20092681.
[7] C. Benalcazar-Parra, Y. Ye-Lin, J. Garcia-Casado, et al., Prediction of labor induction success from the uterine electrohysterogram, J. Sens. (2019) Published Online First, doi:10.1155/2019/6916251.
[8] E. Rohith, V. Sowmya, K.P. Soman, Convolutional neural networks for placenta cell classification, in: Proceedings of the 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies ICICICT 2019, 2019, pp. 1273–1277, doi:10.1109/ICICICT46008.2019.8993145.
[9] D.S.A. Maylawati, M.A. Ramdhani, W.B. Zulfikar, et al., Expert system for predicting the early pregnancy with disorders using artificial neural network, in: Proceedings of the 5th International Conference Cyber IT Service Management CITSM 2017, 2017 Published Online First:, doi:10.1109/CITSM.2017.8089243.
[10] A. Al Housseini, T. Newman, A. Cox, et al., Prediction of risk for cesarean delivery in term nulliparas: a comparison of neural network and multiple logistic regression models, Am. J. Obstet. Gynecol. (2009) Published Online First, doi:10.1016/j.ajog.2009.05.001.
[11] H. Ocak, A medical decision support system based on support vector machines and the genetic algorithm for the evaluation of fetal well-being, J. Med. Syst. (2013) Published Online First, doi:10.1007/s10916-012-9913-4.
[12] J. Spilka, J. Frecon, R. Leonarduzzi, et al., Sparse support vector machine for intrapartum fetal heart rate classification, IEEE J. Biomed. Health Inform. (2017) Published Online First:, doi:10.1109/JBHI.2016.2546312.
[13] J. Alberola-Rubio, J. Garcia-Casado, G. Prats-Boluda, et al., Prediction of labor onset type: spontaneous vs induced; role of electrohysterography? Comput. Methods Programs Biomed. (2017) Published Online First:, doi:10.1016/j.cmpb.2017.03.018.
[14] F. Sereno, J.P. Marques De Sá, A. Matos, et al., Support vector regression applied to foetal weight estimation, in: Proceedings of the International Joint Conference on Neural Networks, 2001, doi:10.1109/ijcnn.2001.939576.
[15] P. Tomas, J. Krohova, P. Dohnalek, et al., Classification of cardiotocography records by Random Forest, in: Proceedings of the 36th International Conference on Telecommunications and Signal Processing, TSP 2013, 2013, doi:10.1109/TSP.2013.6614010.
[16] M. Arif, Classification of cardiotocograms using Random Forest classifier and selection of important features from cardiotocogram signal, Biomater. Biomech. Bioeng. (2015) Published Online First:, doi:10.12989/bme.2015.2.3.173.
[17] M.W.L. Moreira, J. Rodrigues, A.M.B. Oliveira, et al., Predicting hypertensive disorders in high-risk pregnancy using the Random Forest approach, in: Proceedings of the IEEE International Conference on Communications, 2017, doi:10.1109/ICC.2017.7996964.
[18] A. De Ramón Fernández, D. Ruiz Fernández, M.T. Prieto Sánchez, A decision support system for predicting the treatment of ectopic pregnancies, Int. J. Med. Inform. 129 (2019) 198–204, doi:10.1016/j.ijmedinf.2019.06.002.
[19] E. Innovación, Y. P. SOCIAL, Guía de Práctica Clínica sobre la Atención al Parto Normal. *Plan de Calidad para el Sistema Nacional de Salud del Ministerio de Sanidad y Política Social*, Agencia de Evaluación de Tecnologías Sanitarias del País Vasco (2010).
[20] S.R. Sain, V.N. Vapnik, The nature of statistical learning theory, Technometrics (1996) Published Online First:, doi:10.2307/1271324.
[21] I.A. Basheer, M. Hajmeer, Artificial neural networks: Fundamentals, computing, design, and application, J. Microbiol. Methods (2000) Published Online First, doi:10.1016/S0167-7012(00)00201-3.
[22] F. Rosenblatt, The perceptron: a probabilistic model for information storage and organization in the brain, Psychol. Rev. (1958) Published Online First:, doi:10.1037/h0042519.
[23] D.E. Rumelhart, G.E. Hinton, J.L. Mcclelland, A general framework for parallel distributed processing, Parallel Distrib. Process. Explor. Microstruct. Cogn. 1 (1986) 45–76.
[24] L. Breiman, Bagging predictors - springer, Mach. Learn. (1996).
[25] M.K. Siddiqui, M.Z. Islam, M.A. Kabir, A novel quick seizure detection and localization through brain data mining on ECoG dataset, Neural Comput. Appl. 31 (2019) 5595–5608, doi:10.1007/s00521-018-3381-9.
[26] M.K. Siddiqui, X. Huang, R. Morales-Menendez, et al., Machine learning based novel cost-sensitive seizure detection classifier for imbalanced EEG data sets, Int J Interact Des Manuf 14 (2020) 1491–1509, doi:10.1007/s12008-020-00715-3.
[27] M. Lipschuetz, J. Guedalia, A. Rottenstreich, et al., Prediction of vaginal birth after cesarean deliveries using machine learning, Am. J. Obstet. Gynecol. (2020) Published Online First, doi:10.1016/j.ajog.2019.12.267.
[28] N. Burke, G. Burke, F. Breathnach, et al., Prediction of cesarean delivery in the term nulliparous woman: results from the prospective, multicenter Genesis study, Am. J. Obstet. Gynecol. (2017), doi:10.1016/j.ajog.2017.02.017.
[29] A. Al Housseini, T. Newman, A. Cox, et al., Prediction of risk for cesarean delivery in term nulliparas: a comparison of neural network and multiple logistic regression models, Am. J. Obstet. Gynecol. (2009) Published Online First:, doi:10.1016/j.ajog.2009.05.001.
[30] P. Delorme, E. Lorthe, J. Sibiude, et al., Preterm and term prelabour rupture of membranes: a review of timing and methods of labour induction, Best Pract. Res. Clin. Obstet. Gynaecol. (2021), doi:10.1016/j.bpobgyn.2021.08.009.
[31] K.F. Walker, M. Kibuka, J.G. Thornton, et al., Maternal position in the second stage of labour for women with epidural anaesthesia, Cochrane Database Syst. Rev. (2018), doi:10.1002/14651858.CD008070.pub4.
[33] NICE updates guidance on caesareans.(Clinical Guidaline CG132), National Institute for Health and Care Excellence, 2011.