

# Development of a Predictive Model for Induction Success of Labour

Cristina Pruenza<sup>1\*</sup>, María Teulón<sup>2</sup>, Luis Lechuga<sup>2</sup>, Julia Díaz<sup>1</sup>, Ana González<sup>1</sup>

<sup>1</sup> Universidad Autónoma de Madrid and Instituto de Ingeniería del Conocimiento, Madrid (Spain)

<sup>2</sup> Hospital Universitario de Fuenlabrada, Madrid (Spain)

Received 15 July 2016 | Accepted 3 February 2017 | Published 27 March 2017



## ABSTRACT

Induction of the labour process is an extraordinarily common procedure used in some pregnancies. Obstetricians face the need to end a pregnancy, for medical reasons usually (maternal or fetal requirements) or less frequently, social (elective inductions for convenience). The success of induction procedure is conditioned by a multitude of maternal and fetal variables that appear before or during pregnancy or birth process, with a low predictive value. The failure of the induction process involves performing a caesarean section. This project arises from the clinical need to resolve a situation of uncertainty that occurs frequently in our clinical practice. Since the weight of clinical variables is not adequately weighted, we consider very interesting to know a priori the possibility of success of induction to dismiss those inductions with high probability of failure, avoiding unnecessary procedures or postponing end if possible. We developed a predictive model of induced labour success as a support tool in clinical decision making. Improve the predictability of a successful induction is one of the current challenges of Obstetrics because of its negative impact. The identification of those patients with high chances of failure, will allow us to offer them better care improving their health outcomes (adverse perinatal outcomes for mother and newborn), costs (medication, hospitalization, qualified staff) and patient perceived quality. Therefore a Clinical Decision Support System was developed to give support to the Obstetricians. In this article, we had proposed a robust method to explore and model a source of clinical information with the purpose of obtaining all possible knowledge. Generally, in classification models are difficult to know the contribution that each attribute provides to the model. We had worked in this direction to offer transparency to models that may be considered as black boxes. The positive results obtained from both the information recovery system and the predictions and explanations of the classification show the effectiveness and strength of this tool.

## KEYWORDS

Machine learning, Big Data, Electronic Medical Records, Induction of Labour, Predictive Models, Clinical Decision Support System.

DOI: 10.9781/ijimai.2017.03.003

## I. INTRODUCTION

**I**NDUCTION of the labour process, more commonly known as labour induction, is one of the most studied operations within the field of Obstetrics. Over the last two decades, the induction rate has doubled, turning into a fairly common procedure, used in over 20% of gestations [1], [2]. However, good predictive factors of the success of this procedure have not been identified yet, so there are not support tools for the experts' criteria so far. Currently, the decision of inducing is made only on the basis of clinical knowledge itself consisting of protocols, guides and previous experience of certain characteristics of mother and foetus, but in case of making unwise decisions it would cause serious complications [2], [3]. Under these circumstances, it seems interesting to know beforehand the probability of success of the induction in order to dismiss the inductions having a high probability of failure and thus improving results on health, reducing costs derived from medication, hospitalization or qualified staff. Therefore, one of the current challenges in Obstetrics is improving the prediction of successful induction of labour.

In works [4] and [5], the authors present usual variables related to situations where labour has been induced successfully, so these are good predictors and have been defined as a reference model. Nonetheless, some other situations present a lower predictive value than expected. In this paper, the effectiveness of these variables has been evaluated and other models are explored to determine some relevant variables with the aim of building Clinical Decision Support System tool [6].

Also, the healthcare model and, in general, the whole healthcare sector is nowadays one of the fields in which Big Data Technology is having a high impact on, and is experiencing an exponential growth in applications.

In this environment a high percentage of data, clinical evaluations and patient progress information are registered usually in free text fields on the Electronic Medical Records. This information should be processed and transformed into structured and normalised data. In our project, Machine Learning algorithms, Text Mining and Big Data techniques have been used to extract knowledge. A typical difficulty of applying these techniques is that algorithms outcomes are usually difficult to interpret. To prevent that, additional work was done to provide more transparency to the previous algorithms, especially those traditionally considered as black box, such as Neural Networks. Several approximations proposed in the literature were studied in the

\* Corresponding author.

E-mail address: cristina.pruenza@iic.uam.es

Master's thesis [7], the Strumbelj and Kononenko proposal was used in [8], based on the cooperative game theory, which allows us to obtain the contribution of each variable in the classification obtained by the algorithm.

## II. PROPOSED METHODOLOGY

In this project, we developed a tool that is able to exploit, structure and normalize a source of clinical information and that works as an aid for decision making, on the basis of a predictive model. In order to achieve this work, a multidisciplinary team was formed in which clinicians, health care data experts and machine learning researchers worked together. An important step for machine learning to have a meaningful role in healthcare and more specifically in Obstetrics field.

The processes of data acquisition, preparation and validation are essential and, at the same time, the most complex tasks of the project. If there was not structured information, it would not be possible to generate predictive models or to build the validation tool.

This section will be structured as follows. Firstly, the topic related to the collection of data from the patients' medical records will be discussed. Once the necessary variables are obtained, two divergent methodologies will be implemented: (1) expert system based on the rules provided by the obstetrician and (2) machine learning techniques will be briefly explained keeping the typical stages from pre-processing of information to validation of the implemented models. Because most of the times the applied models are complex and it is difficult to understand how the input variables are related to the output of the model, the section ends by pointing out that it is possible to give transparency to the models by measuring the contribution of each of the input variables.

### A. Data Integration

#### 1) Data Collection

The raw data set was provided by *Hospital Universitario de Fuenlabrada*, exported in plain text files directly from the Electronic Medical Records of the Selene platform [9]. Every file contains anonymized information about patients, according to the Spanish Personal Data Protection Act (L.O.P.D.), along with metadata and the type of document within the platform, that is, report, note, form or request.

As a whole, 3,509 reports, 399,646 unstructured notes, 764,783 forms and 235,102 test requests had been used as data sources. All of them in unstructured plain text.

The raw data come from the clinical records of 10,487 patients (healthcare assistance of pregnant women) in a period of time slightly more than 5 years. Most of the data were in free text format.

From the raw data an extraction process was performed to obtain relevant variables useful for later studies. The data extraction phase was performed using text mining techniques. The selected variables (attributes or features) to search in the clinical record was previously defined by an expert physician. A total of 21 variables were sought within each patient's medical history. Fig. 1 shows the attributes for each patient, organised in two categories. Attributes used to make the decision of inducing are in blue, while the object variable (class) of predictive models is in red and may take three possible values: *No induction*, *Induction* or *Caesarea*.

#### 2) Data Preparation

Often the extracted data are incomplete, contain unnecessary or ambiguous information, suffer disruptions due to noise or pose any other difficulty that affects performance of the predictive models. Therefore, it is necessary to pre-process them to avoid future inconveniences.

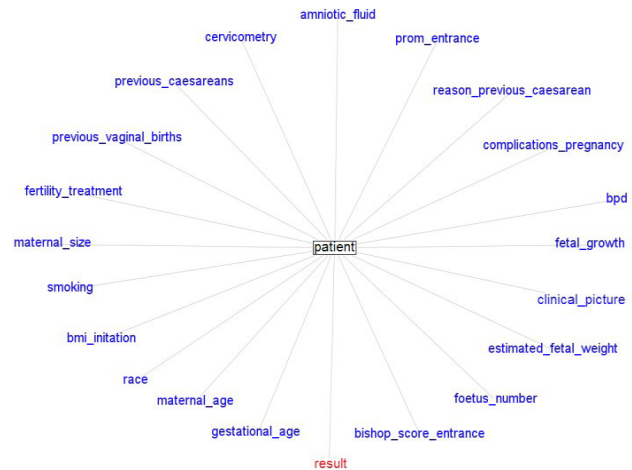


Fig. 1. Graph of attributes of a patient. Attributes used to make the decision of inducing are in blue and the object variable (class) is in red.

Data pre-processing, i.e. cleaning, includes deleting documents that are not classified according to Selene (reports, notes, forms or requests), documents from deliveries assisted elsewhere or from births presenting a gestational age inferior to foetal viability (current limit set of 23 weeks). This filtering process is indispensable to categorise the information into variables, as each one is dealt with particularly and the related information is extracted from a specific set of documents previously defined.

The process of extracting variables out of the patients' data is long and tedious, and needs some collaboration from the expert to be validated. The first step is extracting the terms, followed by a homogenization of capitals and deleting special characters. In order for the process to be quick, we performed a deletion of stopwords and a process of stemming.

As we mentioned before, getting the variables of the data is a compute-intensive phase because it requires a text parsing. Sequential and parallel execution modes were tested. But the runtime of the sequential algorithm was excessive because it was an iterative process. Therefore, the parallel version with threads was used.

#### 3) Validation of the Extracted Variables

In order to validate the goodness of the automatically extracted variables, collaboration of the expert on the field was needed. For the purpose to make the validation process user-friendly, we implemented a web application called as INDUCCESS (INDUCTION and sucCESS), where several experts may check both Electronic Medical Records and the automatically extracted variables representing each patient.

Fig. 2 shows a screenshot of the web application implemented to make such validation. Inside the application it is possible to navigate through the patients and validate or reject the outcomes from the extraction process.

In case of concordance between the automatically extracted variables and what is contained in the patient's medical history, the application executes the predictive or inference model and issues a result suggested for the patient (*No induction*, *Induction* or *Caesarea*). Among the functionality available in INDUCCESS, it is possible to visualize some statistical data and detailed information on the project, as well as information from the institutions collaborating or even send an email seeking advice.



Fig. 2. Screen to check validation of patients from the web application. Nowadays this application is only available in Spanish Language.

Validation of the system has been carried out with several incremental and iterative proofs of concept, starting offline and ending online. It is at this last stage that experts from *Hospital Universitario de Fuenlabrada* take part and access to the web application with the aim of reviewing a random subset of patients. Results obtained from this process have been considered satisfactory, rendering an error of 16.83%. However, we keep on working so that the discrepancy between patient variables and the real information should be minimized.

### B. Decision-making Rules

As it was mentioned before, there are no tools that support the expert in decision making within the field of Obstetrics. With the purpose of ameliorating this situation, we built a baseline model to aid decision making processes based on decision rules from a panel of experts on the field, formulated only according to their own clinical knowledge and experience. This baseline model was used to evaluate the effectiveness of variables and to search for other inference models determining which variables are relevant and improve results when predicting success of inducing labour.

An expert system was designed with the help of the CLIPS tool [10]. CLIPS stands out for providing a strategy of forward chaining inference, that is, it starts with an initial evidence and goes on until a solution is reached. Therefore the usual deductive reasoning of the expert was simulated. Within the system, the attributes representing the patients are part of the basis of facts used by the inference engine to check the knowledge base made of decision rules. However, not all features proposed are highly relevant when making decisions of inducing labour. Consequently the expert (obstetrician) sets the initial weights indicating the relevance of each feature and priorities were assigned to rules accordingly, following the same idea of the certainty factors ( $CF$ ) from the MYCIN system [11]. MYCIN was an expert system to identify bacteria causing severe infections; represents expert reasoning as a set of rules and  $CF$  of each rule is defined as the degree of belief in the hypothesis given the evidence [12]. Then, we use the opposite of  $CF$  for evidence in rules that contain negations,  $CF(\neg F)$ ; the minimum of  $CF$ s for conjunctions,  $CF(F_1 \wedge F_2)$ ; and the maximum of  $CF$ s for disjunctions,  $CF(F_1 \vee F_2)$ . In this project, priorities are calculated using  $CF$ s normalized with the equations (1), (2) and (3):

$$CF(\neg F) = -\frac{2 \cdot \text{weight}(F) - M}{M} \quad (1)$$

$$CF(F_1 \wedge F_2) = -\frac{2 \cdot \min[\text{weight}(F_1), \text{weight}(F_2)] - M}{M} \quad (2)$$

$$CF(F_1 \vee F_2) = -\frac{2 \cdot \max[\text{weight}(F_1), \text{weight}(F_2)] - M}{M} \quad (3)$$

where  $M = \max[\text{weights}]$ ,  $F$  is the feature that defines the rule and  $\text{weight}(\cdot)$  is the weight of the feature defined by the expert.

The attributes used and their weights are shown in Table I and it is fulfilled that the greater the value of the weight the greater is the influence of the variable.

After verifying the coherence of the system and eliminating redundant, unnecessary or conflicting rules, the problem is reduced to work with 35 rules.

The application INDUCESS incorporates the suggested result by the expert system.

### C. Machine Learning Techniques

#### 1) Feature Selection

Typically, there may be some irrelevant or redundant data that, if it is not deleted before training a machine learning model, performance may be affected. That is one of the reasons why the dimension of the original data should be reduced by selecting the most significant characteristics before using predictive models that support decision making.

TABLE I.  
WEIGHTS OF ATTRIBUTES PROPOSED BY THE EXPERTS TO DEVELOP THE BASELINE MODEL

Weights	Features
5-6	clinical_picture prom_entrance bishop_score_entrance estimated_fetal_weight fetal_growth gestational_age previous_caesareans
3-4	previous_vaginal_births foetus_number reason_previous_caesarean
1-2	bmi_initiation amniotic_fluid bpd maternal_size fertility_treatment maternal_age cervicometry complications_pregnancy
0	race smoking

In the present work, we studied a variety of algorithms for feature selection ReliefF [13], mRMR (minimum Redundancy Maximum Relevance) [14], Gain Ratio and Information Gain attribute evaluation [15], CFS (Correlation Feature based Selection) [16] through the Weka tool [17].

#### 2) Classification Algorithms

Machine Learning and Big Data build and study systems that are able to learn from vast amounts of data and to improve classification and prediction processes. In order for these data to be turned into knowledge, they must be processed and analysed with the models, but every model has its idiosyncrasies, so not all of them are suitable to solve any kind of problem.

In particular, in the medical service decision making processes are critical, as a wrong decision might affect people' health directly. Therefore, we analyse advantages and disadvantages of each algorithm in the medical practice. We look for models which offer an additional explanation or information justifying the decision, as it may help healthcare specialists gain some knowledge on the given problem. In [18] machine learning techniques and models traditionally applied in

medicine are reviewed, and requirements to be fulfilled are collected in order to be successful at this field. However, it is hard to know beforehand which method will be the most suitable one, so we tested several classification algorithms and compared their results in a number of experiments.

#### a) Naïve Bayes

Naive Bayes [19] is a probabilistic model that uses of Bayes' theorem in the classifier's decision rule. The Naive Bayes classifier assumes that all predictor variables are conditionally independent given the class.

Bayesian classifiers are one of the favourites in the medical field because they offer great ability to explain their predictions models. They have a good performance, acceptable noise levels are tolerated and have a good level of transparency although not as much as decision trees.

#### b) Decision Trees

Decision trees are similar to systems based decision rules used to represent and categorize a number of conditions that occur in succession. A decision tree is a flowchart-like structure in which each internal node represents a "test" on an attribute, each branch represents the outcome of the test and each leaf node represents a class label. We have used the C4.5 algorithm in the training phase to decide the questions to be formulated in each node of the tree.

Decision trees have high transparency and large capacity to explain each of the predictions.

#### c) Neural Networks

Artificial Neural Networks have become popular in medicine because of their flexibility and dynamism. Multilayer perceptron is an artificial neural network model that maps sets of input data onto a set of appropriate outputs. The algorithm utilizes the backpropagation technique for training the network.

Artificial Neural Networks were typically used as black box classifiers lacking the transparency of generated knowledge and lacking the ability to explain the decisions. However, in this paper we use a technique to explain the predictions emitted by the classifiers, thus providing an algorithm with more transparency and excellent performance.

#### d) Support Vector Machines

Support Vector Machines (SVM) adjust a set of parameters that allow you to set boundaries in the space of n dimensions and approximate functions or separate patterns in different regions of the attribute space.

SVM have good performance but transparency and the ability of explanation are poor. We improved this aspect by applying techniques as in [8].

#### e) Random Forests

Random forests are ensemble learning methods that operate by constructing a multitude of small decision trees at training time and outputting the class that is the mode of the classes of the individual trees. Random forest is considered one of the best performing algorithms, especially problems that have many explanatory variables [20].

However, although able to provide the important variables in the classification, unlike decision trees, its output is difficult to interpret.

### D. Explanation of Classification

Machine learning is becoming increasingly important in certain sectors of science, technology or business. The main purpose of machine learning is creating a model which is able to provide a satisfactory response when information is entered onto it. Medical professionals demand models which are able to explain their predictions.

In this article, we implemented the proposed algorithm first by

[8] and subsequently the master's thesis [7] made an extension. The objective is to estimate the contribution of each attribute to the model. In this way it is possible to give transparency to models that are difficult to explain, thus improving the interpretation of predictions.

## III. RESULTS AND DISCUSSION

In this section, systems proposed previously in subsection II B and II.C were evaluated using as measures the error in terms of percentage classification error (ErrClassif), precision or positive predictive value (Precision), recall or true positive rate (Recall), effectiveness measure (F-measure) and area under a ROC curve (AUC).

In order to guarantee independence between training and testing sets chosen and to get more stable results 10-fold cross-validation was implemented.

### A. Results using Decision-making Rules

Firstly, Table II shows the error reported with the implementation of the expert system, which we have considered the baseline model. As it was explained before, it is based on decision rules that infer from the variables of the data collected (Table I) and validated after the process of extraction.

TABLE II. RESULTS USING DECISION-MAKING RULES (BASELINE MODEL)

ErrClassif (%)	Precision	Recall	F-measure	AUC
41.89	0.567	0.581	0.571	0.646

It is worth mentioning that the errors obtained using the decision rules may not be entirely objective because they may be affected by the experience and knowledge of the expert (obstetrician) who formulates them. In spite of this, we took as reference the 41.89% of classification error.

As previously mentioned, we searched for some other models to improve this result and, moreover, that offer a consistent explanation of the classification obtained, making it easier to be used at the medical field.

### B. Results using Classification Algorithms

In order to reduce the dimensionality of the original set, we chose the algorithm CFS, as according our research in these issues, it selects the most relevant features for inducing, so the predictive model provides better results.

Therefore, we worked with two sets of features to represent patient. On one hand, the complete set of attributes previously shown in Fig. 1, herein called Set 1. This set includes the variables that are used in the 35 rules of the expert system. On the other hand, the six most relevant attributes chosen using CFS shown in Table III, herein called Set 2. It has been observed that four out of the six variables are considered by the expert to be maximum weight (see Table I), i.e. the most relevant ones to determine an induction; while the two remaining ones bear the second highest weight.

TABLE III. ATTRIBUTES OF THE SET 2 OBTAINED WITH CFS OF THE SET 1

CorrelationAttributeEval	
clinical_picture	reason_previous_caesarean
prom_entrance	previous_caesareans
bishop_score_entrance	previous_vaginal_births

Next we will describe the variables corresponding to Set 2, they are the most relevant to predict the performance.

- Clinical\_picture is related to the conditions of the fetus, the mother or both, is a nominal variable whose values are: 1) chronologically prolonged gestation (CPG), 2) premature rupture of membrane

(PRM), 3) intrauterine growth restriction (IUGR), 4) small fetus for gestational age (SGA), 5) oligoamnios -decrease of the amount of amniotic fluid-, 6) altered fetal well-being (AFWB), 7) hypertensive disorders during the pregnancy (HDP) -chronic hypertension, preeclampsia, eclampsia-, 8) diabetes, 9) maternal medical pathology (MMP), 10) other and 11) NA.

- Prom\_entrance is a binary variable that indicates whether the patient entered for premature rupture of the membrane.
- Bishop\_score\_entrance is the patient's value at the time of entry. It is a pre-labor scoring system to assist in predicting whether induction of labor will be required. The duration of labor is inversely correlated with the Bishop score; a score that exceeds 8 describes the patient most likely to achieve a successful vaginal birth. Bishop scores of less than 6 usually require that a cervical ripening method be used before other methods.
- Reason\_previous\_caesarean is a nominal variable which values refers to 1) risk of fetal well-being (RFWB), 2) induction of labor failure (IOLF), 3) No parturition progress (NPP), 4) pelvic-cephalic disproportion (PCD), 5) breech birth (BB) 6) other and 7) NA.
- Previous\_caesareans is a binary variable which indicates wether the patient had previous caesareans.
- Previous\_vaginal\_births indicates the number of previous vaginal birth.

With the purpose of improving the error obtained with the reference model (expert system), we applied the several classification algorithms to both sets of attributes, Set 1 and Set 2. The results are collected in Table IV and Table V, respectively.

Table IV shows that the results are better than those obtained with the reference model. We obtained the best result with Neural Network reaching a classification error of 26.90%.

TABLE IV. RESULTS USING CLASSIFICATION ALGORITHMS (SET 1)

SET 1					
Algorithm	ErrClassif (%)	Precision	Recall	F-measure	AUC
Naïve Bayes	34.96	0.683	0.650	0.659	0.808
Decision tree	27.99	0.693	0.720	0.698	0.799
Neural Network	26.90	0.721	0.731	0.725	0.844
SVM	31.17	0.658	0.688	0.657	0.718
Random forest	27.13	0.709	0.729	0.714	0.848

Results obtained with Set 2 (Table V) show an improvement in respect of Set 1. We have obtained the best result with the same methodology, Neural Network (25.16% classification error).

TABLE V. RESULTS USING CLASSIFICATION ALGORITHMS (SET 2)

SET 2					
Algorithm	ErrClassif (%)	Precision	Recall	F-measure	AUC
Naïve Bayes	29.09	0.696	0.709	0.700	0.832
Decision tree	27.38	0.692	0.726	0.683	0.789
Neural Network	25.16	0.736	0.748	0.738	0.867
SVM	26.59	0.712	0.734	0.702	0.762
Random forest	27.79	0.693	0.722	0.694	0.836

The worst results were obtained with the Naïve Bayes algorithm in both the two subsets. This leads us to speculate that the attributes comprising the two subsets are not independent which causes the worst results in relation to the proven methods. Fig. 3 shows that in general, better results are obtained with less complex models, according to the principle of Occam's razor.

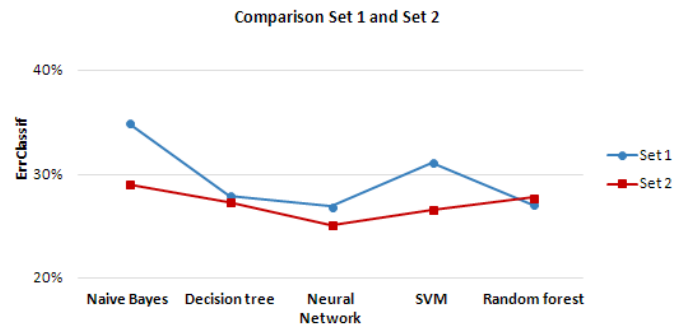


Fig. 3. Comparison between the classification error of Set 1 and Set 2 for the algorithms used.

### C. Explanation of Classification

In this subsection we provide a system to explain the classification obtained from the machine learning models. An algorithm has been implemented to provide transparency to the Neural Network and SVM models, both considered as black box [7], [8].

This system provides explanations to predictions of the instances. Afterwards, explanations are averaged to obtain the contributions of each value (or range of values) to a specific attribute and, in turn, to obtain the global contributions of each attribute to class prediction. For this process, we considered positive and negative contributions independently; otherwise, the contribution of a value or an attribute may be almost non-existent whereas it is very influential in both ways.

In order to simplify the analysis and visualization of results, we tested with the Set 2 selecting only four out of six attributes. Therefore, the dataset used is composed of 10,487 instances including both numerical and nominal attributes: clinical\_picture, prom\_entrance, bishop\_score\_entrance and reason\_previous\_caesarean. The object variable (class) is the decision chosen before the labour process starts and, as it has been stated in the paper, it may take three possible values: *No induction*, *Induction* or *Caesarea*. On the following figures they will be referred to as class 1, 2 and 3, respectively.

Fig. 4 depicts the global contributions of the four selected attributes for Neural Network model. It can be seen that the attributes specialize

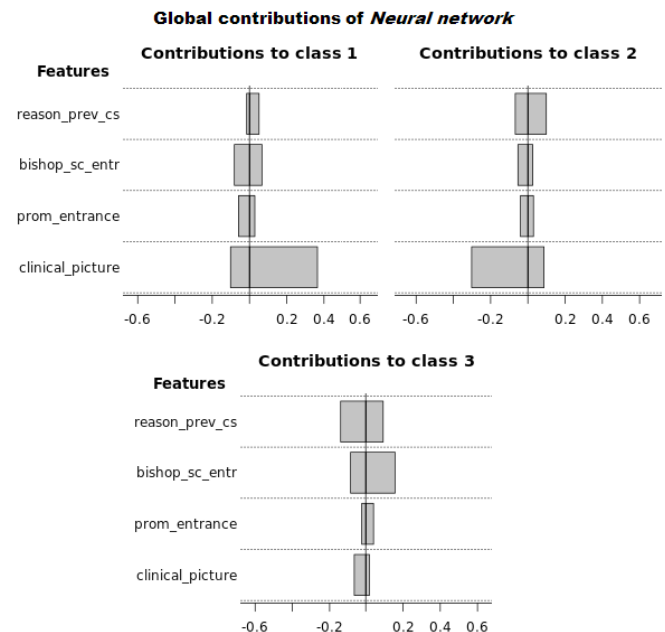


Fig. 4. Global contributions of reason\_previous\_caesarean, bishop\_score\_entrance, prom\_entrance, clinical\_picture of Neural Network to prediction of each class.

in the target. In particular, the attribute `clinical_picture` influences the prediction of the target variables `No induction` or `Induction`, but obviously in a different sense, i.e. in a positive way for `No Induction` and negatively for `Induction`. This attribute corresponds to an indicator that works as a support for the expert to determine whether or not to perform the induction of the labour. As it was stated above, this variable is one of the most relevant ones for the decision which agreed with the experts (highest weight in Table I). This reasoning is supported by the image on the top right in Fig. 4, which includes the contributions to predict class 2 (`Induction`), where `clinical_picture` affects negatively and obviously for the class 1 (`No Induction`) the influence is positive. For class 3 (`Caesarea`), it may observe that all attributes are influential and, although contributions are low, `bishop_score_entrance` stand out with positive values.

Fig. 5 shows the contributions of the Neural Network model for `clinical_picture` (top panel) and `bishop_score_entrance` (bottom panel), attribute differentiated by value segments. At the top of each graph is depicted the average contribution, both positive and negative. For the variable `clinical_picture`, only `No Induction` and `Induction` classes are represented. It may see that on average this variable has a positive influence for `No Induction` and in a negative way for `Induction`. However the contribution depends on the input of the variable. For the case of `No Induction` class is positive the presence of CPG, diabetes, and the influence is negative for most of the conditions related with the fetus, i.e. PRM, IUGR, SGA, oligoamnios, AFWB, HDP. The opposite occurs with the `Induction` class.

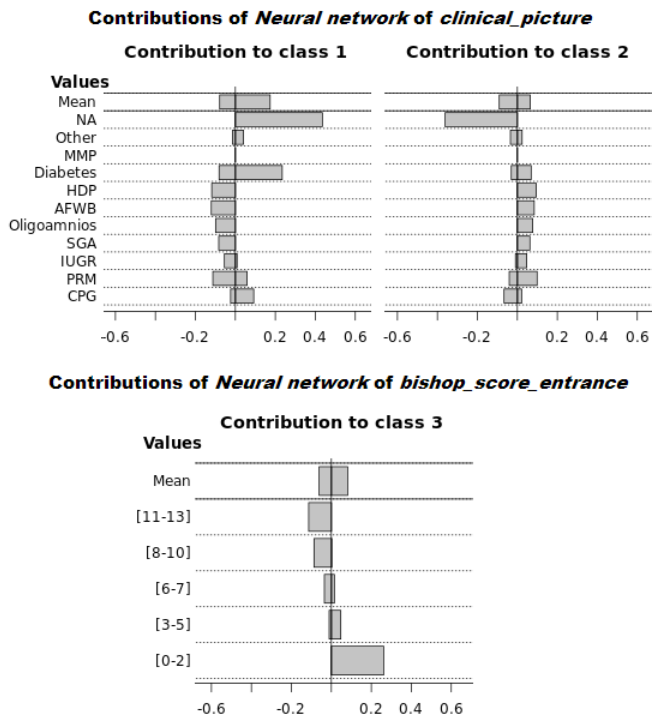


Fig. 5. Contributions of the values of `clinical_picture` (top panel), `bishop_score_entrance` (bottom panel) using Neural Network to prediction. The first bar Mean represents the average of all ranges of values.

However, the image at the bottom panel shows that the low values of the `bishop_score_entrance` feature are the most influential in order to decide `Caesarea`. It matches one of the decision rules provided ('If `bishop_score_entrance` ≤ 6 then `Caesarea`'). The work [21] suggests that a score of 5 or less indicates that the labour is unlikely to start without induction. This agrees with the results show on the bottom panel of Fig. 5.

On the other hand, we included global contributions using SVM in

Fig. 6. In this case, despite including lower contributions than using Neural Network, we observe that for `No induction` and `Caesarea` classes the influential attribute is the same, which is, again `clinical_picture` in `No induction` class and `bishop_score_entrance` in `Caesarea` class. On the contrary, SVM and Neural Network disagreed with the prediction of `Induction` class. In SVM case, it is `prom_entrance` the most influence variable, despite the fact that `reason_previous_caesarean` has also positive values.

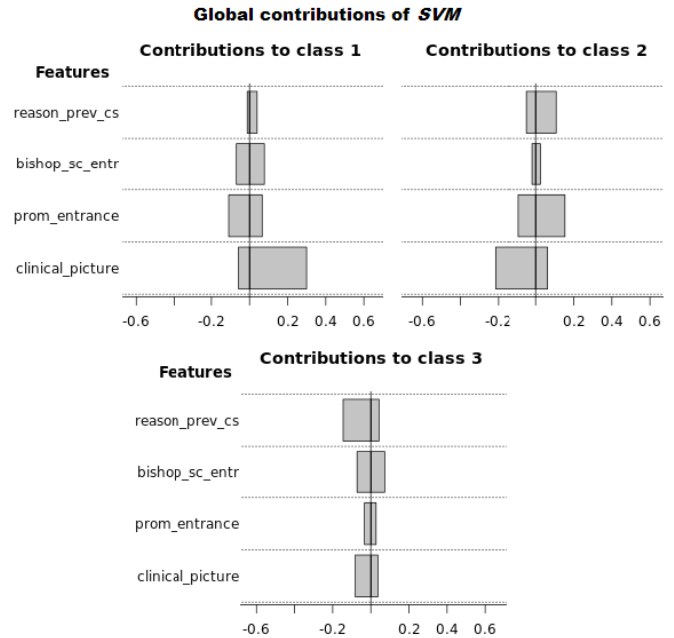


Fig. 6. Global contributions of `reason_previous_caesarean`, `bishop_score_entrance`, `prom_entrance`, `clinical_picture` of SVM to prediction of each class.

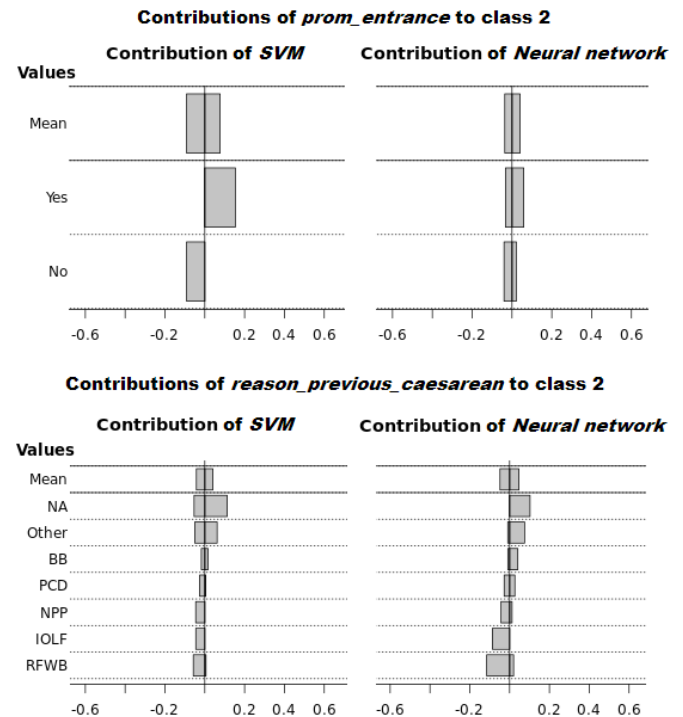


Fig. 7. Contributions of the values of `prom_entrance` (upper panel), `reason_previous_caesarean` (lower panel) for both SVM and the Neural Network with respect to the prediction of the `Induction` class. The first bar Mean represents the average of all ranges of values.

With the aim of determining which attribute of the two previously discussed, `prom_entrance` or `reason_previous_caesarean`, is more relevant in the prediction, Fig. 7 shows the explanations with Neural Network in the graphs on the right hand side and with SVM in the graphs on the left hand side.

Regarding `prom_entrance`, the graphs on the top in Fig. 7 show that most instances (all of them in SVM) of *Induction* class are confined in the option Yes. Nonetheless, low contributions in the case of Neural Network may indicate this attribute is not really influential in order to determine this class or, on the contrary, the model might not have captured the domain of the problem properly, as in this case we obtained `reason_previous_caesarean`. However, the fact that there are both negative and positive contributions for the two ranges of values reveals the extreme complexity of the problem due to the influence of many other factors in the decision.

The graphs at the bottom show that `reason_previous_caesarean` is also an influential attribute and it is more probable that in both algorithms the *Induction* class is assigned to an instance with values BB, other, NA. Most instances take NA value, which leads to think that, although the knowledge of the expert tells us that `reason_previous_caesarean` variable is determinant for a cesarean section, in the absence of this information for a patient (NA value), the prediction of cesarean section is discarded. The prediction is leaning towards any of the two others possible results, in this case we are showing the case of *Induction*.

All previous contributions prove what the model has learned from training. There is no direct reference to the real distribution of instances in the space of attributes. In some of the situations depicted, the trained models have captured the real domain of the problem properly and contributions, besides explaining how the model works, reflect this field quite accurately.

#### IV. CONCLUSION AND FUTURE WORK

In this paper we have designed a system to exploit information compiled in the Electronic Medical Records about pregnancies women. The goal is to extract value out of data.

Five principles were pursued in this project: (1) accuracy, models respond correctly, (2) interpretability, responds to the question why a particular action is recommended, (3) actionable, to reduce patient risk, (4) credible, consistent with what is known in the clinical literature and (5) robust, capable of adopting changes over time and population.

A computer system was built which incorporates two divergent principles. Firstly, a Clinical Decision Support System based on decision rules provided by a panel of experts in Obstetrics and secondly, methodology based on learning techniques, Big Data and algorithms was implemented.

Finally, we have verified that a small number of variables is sufficient to obtain robust models. In addition, attempts have been made to obtain transparency in models or algorithms difficult to interpret and thus be able to obtain new rules of behavior.

Experimental results with this dataset indicate that, if there is no reason why the expert might recommend induction, the result should be *No induction*. For *Induction*, it is required that the patient had not had any previous caesarean, or that there has been a premature rupture of membranes (`prom`) when was admitted. In order for a *Caesarea* to be determined, typically the Bishop score must be less than 6. These explanations make these models more transparent and may help complement knowledge or discover relationships among data that were so far unnoticed.

The implemented system has proved to be of interest and useful to the expert in decision making. It is not only a new tool for access

and validation of clinical information, but a new line of work has been created, where the application developed can be used in clinical practice in real time by expert medical personnel, hoping to improve their results. The applied methodology can be extrapolated to any other branch of Medicine.

#### ACKNOWLEDGMENT

The authors are extremely grateful to *Hospital Universitario de Fuenlabrada* and *Instituto de Ingeniería del Conocimiento*, Madrid. This work was partially supported by the Spanish CICYT project TIN2013-42351-P.

#### REFERENCES

- [1] C. J. Verhoeven et al., "Validation of models that predict Cesarean section after induction of labor," *Ultrasound Obstet Gynecol*, vol. 34, no. 3, pp. 316–321, Sep. 2009.
- [2] N. Baños, F. Migliorelli, E. Posadas, J. Ferreri, and M. Palacio, "Definition of Failed Induction of Labor and Its Predictive Factors: Two Unsolved Issues of an Everyday Clinical Situation," *Fetal Diagn Ther*, vol. 38, no. 3, pp. 161–169, Jun. 2015.
- [3] F. Almachi and S. Lissette, "Eficacia y seguridad de la inducción del trabajo de parto con misoprostol en pacientes con embarazo a término con indicación de inducción y cérvix desfavorable," *Escuela de Obstetricia - Universidad de Guayaquil*, 2013.
- [4] E. Mozurkewich, J. Chilimigras, E. Koepke, K. Keeton, and V. King, "Indications for induction of labour: a best-evidence review," *BJOG: An International Journal of Obstetrics & Gynaecology*, vol. 116, no. 5, pp. 626–636, Apr. 2009.
- [5] American College of Obstetricians and Gynecologists, "ACOG Practice Bulletin: Clinical Management Guidelines for Obstetrician-Gynecologists," *Obstet Gynecol*, vol. 114 (2 Part 1), no. 107, pp. 386–397, 2009.
- [6] D. F. Lobach et al., "Increasing Complexity in Rule-Based Clinical Decision Support: The Symptom Assessment and Management Intervention," *JMIR Medical Informatics*, vol. 4, no. 4, pp. e36, Nov. 2016.
- [7] J. D. Sampedro, "Estudio y aplicación de técnicas de aprendizaje automático orientadas al ámbito médico: estimación y explicación de predicciones individuales," *EPS - UAM*, 2012.
- [8] E. Strumbelj and I. Kononenko, "An efficient explanation of individual classifications using game theory," *Journal of Machine Learning Research*, vol. 11, pp. 1–18, 2010.
- [9] "Cerner Selene Hospitales," 2015. [Online]. Available: <http://www.cerner.com/Soluciones/Sistemas-de-Informacion/Cerner-Selene/Selene-Hospitales/?LangType=1034> [Accessed 18 Feb. 2016].
- [10] "CLIPS, NASA Johnson Space Center Std.," 1984. [Online]. Available: <http://clipsrules.sourceforge.net/> [Accessed 7 Apr. 2016].
- [11] B. Buchanan and E. Shortliffe, "Rule-based expert systems: the MYCIN experiments of the Stanford heuristic programming project," *Addison-Wesley*, NY, vol. 18, pp. 14–3, 1984.
- [12] A. J. Gonzalez and D. D. Dankel, *The engineering of knowledge-based systems: theory and practice*. Prentice-Hall, Inc., 1993.
- [13] I. Kononenko, "Estimating attributes: analysis and extensions of RELIEF," in *Machine Learning: ECML-94*. Springer, 1994, pp. 171–182.
- [14] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," *Journal of bioinformatics and computational biology*, vol. 3, no. 02, pp. 185–205, Apr. 2005.
- [15] A. Roy et al., "A comparative study of feature ranking methods in recognition of handwritten numerals," in *Artificial Intelligence and Evolutionary Algorithms in Engineering Systems*, Springer, New Delhi, 2015, pp. 473–479.
- [16] M. A. Hall, "Correlation-based feature selection for machine learning," *The University of Waikato*, 1999.
- [17] M. Hall et al., "The WEKA data mining software: an update," *ACM SIGKDD explorations newsletter*, vol. 11, no. 1, pp. 10–18, Nov. 2009.
- [18] I. Kononenko, "Machine learning for medical diagnosis: history, state of

the art and perspective,” Artificial Intelligence in medicine, vol. 23, no. 1, pp. 89-109, Aug. 2001.

- [19] D. A. Morales et al., “Bayesian classification for the selection of in vitro human embryos using morphological and clinical data,” Computer methods and programs in biomedicine, vol. 90, no. 2, pp. 104-116, 2008.
- [20] L. Breiman, “Random forests,” Machine learning, vol. 45, no. 1, pp. 5- 32, Oct. 2001.
- [21] J. Tenore, “Methods for cervical ripening and induction of labor,” American family physician, vol. 67, no. 10, pp. 2123-2128, May. 2003.



Cristina Pruenza García-Hinojosa

Cristina Pruenza García-Hinojosa has a Degree in Mathematics and Computer Engineering from Universidad Autónoma de Madrid (2009-2014) and she has got a Master’s Degree in Computer Engineering and a Master’s Degree in ICT Research and Innovation in Computational Intelligence at the Polytechnic School, Universidad Autónoma de Madrid (2014-2016). She is currently data analyst at the Instituto de Ingeniería del Conocimiento (IIC-UAM) in the Health and Energy Predictive Analytics group, where she is currently working in the following innovation lines: Big Data, modelling, data analysis, evaluation of wind production forecasting, e- Health among others. Pruenza García-Hinojosa has participated in several projects related to e-Health and in the ADNI project research on Alzheimer’s and some possible tools to treat MRIs.



María Teulón

María Teulón is Doctor of Medicine and Surgery, Gynecology and Obstetrics Specialist and bi-Master Health System and Hospital Management. At present, she is the Coordinator of Gynecology and Obstetrics Dept at, University Fuenlabrada Hospital, Madrid (Spain). She has spent most of her career as a specialist in Obstetrics and Fetal Medicine. She has participated in several research projects in this interest area, highlighting: Fetal Mvision-MIT Project (Placental RNM) or Standards of Appropriateness Use of Caesarean Section (support tool in clinical decision making).



Luis Lechuga-Suárez

Luis Lechuga-Suárez received the B.Sc. degree in physics from the Universidad Complutense, Madrid, Spain, in 1988. He was Certified in HL7 2.x by the International Standards Organization HL7.ORG. After developing his career in the IT Department at several companies like Telefonica Spain and Madrid Stock Exchange, he has been responsible for Systems, Communications, and Integrations at Fuenlabrada University Hospital, Madrid, Spain, since 2003, where he also leads the ICT R&D section. He has participated in the implementation, standardization, interoperability and Big Data projects of electronic health records.



Julia Díaz

Julia Díaz is Ms Degree in Mathematics, PhD in Computer Science both from Universidad Autónoma de Madrid (UAM-Spain) and General Management Program from IESE-Universidad de Navarra (Spain). At present she is Senior Innovation Manager in a private R&D+i institution named Instituto de Ingeniería del Conocimiento (IIC-UAM) dedicated to extracting knowledge on the basis of high volumes of heterogeneous data (Big Data) and optimizing business processes in areas such as healthcare and energy. She also is Part Time PhD Professor in Computer Sciences in the UAM and Professor in Big Data & Data Sciences Master in UAM and ESADE.



Ana González

Ana González is Ms Degree in Chemistry and PhD in Computer Science both from Universidad Autónoma de Madrid (UAM-Spain). At present she is Lecturer at the Department of Computer Science of the Escuela Politécnica Superior (EPS) at UAM. She is also a member of the Automatic Learning Research Group at the EPS which performs research on Machine Learning and Data Mining applications. Dr. González is also collaborating with a private R&D+i institution named Instituto de Ingeniería del Conocimiento (IIC-UAM) in the area of healthcare.