Master in Artificial Intelligence

Master Thesis

A comprehensive probabilistic model for the embryo selection problem

Author:
Olga Valls Murcia

Supervisors:
Jerónimo Hernández-González
Jesús Cerquides

Tutor:
Javier Béjar

Facultat d'Informàtica de Barcelona (FIB)
Universitat Politècnica de Catalunya (UPC)

Facultat de Matemàtiques i Informàtica de Barcelona (UB)
Universitat de Barcelona (UB)

Escola Tècnica Superior d'Enginyeria (URV)
Universitat Rovira i Virgili (URV)

January 19, 2021

Abstract

Assisted reproductive technologies (ARTs) are a set of invasive medical techniques that attempt to induce a pregnancy. In vitro fertilization (IVF) is the most common and effective type of ART. Embryo selection is a difficult and complex task. There is a morphological evaluation criteria and a categorization into scales for each of the various embryo stages. From these results, clinicians have to select which embryos to transfer, as the clinical procedure can produce excess embryos. The transferred embryos have to be carefully selected among the ones that show best quality according to this morphological classification, as the aim of the process is to achieve a pregnancy. In this project, we present a novel probabilistic graphical model that, for the first time, accounts for the uncertainty that represents all the unknown factors that can drive to a failure even though all the components that take part in the ART process seem to be favorable. In an ARTs' dataset it is not always possible to know which embryo was implanted. Among others, this uncertainty source forces us to use an EM strategy, as well as the consideration of hidden variables in our model. The experiments carried out show that much more information can be obtained from this type of model than from previous simpler approaches. The database for this work have been collected by the Unit of Assisted Reproduction of the Hospital Donostia (Spain) throughout 18 months (January 2013–July 2014) where 604 patients participated in the IVF-ICSI program compiling a total number of 3125 embryos.

Contents

1 Introduction

1.1 Approximation to the embryo selection problem

Assisted reproductive technologies (ARTs) are a set of invasive medical techniques that attempt to induce a pregnancy. In vitro fertilization (IVF) is the most common and effective type of ART. Each trial of a treatment is known as a cycle. The woman must follow a treatment of ovarian stimulation for some weeks to induce the development of multiple follicles with a large number of oocytes. Those oocytes are retrieved, and the matures ones are fertilized. The resulting embryos are cultured for several days and clinicians must select which embryos will be transferred to the woman's uterus[9].

Embryo selection is a difficult and complex task. There is a morphological evaluation criteria and a categorization into scales for each of the various embryo stages. The first decision to be made in dealing with embryo quality is the selection of the working tool. The most common method is the use of morphological parameters, although other criteria, such as the study of aneuploides, are proving to be also effective[9]. The second decision is choosing the developmental stages for these morphological parameters. The most common practice is to use the second or third day of culture for the embryo or the fifth or sixth for the bastocyst stage.

Once all the morphological data has been collected the final embryo classification needs to be decided. The most extended classification option is by category, although it also exists a scoring measurement. From these results, clinicians have to select which embryos to transfer, as the clinical procedure can produce excess embryos. The transferred embryos have to be carefully selected among the ones that show best quality according to this morphological classification, as the aim of the process is to achieve a pregnancy. Multiple transference is considered risky for the woman as well as for the developing fetus(es). Besides, there exist legal restrictions limiting the maximum number of transferred embryos (in Spanish law this is limited to three).

This process is physically and psychologically tough, specially for women, and does not guarantee that any of the transferred embryos implant in the uterus. The latest report from the Spanish Society of Fertilization (SEF), published in 2018, shows that only 35.6% of the IVF processes end up in pregnancy[11].

The Artificial Intelligence approach to ARTs uses Machine Learning techniques (see Section 2) to try to solve the embryo selection problem, as well as other problems presented in IVF. In the case of the embryo selection, different features of the embryos at different stages of the culture, and some characteristics of the cycle, are used to train different classifiers and predict whether a transferred embryo will succeed in implantation. This would relieve some of the stress to the couples wanting to become parents, as the predictions would assist clinicians in the selection of the embryos that most likely would lead to pregnancy.

All these techniques take embryos and cycles' characteristics, and the medical treatment that has been followed, as responsible for the success or failure of an IVF process. Yet, there is a recurrent problem in assisted reproduction units: there exist unknown factors that affect the success of an ART cycle. Many times, the uterus of the patient is healthy and has received the proper ovarian stimulation, and the embryos have high quality, nevertheless, none of the transferred embryos implant.

In this project, we present a novel probabilistic graphical model (see Section 4.4) that, for the first time, accounts for the uncertainty that represents all the unknown factors that can drive to a failure even though all the components that take part in the ART process seem to be favorable. In an ARTs' dataset it is not always possible to know which embryo was implanted. Among others, this uncertainty source forces us to use an EM strategy, as well as the consideration of hidden variables in our model. The experiments carried out show that much more information can be obtained from this type of model than from previous simpler approaches.

We have developed a learning algorithm specifically for this model. It is based on the Expectation-Maximization (EM) strategy and uses two probabilistic classifiers to approximate the probabilistic distribution of the quality of embryos and cycles given the respective descriptive features. This is a comprehensive model, as cycles, embryos and also those unknown factors (represented by a Bernoulli distribution) are taken into account.

This model adds more detail describing reality as it shows the likelihood of a successful result when all the components are favorable. It represents the relationship between the cycles, the embryos belonging to the cycles, and this uncertainty factor that affects the outcome and helps define the behaviour of the whole system.

1.2 Motivation

Embryo implantation is a complex process involving maternal hormonal changes, immune responses and maturational events in the embryo. A pregnancy could fail when these events are not synchronized[**rif04**]. Despite the great improvements in ovarian stimulation protocols and fertilization procedures, implantation rates per embryo remain at approximately 15% and many couples are still left frustrated following multiple failed attempts[**rif02**]. Recurrent implantation failure (RIF) is a condition resulting from repetitive unsuccessful cycles of IVF or intracytoplasmic sperm injection (ICSI) treatment [**rif03**], and it is the clearest evidence that there exist still unknown factors that affect the success of an ART cycle.

A percentage of 35.6% of success in an IVF process can be seen as hope for couples who expect to have a baby, as well as a failure for scientists who try to perfect a technique. The unknown factors that affect those processes are on continuous study, and integrating its representation in a model adds information

about the data the clinicians are working with. Knowing how a given set of cycles and embryos behaves is key for decision, not only for embryo selection but, for instance, the study of different morphological characteristics or modification of the ovarian stimulation treatment.

The ASEBIR committee for "the definition of morphological evaluation criteria and their categorization, from Oocyte to Blastocyst"[9] was created in 2004 to try to respond to the need for the unification of embryo evaluation criteria. Its goal was to set out a proposal for morphological evaluation criteria and their categorization into scales for each of the various embryo stages.

This categorization is a grading system that divides the expected implantation potential of the embryo into 4 categories from A (an embryo with optimal quality and the best implantation potential) to D (a poor-quality embryo with a low chance of implantation). Assigning an embryo to a particular category depends on its morphological parameters. The final selection of the embryos is performed taking into account this categorization and the number of embryos the clinicians decide to transfer. The low percentage of success in IVF shows that nowadays clinicians have not found a proper algorithm for such task.

The model proposed, that shows the likelihood that the transfer of any selected embryo willing to implant to a healthy uterus results in pregnancy, can be used as a support for the clinicians to select the number of embryos to transfer depending on how high this likelihood is. A low likelihood would tell that, even with perfect conditions, that would be difficult for a high quality embryo to implant. Thus, likelihood would be something to consider.

1.3 Structure of the document

This work starts introducing the state of the art in this subject (see section 2) where different approaches to solve the embryo selection problem are described, along with other problems related to IVF processes.

Afterwards a description of the features for each of datasets used to create the model are presented (see section 3). There is a dataset for cycles and a related one for embryos, both provided by the University of the Basque Country UPV/EHU. Not all the data provided is used and some transformations for some of the features are performed.

The work continues with a broader description of the proposed model and the learning technique (see section 4) used to predict the probability of implantation for an embryo selected to be transferred. First, an introduction to the probabilistic model for IVF and to the general probability model for ARTs is presented. An ART can carry out more than one transfer per cycle, although this work only consider one transfer per cycle. In the graphical description of the model the reader can see how the different variables of this system interact, discovering the role of the latent variables in the whole system, the observed variables and also the hyper-parameters for each of the classifiers and the Bernoulli distribution.

Next, the derivation of the EM algorithm used to learn our particular model is presented which provides a measure to evaluate how is the data in the system in terms of chances to succeed in the process in perfect conditions.

Some experimentation has been performed (see section 5) where the model proposed is tested with different probabilistic classifiers to be able to see which one is able to define the reality of the datasets more accurately, and is also able to perform better at predicting the success of an embryo selected to be transferred. As the EM algorithm needs some iterations to perform, and it is tested with different probabilistic classifiers, the evolution of this convergence has been studied with different measures. The loss and the recall, along with the probability of implanting given perfect conditions is also computed for all of these probabilistic classifiers. In order to see how good the proposed model would perform, it is compared with a baseline model which consists only in a probabilistic model. The latest also predicts the probability of implantation for an embryo selected to be transferred, and the AUC-ROC score is computed for each of the probabilistic classifiers.

As a closure for this work, some conclusions (see section 6) are discussed and ideas for future work (see section 6.1) presented. Machine learning methodologies applied to heterogeneous sets of data, and a proper modelling of the system that this data is representing, can assist clinicians in decision making and improve performance of IVF processes.

2 State of the art

Ranking algorithms[**paperdeu**], statistical models, ensemble techniques, neural networks, classification and regression tree, and regression analysis, discriminant analysis and case based reasoning systems are some of the IA techniques used in in-vitro fertilization (IVF)[**papernou**]. However, there is not a single AI method that is useful for solving a particular problem and fits in all cases[**papercinc**]. We have to consider the data and labels available, and the output desired. It is important to carefully analyse the dataset and model the problem properly.

AI methods are being investigated as a promising means for improving embryo selection and predicting implantation and pregnancy outcomes. As happend in any field, there are studies that wonder about the inclusion of "machines" to perform such tasks[5]. On the other hand traditional morphokinetic grading by trained clinicians can be subjective and variable, but other complementary techniques, such as time-lapse imaging, which would be more objective, require costly equipment, and they have not demonstrated sufficient predictive ability[**papercatorze**].

These objective, standardized and efficient tools for evaluating human embryos are demanded in laboratories, not only for embryo selection but for other needs, such as assessing patient reproductive potential and individualizing gonadotropin stimulation protocols[8]

Images can be used to feed a neural network. Specifically the texture descriptors from a given image, based on morphological analysis of the embryos[**paperset**]. The features incorporated in the texture of images are not usually perceived by the human eye, and might be very useful for the recognition of viable embryos. There are works[7] that are capable to classify, from images, embryos, pronuclei or oocytes suitable for procreation.

Deep Neural Networks can be applied to combine spatial and temporal information to predicting blastocits quality, given the information provided by time-lapse imaging[**paperquatre**]. A Convolutional Neural Network is trained to predict inner cell mass (ICM) and trophectoderm (TE) grades from a single image frame, and a Recurrent Neural Network is used to incorporate temporal information for multiple frames.

A Bayesian Network model[**paperdos**], based on EU assumption, can improve parameter estimates. The EU model estimates the probability of pregnancy after the transfer of a single embryo, assuming the independence of viability and receptivity; when dealing with the transfer of multiple embryos, each embryo is assumed to implant independently from the others. It takes a reduced subset of feature variables related to embryo morphology and clinical data of patients, which is the same that expert embryologists take into account in normal practice[**papersis**]. Different Bayesian classifiers take into account diverse dependencies between variables; using information extracted from embryo images[3], the viability to succeed in implantation on woman's uterus can be predicted.

In a multi-variate data analysis, all the collected features, describing cycles and embryos, are considered[**papertres**]. As mentioned before, it is important to create an appropriate modeling of the problem to improve Machine Learning techniques for embryo selection. Learning from label proportions considers all the availability data, as there are also embryos whose fate can not be certainly established; those are the cases where the number of embryos implanted are less than the number of embryos transferred. This is incomplete information that is used to train the classification models. Once again Bayesian classifiers are used, as this is a model that can be calibrated to balance the contribution of each predictive feature.

The Area Under the ROC curve is a measure that is used in some works[1] [**papervintiquatre**] to evaluate Machine Learning Algorithms. It is a good way of visualizing a classifier's performance in order to select a suitable operating point or decision threshold.

3 Data

The database have been collected by the Unit of Assisted Reproduction of the Hospital Donostia (Spain) throughout 18 months (January 2013–July 2014). 604 patients participated in the IVF-ICSI program compiling a total number of 3125 embryos.

It is composed of two spreadsheets, one for cycles and another for embryos, related by a one-to-n relationship: for one cycle, the n embryos transferred in that procedure. Each cycle is described by 25 features, including characteristics of the patient, stimulation treatment and statistics of the associated embryos. Each embryo is described by 20 features: oocyte/embryonic morphological characteristics and quality grades.

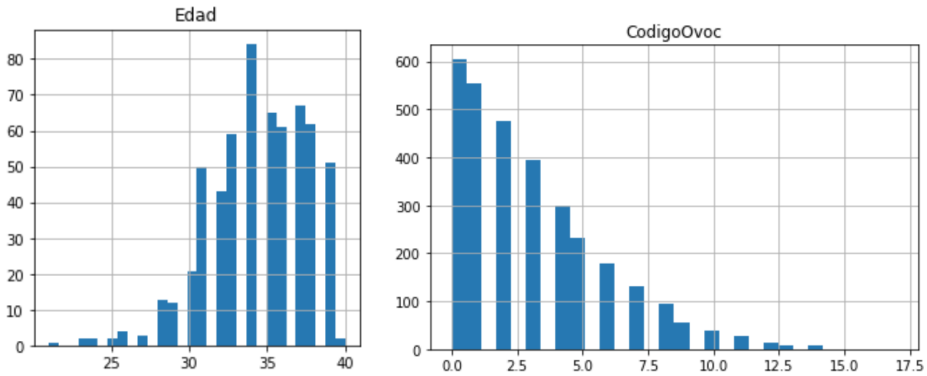
Both datasets have a feature, TasaExito, that determines the percentage of success of the cycle or the embryo. In the case of the cycles, this feature indicates the percentage of transferred embryos belonging to that cycle that have succeed in implantation. In the case of the embryos, this feature indicates whether that transferred embryo has been implanted or if it belongs to a selection of transferred embryos where some of them have been implanted and some do not.

As detailed in Table 1, 412 cycles failed to induce pregnancy (839 embryos), and only in 57 cycles **all** the transferred embryos resulted implanted (108 embryos). In the remaining 135 cycles, only a subset of the transferred embryos were implanted (307 embryos) but it is not possible to know which specific embryos are the ones implanted. There are, still, 1871 embryos not selected for transfer, which are low quality embryos or surplus, due to the maximum number of embryos that the law allows to transfer (in Spain this number is limited to three).

Failed pregnancy	412 cycles	839 embryos	
All embryos implanted	57 cycles	108 embryos	
Some embryos implanted	135 cycles	307 embryos	
	604 cycles	1254 embryos	
		1871 embryos	not selected for transfer
		3125 embryos	in the dataset

Table 1: Success number of cycles and related number of embryos

According to this database, women that undergo an IVF process ranges from 21 to 40 years old, being the average 34.43 (see Figure 1a). The average number of embryos in a cycle is 5 although a variability from 1 to 18 is seen (see Figure 1b).



(a) Range of women's ages (b) Number of embryos in cycles

Figure 1: Range of ages from women undergoing an IVF process, and count for the number of embryos in cycles

More important than the number of embryos per cycle is the quality of the embryos. The ASEBIR grading system assign each embryo a category based on different morphological qualities. These are the categories considered:

- Category A: An embryo with optimal quality and the best implantation potential.
- Category B: An embryo of good quality and a high implantation potential.
- Category C: An acceptable embryo with an average chance of implantation.
- Category D: A poor-quality embryo with a low chance of implantation.

Figure 2 shows that more than half of the embryos belong to the lowest quality categories (37.056% for category C 22.464% for category D). The best quality embryos represent a 24.352% for category A and 16.032% for category B. The remaining 0.096% are embryos without assigned category.

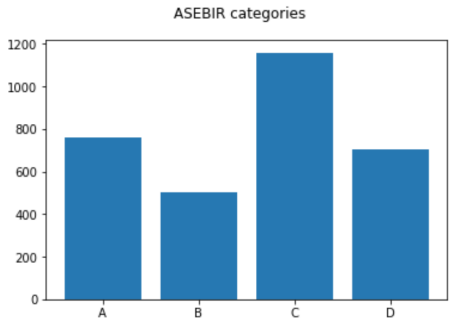


Figure 2: Count for ASEBIR categories

3.1 Dataset of cycles

The features collected for each of ART cycles are shown in the following table (see Table 2):

Feature	Description	Possible values
Codigo	Code of the cycle	*numeric*
TEsteril	Time since infertility was detected	*numeric*
Features related to Female		
Indicac	Indication of the cycle	endometriosis, fracasoia, tubarico, masculino, mixto, otros, desconocido
Edad	Age of the patient	*numeric*
IMC	Body mass index	*numeric*
EmbPrev	Whether the patient has got previous pregnancies	No, Yes
AboPrev	Whether the patient has got previous miscarriages	No, Yes
FSH	Quantity of follicle-stimulating hormone	*numeric*
Ciclosprevios	Number of previously undergone ART cycles	*numeric*
AMH	Quantity of anti-mullerian hormone	*numeric*
folAntral	Number of antral follicles	*numeric*
E2	Quantity of estradiol	*numeric*
P4	Quantity of progesterone	*numeric*
lEnd	Endometrial thickness	*numeric*
Features related to Male		
caSemen	Quality of the semen	A, N, O, OA, OAT
REM	Total pregressive sperm recovery	*numeric*
Features related to Stimulation		
Protocol	Stimulation protocol	PC+Agon, PC+Ant, PL
Estimul	Stimulation treatment	FSHrec, FSHrec+hMG, FSHur, hMG, FSHur+hMG, FSH+Lhrec
dEst	Number of days of stimulation	*numeric*
unidFSH	Units of FSH	*numeric*
unidLH	Units of LH	*numeric*
Features related to Embryos		
nEmbObten	Number of embryos	*numeric*
TasaFertil	Number of embryos / Number of mature oocytes (MII state)	*numeric*
nEmbTrans	Number of transferred embryos	*numeric*
TasaExito	Percentage of success in the transfer, given all the embryos belonging to the same cycle	*numeric*

Table 2: Features collected for each ART cycle

3.2 Dataset of embryos

The features collected for each of the embryos are shown in the following table (see Table 3):

Feature	**Description**	**Possible values**
CodigoCiclo	Code of the cycle the embryo belongs to	*numeric*
CodigoOvoc	Code of the embryo	*numeric*
Tecnica	Fertilization technique	IVF, ICSI
Features related to Oocytes		
Vac	Presence of vacuoles	No, Few, Many
REL	Presence of smooth endoplasmic reticulum clusters	No, Yes
Epv	Description of the perivitelline space	Normal, Augmentado
CP	Description of the first polar body	Normal, Anormal
PN	Tesarik and Greco's pronuclear grade	*numeric*
Features related to Oocytes at D+1		
CP.1	Number of polar bodies	*numeric*
Z	Scott's pronuclear grade	Z1, Z2, Z3, Z4
Features related to Oocytes at D+2		
nCel+2	Number of cells	*numeric*
frag+2	Percentage of cell fragmentation	*numeric*
simet+2	Are the blastomeres symmetric?	No, Yes
ZP+2	Zona pellucida	Normal, Abnormal
vac+2	Presence of vacuoles	No, Few, Many
multiNuc+2	Presence of multi-nucleation in a cell	No, Yes
CALIDAD+2	ASEBIR quality grade	A, B, C, D
Features related to Embryos		
Transfer	Embryo selected to be transferred	No, Yes
Vitrificado	Surplus' embryos to froze	No, Yes, Surplus
TasaExito	Percentage of success in the transfer, given all the embryos belonging to the same cycle.	*numeric*

Table 3: Features collected for each oocyte/embryo

3.3 Pre-processing

In the following lines a description of how the data stored in both datasets is processed before being used to create the proposed model (see Section 4) is reported.

Both datasets have heterogeneous features and, to create the model, all the categorical features need to be converted into numerical features. There are binary categorical variables can be converted into numerical by assigning 0 to one of the values and 1 to the other one (i.e, No=0 and Si=1, or Anormal=0 and Normal=1). Multi-value categorical variables need to be converted via one-hot encoding, which will produce one new variable for each of the possible categorical values (see Tables 4 and 5).

Features related to Cycles	
Original	**Converted numerical feature**
Binary categorical variables	
EmbPrev	0, 1
AboPrev	0, 1
Multi-value categorical variables	
Indicac	Indicac_endometriosis, Indicac_fracasoia, Indicac_tubarico, Indicac_masculino, Indicac_mixto, Indicac_otros, Indicac_desconocido
caSemen	caSemen_A, caSemen_N, caSemen_O, caSemen_OA, caSemen_OAT
Protocol	Protocol_PC+Agon, Protocol_PC+Ant, Protocol_PC+PL
Estimul	Estimul_FSH+Lhrec, Estimul_FSHrec, Estimul_FSHrec+hMG, Estimul_FSHur, Estimul_FSHur+hMG, Estimul_hMG

Table 4: Categorical features from cycles converted into numerical features

Features related to Embryos	
Original	**Converted numerical feature**
Binary categorical variables	
REL	0, 1
Epv	0, 1
CP	0, 1
Simet+2	0, 1
ZP+2	0, 1
multiNuc+2	0, 1
Multi-value categorical variables	
Tecnica	Tecnica_IVF, Tecnica_ICSI
Vac	Vac_No, Vac_Escasas, Vac_Abundantes
Z	Z_Z1, Z_Z2, Z_Z3, Z_Z4
Vac+2	Vac+2_No, Vac+2_Escasas, Vac+2_Abundantes
Calidad+2	Calidad+2_A, Calidad+2_B, Calidad+2_C, Calidad+2_D

Table 5: Categorical features from embryos converted into numerical features

Even so, some of these variables will not be used as their possible values are already reflected in the value of some other variables (i.e, Indicac_otros and Indicac_desconocido would be activated when the rest of Indicac's variables are not, which would be redundant).

Moreover, general variables have been created to connect embryos with their belonging cycles and the count of embryos transferred and implanted (see Table 8). The use of those variables will also make the use of some features useless. And there also variables that, for some reason, do not provide relevant information in defining the model (see Tables 6 and 7).

The features removed for the dataset of cycles are the following (see Table 6):

Features related to Cycles	
Feature	**Reason why this feature is discarded**
Codigo	The relation between cycles and embryos will be recorded in general variables
AMH	Too skewed variable
nEmbTrans	Information recorded in general variables
TasaExito	Information recorded in general variables
Indicac_desconocido	Redundant
Indicac_otros	Redundant

Table 6: Features discarded from the dataset of cycles

The features removed for the dataset of embryos are the following (see Table 7):

Features related to Embryos	
Feature	**Reason why this feature is discarded**
CodigoCiclo	The relation between cycles and embryos will be recorded in general variables
CodigoOvoc	The relation between cycles and embryos will be recorded in general variables
Transfer	This information will be recorded in general variables
Vitrificado	Feature not relevant
TasaExito	This information will be recorded in general variables
Vac_No	Redundant
Vac+2_Normal	Redundant

Table 7: Features discarded from the dataset of embryos

The variables created to keep a record of the relations of the embryos with its belonging cycles, and used for computation when creating the model, are the following (see Table 8):

Variable	**Description**
num_cycles	number of cycles
num_embryos	number of embryos
embryo_belong_to_cycle	which cycle each embryo belongs to
embryo_was_transfered	whether an embryo was transfered
embryo_was_implanted	whether an embryo was implanted
num_emb_transf_per_cycle	number of embryos transferred per cycle
num_emb_implanted_per_cycle	number of embryos implanted per cycle
cycle_has_embryos	which embryos belong to each cycle
cycle_has_trans_embryos	which are the embryos belonging to one cycle that have been transferred

Table 8: Variables created to keep a record of the relations of the embryos with its belonging cycles

After all these processes the dataset of cycles has 36 features and the dataset of embryos has 24 features. They all are important as they take part in the definition of the model that will show the behaviour of the whole system of cycles, embryos, and the unknown factors that cause an implantation to fail when there are perfect conditions.

The success rate (feature TasaExito) indicates the percentage of transferred embryos, belonging to the same cycle, that have been implanted, and will determine which class an embryo is assigned (0, 1, -1):

- TasaExito == 1: all the embryos selected for transfer have been implanted; class 1 will be assigned.
- TasaExito == 0: none of the embryos selected for transfer have been implanted; class 0 will be assigned.
- Otherwise: only a percentage of the embryos selected for transfer have been implanted and, although it is not possible to know which one(s) were successful, they also provide information about the behaviour of the system. class -1 will be assigned.

4 Proposed model and learning technique

The objective of the model proposed is to describe reality taking into account as many factors as possible. The datasets of cycles and embryos are linked, as each cycle produces a set of embryos. Both, cycles and embryos have its specific morphological characteristics.

The characteristics of the cycle of a woman that undergoes an IVF process can be enhanced, for instance, with the ovarian stimulation, or some other factors out of the scope of the work at hand. All the tasks conducted to improve the characteristics of one cycle to achieve pregnancy will also be called *configuration of the cycle*. Clinicians play a part in performing some modifications that will alter the next cycle the woman will have and the embryos related to it.

However there are factors, out of the scope of cycles and embryos, that affect whether an embryo selected to transfer will be implanted or will fail. Those are factors that do not depend on the quality of the cycle neither the embryos', and can be named *unknown factors*, as we do not know their origin or relation with the cycles or embryos, but have a role in implantation or failure.

For this work, it has been decided that the random manner those unknown factors show, can be modelled by a Bernoulli distribution, θ, which is split into two distributions, θ_1 and θ_0.

The first one, θ_1, codifies the perfect conditions: the probability that in a cycle that has been properly configured, an high quality embryo, selected for transfer, is implanted. In a perfect world, where there were no unknown factors that caused a failure, θ_1 would be set to 1 and the pregnancy will always occur. However this is not the case for this work, where $\theta_1 < 1$ and no one can assure when a pregnancy will take place.

To second one, θ_0, exists to complement θ_1, and codifies quite the opposite to θ_1. This is needed so that the model can be exemplified, and represents the cases when the embryo selected for transfer has low quality, the cycle the embryo belongs to has a bad configuration, or, simply, the embryo has not been selected to be transferred. Its value is always 0, as in this cases there is no chance for pregnancy.

4.1 Notation

The following notation has been created for this work:

c	index for cycle
e	index for embryo
C	Set of cycles
E_c	set of embryos associated to cycle c
S^c	set of embryos selected for transfer in cycle c

$\boldsymbol{x}_e$	characteristics of embryo e
$\boldsymbol{v}_c$	characteristics of cycle c

$\boldsymbol{w}_e$	Boolean random variable that represents whether embryo e is willing to implant
$\boldsymbol{r}_c$	Boolean random variable that represents whether a cycle c is willing to let embryos implant
$\boldsymbol{i}_e^c$	Boolean random variable that represents whether the uterus of the patient is willing to accept embryo e in cycle c
$\boldsymbol{y}_c$	Integer random variable that represents the number of embryos implanted in the cycle c

4.2 A probabilistic implantation model for IVF

A widely accepted assumption for IVF is that the individual characteristics of an embryo ($\boldsymbol{x}_e$), such as the embryo morphokinetic traits, are relevant in order to predict the probability of an embryo implanting in the uterus.

Here we assume that this can be quantified by means of a probability distribution

$$p(\boldsymbol{w}_e \mid \boldsymbol{x}_e; \alpha) \tag{1}$$

that measures the probability of the embryo to implant provided that the uterus is willing to accept it. That is, the probability for the embryo to implant in a "perfect uterus".

On the other hand, it is also accepted that the individual characteristics of a patient and the treatment performed exert an influence into the likelihood of an embryo to implant into her uterus. We also consider the simplifying assumption that only one transfer (of multiple embryos) is carried out in each cycle.

Here we assume that this can be quantified by means of a probability distribution

$$p(\boldsymbol{r}_c \mid \boldsymbol{v}_c; \beta) \tag{2}$$

which encodes how the characteristics of the patient influence the ratio of acceptance of embryos. An embryo willing to be implanted in the uterus of a willing-to-accept cycle, following

$$\boldsymbol{i}_e^c \sim Bernoulli(\theta_{w_e * r_c * s_e^c}) \tag{3}$$

where s_e^c is 1 when and embryo e is transferred in cycle c; 0 otherwise.

This implies that we accept the (highly unlikely but practical) assumption that the probability of acceptance by the uterus is statistically independent from the embryo characteristics. Provided that these hypothesis hold, we can assess the number of embryos implanted as:

$$\boldsymbol{y}_c = \sum_{e \in E_c} \boldsymbol{i}_e \tag{4}$$

4.3 General probability model for ARTs

Despite the fact that each cycle can carry out more than one transfer, this work only considers one transfer (of multiple embryos) for each cycle (see Section 4.2).

The system we expect to model is as shown in the following Figure 3:

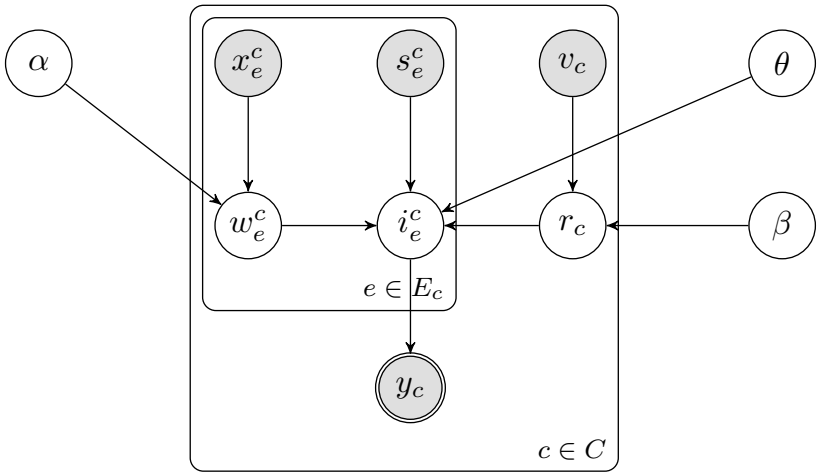


Figure 3: Graphical description of the simplified model. Shadowed nodes represent observed variables. Double line denotes a deterministic variable.

Shadowed nodes (x, v, s and y) represent observed variables. x and v define the characteristics of embryos and cycles. Their values can be directly measured. s determines if an embryo was selected; that is, a physicians' decision. And y is the outcome of the cycle: the number of implanted embryos.

White nodes inside the cycle box (w, r and i) represent latent variables, the value of which need to be inferred, which play an important role in defining the representation of the world, as they are connected to the observed variables.

White nodes outside the cycle box (α, β and θ) are the hyper-parameters for the cycles' and embryo's classifiers (Eq. 8 and 9) and for Bernoulli model (Eq. 10).

The joint probability is

$$p(\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{v}, \boldsymbol{r}, \boldsymbol{s}, \boldsymbol{i}, \boldsymbol{y}; \alpha, \beta, \theta) = p(\boldsymbol{w}|\boldsymbol{x}; \alpha)p(\boldsymbol{x})p(\boldsymbol{r}|\boldsymbol{v}; \beta)p(\boldsymbol{v})p(\boldsymbol{s})p(\boldsymbol{y}|\boldsymbol{i})p(\boldsymbol{i}|\boldsymbol{w}, \boldsymbol{r}, \boldsymbol{s}; \theta) \tag{5}$$

The relationship between i and y is deterministic and, by marginalizing $\boldsymbol{i}$ out, we can reduce the previous expression to the assignments of $\boldsymbol{i}$ that match the corresponding y_c outcome:

$$p(\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{v}, \boldsymbol{r}, \boldsymbol{s}, \boldsymbol{y}; \alpha, \beta, \theta) = \sum_{\breve{\boldsymbol{i}} \in \mathcal{I}_{\boldsymbol{s},\boldsymbol{y}}} p(\boldsymbol{w}|\boldsymbol{x}; \alpha)p(\boldsymbol{x})p(\boldsymbol{r}|\boldsymbol{v}; \beta)p(\boldsymbol{v})p(\boldsymbol{s})p(\breve{\boldsymbol{i}}|\boldsymbol{w}, \boldsymbol{r}, \boldsymbol{s}; \theta) \tag{6}$$

where $\mathcal{I}_{\boldsymbol{s},\boldsymbol{y}}$ is the set of valid vectors that assign value to all the embryos (implanted or not) according to the known outcomes $\{y_c\}_{c=1}^{N_C}$ and the selections $\{s_\cdot^c\}_{c=1}^{N_C}$. A valid vector assigns value i_e^c to all non-transferred embryos ($s_e^c = 0$) and, for every cycle c, $\sum_{e \in E_c} i_e^c = y_c$.

To exemplify this, let's consider a cycle with 10 embryos, where $\boldsymbol{e_1}$ and $\boldsymbol{e_7}$ are the 2 embryos selected to transfer and only one of them is implanted. As it is not possible to know which of the embryos is the one that is implanted, there would be 2 possible vectors defining the possible implants, with value 1 for and embryo implanted and 0 otherwise. The embryos not selected for transfer would be set to 0. The two vectors defining the possible implants would be:

- vector 1: [**1**, 0, 0, 0 , 0, 0, **0**, 0, 0, 0]
- vector 2: [**0**, 0, 0, 0 , 0, 0, **1**, 0, 0, 0]

The complete probability of generating the data from $\boldsymbol{x}$, $\boldsymbol{v}$ and $\boldsymbol{y}$, that is, the likelihood of the observed data, is

$$\begin{aligned}
p(\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{s}, \boldsymbol{y}; \alpha, \beta, \theta) &= \sum_{\boldsymbol{w},\boldsymbol{r}} p(\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{v}, \boldsymbol{r}, \boldsymbol{s}, \boldsymbol{y}; \alpha, \beta, \theta) \\
&= \sum_{\boldsymbol{w}} \sum_{\boldsymbol{r}} \sum_{\breve{\boldsymbol{i}} \in \mathcal{I}_{\boldsymbol{s},\boldsymbol{y}}} p(\boldsymbol{w}|\boldsymbol{x}; \alpha)p(\boldsymbol{x})p(\boldsymbol{r}|\boldsymbol{v}; \beta)p(\boldsymbol{v})p(\boldsymbol{s})p(\breve{\boldsymbol{i}}|\boldsymbol{w}, \boldsymbol{r}, \boldsymbol{s}; \theta) \\
&= p(\boldsymbol{x})p(\boldsymbol{v})p(\boldsymbol{s}) \sum_{\boldsymbol{r}} p(\boldsymbol{r}|\boldsymbol{v}; \beta) \sum_{\breve{\boldsymbol{i}} \in \mathcal{I}_{\boldsymbol{s},\boldsymbol{y}}} \sum_{\boldsymbol{w}} p(\breve{\boldsymbol{i}}|\boldsymbol{w}, \boldsymbol{r}, \boldsymbol{s}; \theta)p(\boldsymbol{w}|\boldsymbol{x}; \alpha)
\end{aligned}$$

from which we can define the probability of generating the outcome $\boldsymbol{y}$ as

$$p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{s}; \alpha, \beta, \theta) = \sum_{\boldsymbol{r}} p(\boldsymbol{r}|\boldsymbol{v}; \beta) \sum_{\breve{\boldsymbol{i}} \in \mathcal{I}_{\boldsymbol{s},\boldsymbol{y}}} \sum_{\boldsymbol{w}} p(\breve{\boldsymbol{i}}|\boldsymbol{w}, \boldsymbol{r}, \boldsymbol{s}; \theta)p(\boldsymbol{w}|\boldsymbol{x}; \alpha) \tag{7}$$

By assuming independence among instances given the parameters, we add more structure:

$$p(\boldsymbol{w}|\boldsymbol{x};\alpha) = \prod_{c=1}^{N_C} \prod_{e \in E_c} p(w_e^c|x_e^c;\alpha) \tag{8}$$

where $p(w_e^c|x_e^c;\alpha)$ is the *a priori* probability that an embryo x is willing to get implanted, controlled by parameter α. We call this the **embryo** model or **classifier**.

$$p(\boldsymbol{r}|\boldsymbol{v};\beta) = \prod_{c=1}^{N_C} p(r_c|v_c;\beta) \tag{9}$$

where $p(r_c|v_c;\alpha)$ is the *a priori* probability that a cycle v is prepared for implantation, controlled by parameter β. We call this the **cycle** model or **classifier**.

And

$$p(\boldsymbol{y}|\boldsymbol{w},\boldsymbol{r},\boldsymbol{s};\theta) = \prod_{c=1}^{N_C} \sum_{\check{\boldsymbol{i}} \in \mathcal{I}_{\boldsymbol{s}_\cdot^c, y_c}} \prod_{e \in E_c} p(\check{i}_e^c|w_e^c, r_c, s_e^c;\theta) \tag{10}$$

where $p(\check{i}_e^c|w_e^c, r_c, s_e^c;\theta)$ is the probability that embryo e gets implanted in cycle c, which we model by means of a **Bernoulli distribution**.

We are interested in finding the parameters $\langle \alpha, \beta, \theta \rangle$ which maximize the likelihood in Eq. 5. Equivalently, we look for the parameters that maximize the probability of the outcome $\boldsymbol{y}$ given by Eq. 7,

$$\alpha^*, \beta^*, \theta^* = \arg\max_{\alpha,\beta,\theta} p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{s};\alpha,\beta,\theta) \tag{11}$$

4.4 EM algorithm

An expectation–maximization (EM) algorithm is an iterative method to find (local) maximum likelihood of parameters in statistical models, where the model depends on unobserved latent variables. In our model, those latent variables are the ones that we do not know: w_e^c, r_c and i_e^c, and the model hyper-parameters α, β, θ.

The EM iteration alternates between performing an expectation (E) step, and a maximization (M) step.

- E step: the weights for cycles and embryos are computed taking into account all the current fit of the model and all the hidden variables (Eqs. 12, 13 and 14).
- M step: computes parameters maximizing the expected log-likelihood found on the E step.

These parameter-estimates are then used to determine the distribution of the latent variables in the next E step.

The specific EM algorithm for our particular model is defined in the following terms:

Note that we have a set of parameters $\eta = \langle \alpha, \beta, \theta \rangle$ such that

$$p(h, z; \eta) = p(h, z; \alpha, \beta, \theta) = p(h; \theta) p(z|h; \boldsymbol{\alpha})$$

or, specifically,

$$p(\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{r}|\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{s}; \alpha, \beta, \theta) = p(\boldsymbol{w}|\boldsymbol{x}; \alpha) p(\boldsymbol{r}|\boldsymbol{v}; \beta) p(\boldsymbol{y}|\boldsymbol{w}, \boldsymbol{r}, \boldsymbol{s}; \theta)$$

To compute the **E-step**, we need the likelihood of a cycle:

$$p(y_c|\boldsymbol{i}^c) p(\boldsymbol{i}^c|\boldsymbol{w}^c, r_c, \boldsymbol{s}^c; \theta) p(\boldsymbol{w}^c|\boldsymbol{x}^c; \alpha) p(r_c|v_c; \beta)$$

If we marginalize out the variables $\boldsymbol{i}^c$:

$$\Big(\sum_{\boldsymbol{i}^c} p(y_c|\boldsymbol{i}^c) p(\boldsymbol{i}^c|\boldsymbol{w}^c, r_c, \boldsymbol{s}^c; \theta) p(\boldsymbol{w}^c|\boldsymbol{x}^c; \alpha) \Big) p(r_c|v_c; \beta)$$

and

$$\Big(\sum_{\boldsymbol{i}^c \in \mathcal{I}_{\boldsymbol{s}^c_\cdot, y_c}} p(\boldsymbol{i}^c|\boldsymbol{w}^c, r_c, \boldsymbol{s}^c; \theta) p(\boldsymbol{w}^c|\boldsymbol{x}^c; \alpha) \Big) p(r_c|v_c; \beta)$$

and

$$\Big(\sum_{\boldsymbol{i}^c \in \mathcal{I}_{\boldsymbol{s}^c_\cdot, y_c}} \prod_e p(i^c_e|w^c_e, r_c, s^c_e; \theta) p(w^c_e|x^c_e; \alpha) \Big) p(r_c|v_c; \beta)$$

The **expected value of $\boldsymbol{r_c}$** is:

$$q(r_c = r) \propto \Big(\sum_{\boldsymbol{i}^c \in \mathcal{I}_{\boldsymbol{s}^c_\cdot, y_c}} \prod_e \sum_{w^c_e} p(i^c_e|w^c_e, r_c = r, s^c_e; \theta) p(w^c_e|x^c_e; \alpha) \Big) p(r_c = r|v_c; \beta) \tag{12}$$

Note that when there is a pregnancy ($y_c \geq 1$), $q(\mathrm{r}_c{=}1){=}1$ and $\mathrm{q}(r_c{=}0){=}0$: the probability that the cycle is willing to accept an embryo is 1, and the probability that the cycle is not willing to accept an embryo is 0.

The **expected value of $\boldsymbol{w^c_e}$** is:

$$q(w^c_e = w) \propto \sum_{r_c} \Bigg(\sum_{\boldsymbol{i}^c \in \mathcal{I}_{\boldsymbol{s}^c_\cdot, y_c}} p(i^c_e|w, r_c, s^c_e; \theta) p(w|x^c_e; \alpha) \cdot \prod_{e' \neq e} \sum_{w^c_{e'}} p(i^c_{e'}|w^c_{e'}, r_c, s^c_{e'}; \theta) p(w^c_{e'}|x^c_{e'}; \alpha) \Bigg) p(r_c|v_c; \beta) \tag{13}$$

Note that when the number of transferred embryos is the same as the number of implanted embryos, $\mathrm{q}(w^c_e){=}1$: the probability for all the embryos of the cycle to be willing to implant is 1. However, when the number of transferred embryos is different than the number of implanted ones, this value is unknown and its expected value has to be computed.

The **expected value of $\boldsymbol{i^c}$** is:

$$q(\boldsymbol{i}^c = \boldsymbol{i}) \propto \sum_{r_c} \Big(\prod_e \sum_{w_e^c} p(i_e | w_e^c, r_c, s_e^c; \theta) p(w_e^c | x_e^c; \alpha) \Big) p(r_c | v_c; \beta) \tag{14}$$

where $\boldsymbol{i} \in \mathcal{I}_{\boldsymbol{s}_:^c, y_c}$.

This is the probability for each of the i^c vectors to be the one that defines the embryos implanted. In the example posted after Eq. 6, each of the vectors would have a probability to be the one that defines the single embryo that is implanted among the two transferred.

To compute the update of the model parameters α, β, θ (**M-step**) we want

$$\arg\max_{\alpha,\beta,\theta} \mathbb{E}_{(\boldsymbol{w},\boldsymbol{r})\sim q} \log p(\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{r} | \boldsymbol{x}, \boldsymbol{v}, \boldsymbol{s}; \alpha, \beta, \theta)$$

Let us imagine that we do know the real value of all hidden variables. Thus, the likelihood would be

$$\prod_c \prod_{\boldsymbol{i}^{c'}} \left[\prod_{r_{c'}} \Big[p(r_{c'} | v_c; \beta) \prod_e \prod_{w_{e'}^c} \big[p(i_{e'}^c | w_{e'}^c, r_{c'}, s_e^c; \theta) p(w_{e'}^c | x_e^c; \alpha) \big]^{\mathbb{I}[w_{e'}^c = w_e^c]} \Big]^{\mathbb{I}[r_{c'} = r_c]} \right]^{\mathbb{I}[\boldsymbol{i}^{c'} = \boldsymbol{i}^c]}$$

and the log-likelihood:

$$\sum_c \sum_{\boldsymbol{i}^{c'}} \mathbb{I}[\boldsymbol{i}^{c'} = \boldsymbol{i}^c] \Bigg[\sum_{r_{c'}} \mathbb{I}[r_{c'} = r_c] \Big[\log p(r_{c'} | v_c; \beta) + \sum_e \sum_{w_{e'}^c} \mathbb{I}[w_{e'}^c = w_e^c] \big[\log p(i_{e'}^c | w_{e'}^c, r_{c'}, s_e^c; \theta) + \log p(w_{e'}^c | x_e^c; \alpha) \big] \Big] \Bigg]$$

But, if the real values are unknown, we need to resort the expected values as,

$$\sum_c \sum_{\boldsymbol{i}^{c'} \in \mathcal{I}_{\boldsymbol{s}_:^c, y_c}} q(\boldsymbol{i}^{c'}) \Bigg[\sum_{r_{c'}} q(r_{c'}) \Big[\log p(r_{c'} | v_c; \beta) + \sum_e \sum_{w_{e'}^c} q(w_{e'}^c) \big[\log p(i_{e'}^c | w_{e'}^c, r_{c'}, s_e^c; \theta) + \log p(w_{e'}^c | x_e^c; \alpha) \big] \Big] \Bigg]$$

Note that the variables i follow a Bernoulli distribution:

$$i_e^c \sim \ Bernoulli(\theta_{r_c \cdot w_e^c \cdot s_e^c})$$

where, in practice, θ_0 fixed to $\theta_0 = 0$ (whenever r_c, w_e^c, or s_e^c are zero –no transfer, or bad cycle/embryo) and θ_1 determines the probability of implantation

in perfect conditions. To find the parameter θ_1, we derive the log-likelihood with respect to θ_1, and set it to 0:

$$\frac{\partial \left[\sum_c \sum_{\boldsymbol{i}^{c'} \in \mathcal{I}_{\boldsymbol{s}^c_{\vdots}, y_c}} q(\boldsymbol{i}^{c'}) \left[q(r_c = 1) \Big[\sum_e q(w_e^c = 1) \big[i^c_{e'} \log \theta_1 + (1 - i^c_{e'}) \log(1 - \theta_1) \big] \Big] \right] \right]}{\partial \theta_1} = 0$$

$$\sum_c \sum_{\boldsymbol{i}^{c'} \in \mathcal{I}_{\boldsymbol{s}^c_{\vdots}, y_c}} \sum_e q(\boldsymbol{i}^{c'}) q(r_c = 1) q(w_e^c = 1) \left[\frac{i^c_{e'}}{\theta_1} - \frac{(1 - i^c_{e'})}{(1 - \theta_1)} \right] = 0$$

$$\sum_c \sum_{\boldsymbol{i}^{c'} \in \mathcal{I}_{\boldsymbol{s}^c_{\vdots}, y_c}} \sum_e q(\boldsymbol{i}^{c'}) q(r_c = 1) q(w_e^c = 1) \left[(1 - \theta_1) i^c_{e'} \right] =$$

$$\sum_c \sum_{\boldsymbol{i}^{c'} \in \mathcal{I}_{\boldsymbol{s}^c_{\vdots}, y_c}} \sum_e q(\boldsymbol{i}^{c'}) q(r_c = 1) q(w_e^c = 1) \left[(1 - i^c_{e'}) \theta_1 \right]$$

$$\sum_c \sum_{\boldsymbol{i}^{c'} \in \mathcal{I}_{\boldsymbol{s}^c_{\vdots}, y_c}} \sum_e q(\boldsymbol{i}^{c'}) q(r_c = 1) q(w_e^c = 1) i^c_{e'} = \sum_c \sum_{\boldsymbol{i}^{c'} \in \mathcal{I}_{\boldsymbol{s}^c_{\vdots}, y_c}} \sum_e q(\boldsymbol{i}^{c'}) q(r_c = 1) q(w_e^c = 1) \theta_1$$

$$\theta_1 = \frac{\sum_c \sum_{\boldsymbol{i}^{c'} \in \mathcal{I}_{\boldsymbol{s}^c_{\vdots}, y_c}} \sum_e q(\boldsymbol{i}^{c'}) q(r_c = 1) q(w_e^c = 1) i^c_{e'}}{\sum_c \sum_{\boldsymbol{i}^{c'} \in \mathcal{I}_{\boldsymbol{s}^c_{\vdots}, y_c}} \sum_e q(\boldsymbol{i}^{c'}) q(r_c = 1) q(w_e^c = 1)} \tag{15}$$

This is the value that gives meaning to the model, as it expresses its goodness. It shows the probability that a high-quality embryo selected for transfer in a well-configured cycle, results in pregnancy. The higher this probability is, the more reliable the model is predicting outcomes for new cycles.

The resulting method is shown in Algorithm 1:

Algorithm 1 Our EM algorithm

1: **procedure** EM($\alpha^{(0)}, \beta^{(0)}, \theta^{(0)}$)
2: $\alpha, \beta, \theta \leftarrow \alpha^{(0)}, \beta^{(0)}, \theta^{(0)}$
3: **while** q not converged **do**
4: $q \leftarrow p(i, w, r | y, x, v; \alpha, \beta, \theta)$ ▷ Update q: E-step (Eqs. 12, 13, 14)
5: $\alpha \leftarrow \arg\max_\alpha \mathbb{E}_{\boldsymbol{w} \sim q} \log p(\boldsymbol{w} | \boldsymbol{x}; \alpha)$ ▷ Update α: M1-step
6: $\beta \leftarrow \arg\max_\beta \mathbb{E}_{\boldsymbol{r} \sim q} \log p(\boldsymbol{r} | \boldsymbol{v}; \beta)$ ▷ Update β: M2-step
7: $\theta \leftarrow \arg\max_\theta \mathbb{E}_{\boldsymbol{i} \sim q} \log p(y | ...; \theta)$ ▷ Update θ: M3-step (Eq. 15)
8: **end while**
9: **return** $\eta = \langle \theta, \boldsymbol{\alpha} \rangle$
10: **end procedure**

In this method, we propose to roughly model $p(\boldsymbol{w}|\boldsymbol{x}; \alpha)$ and $p(\boldsymbol{r}|\boldsymbol{v}, \beta)$ by means of probabilistic classifiers. Thus, steps 5 and 6 in Algorithm 1 would be just the learning steps of those classifiers, given the corresponding weights of Equations 13 and 12 respectively. Step 7 is the calculation of the probability of implant in perfect conditions (Eq. 15).

4.5 EM algorithm for IVF model

As EM algorithm iterates, the weights and probabilities are computed and the final model will, eventually, be defined. The algorithm starts by filling the gaps for the latent variables with random values that, as iterations go on, will be adjusted and will finally stabilize. At each iteration a new definition of the reality is being defined.

The EM algorithm is computed on the training set until the system converges, as this is a learning strategy that wants to understand the behaviour of the dataset. While the procedure does not converge, the Estimation and the Maximization steps will be iterated and the model will evolve. Once it has converged, the model is stable. As a product of this learning procedure, the classifier for cycles and the classifier for embryos will have been learnt, and will be capable to predict outcome from new data and estimate their performance.

The main steps for the algorithm are:

1. Initialization: weights for cycles and embryos are initialized.
2. E-step: estimates the new weights from the whole formulas of the system.
3. M-step: model learning with the estimated values from E-step.

Steps 2 and 3 will iterate until the weights converge. The final θ_1 will be computed providing the probability that, given a properly-configured cycle, a good-quality embryo selected to transfer will implant.

4.5.1 Initialization step

There are two strategies for the initialization of the EM algorithm: (i) estimate the probabilities (q_r for cycles, q_w for embryos and q_i for implantation vectors) and use them to learn the first fit of the model, or (ii) assign random values to parameters α, β and θ_1.

In this work, the first strategy is taking into account **the model**, composed by cycles, embryos, and unknown factors the probabilities that need to be initialized are:

- $q(w = 1)$: probability that an embryo is classified as class 1.
- $q(w = 0)$: probability that an embryo is classified as class 0.
- $q(r = 1)$: probability that a cycle is classified as class 1.
- $q(r = 0)$: probability that a cycle is classified as class 0.
- $q(i^c = 1)$: probability that a specific implantation vector is classified as class 1.
- $q(i^c = 0)$: probability that a specific implantation vector is classified as class 0.

Using the dataset completed with the previously computed weights, the classifiers for the cycles and the embryos can be learnt again and θ_1 computed.

4.5.2 Estimation step

With the current fit of the model, the probability distribution for each of the classifiers ($p(r|v)$ for cycles and $p(w|x)$ for embryos) can be calculated.

These probability distributions are now used to update the weights (q_r, q_w and q_i) that are computed with Eqs. 12, 13 and 14.

4.5.3 Maximization step

Using the dataset completed with the previously computed weights (q_r, q_w and q_i), the classifiers for the cycles and the embryos can be learnt again. θ_1 is computed at this step too. At this point, we can consider that a new fit of the model have been learnt.

5 Experiments

Given the model described in section 4.5, we aim to perform a solid validation, in which we test each of the probabilistic classifiers, to ensure that we choose the most suited for the selection of the best embryos; the ones with more probability of implantation once transferred to the uterus in one cycle.

Recall that we assume we have two types of embryos (class 1 and class 0) and two types of cycles in a uterus (class 1 and class 0), being:

- class 1: Positive embryos or cycles. Embryos willing to implant in the uterus, or cycles in which the uterus is receptive to the implantation of a selected embryo.
- class 0: Negative embryos or cycles. Embryos not willing to implant in the uterus, or cycles in which the uterus is not receptive to the implantation of a selected embryo.

5.1 Goals

Our goal here is to evaluate the behaviour of the model. Using test data, we will detect whether there could be benefits from using our model, that uses EM strategy, to estimate the implantation probability of an embryo in an uterus, with perfect conditions, against a baseline probabilistic classifier.

A cycle is considered to be in perfect conditions if the uterus is receptive to the implantation of embryos, and the embryos selected to be transferred are willing to be implanted in the uterus.

We present two sets of experiments:

- Experiment 1: Find the best model in EM strategy: Which is the probabilistic classifier that provides higher positive recall and lower level proportion loss.
- Experiment 2: Compare the model with a baseline classifier, which uses a non-EM strategy at all.

5.2 Probabilistic classifiers

Given the heterogeneity of the data, there is not a probabilistic classifier that we can advocate for as the best to use for the selection of the best embryos in a cycle. Each variable, due to its own characteristics, will be suitable for a different probabilistic classifier.

The following probabilistic classifiers are considered in this work:

- LR: Linear regression is a type of regression analysis where the number of independent variables is one and there is a linear relationship between the independent(x) and dependent(y) variable.

- LRCV: Logistic Regression that uses Cross-Validation to optimize the hyper-parameters such as the regularization strength.
- RF200 and RF500: Random Forest with 200 or 500 trees. It is a meta estimator that fits a number of decision tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting
- GBOOST: Gradient Boosting for classification. It builds an additive model where it allows for the optimization of arbitrary differentiable loss functions. In each stage n (number of classes) regression trees are fit on the negative gradient of the binomial or multinomial deviance loss function. Binary classification is a special case where only a single regression tree is induced.
- DTREE: Decision Tree classifier.
- ETREE: Extremely randomized tree classifier. When looking for the best split to separate the samples of a node into two groups, random splits are drawn for each of the max_features.
- ETREES: Extra-trees classifier that fits a number of randomized decision trees (extra-trees) on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting.

5.3 Performance measures

In both experiments some of the following performance measures will be computed:

- Euclidean Distance: As this is computed between two 1-D arrays, we compute this for each variable's weight or probability between two iterations of the EM strategy, then we compute the mean for all the variables. [4]

  The Euclidean distance between 1-D arrays u and v, is defined as:

$$(\sum(w_i|(u_i - v_i)|^2))^{1/2} \tag{16}$$

- Relative Entropy: As this is an element-wise computation, we compute this for each element of each variable's weight or probability between two iterations of the EM strategy, add the results, then we compute the mean for all the variables. [6]

  Element-wise function for computing relative entropy

$$rel\_entr(x, y) = \begin{cases} x \cdot \log(\mathrm{x/y}) & \mathrm{x}>0,\ \mathrm{y}>0 \\ 0, & \mathrm{x}=0\ \mathrm{y}\geq 0 \\ \infty & \text{otherwise} \end{cases} \tag{17}$$

- Level Proportion Loss: *bag_loss (lp_loss)*: Applied to embryos. Given a set of transferred embryos, when not all the embryos are implanted and my prediction is the same as the real data, the loss is 0. The aim is that the implanted proportion for the prediction and for reality is the same.

- Pseudo-Recall: *ps_recall*: Applied to cycles. Its is and estimation of the real recall as it is the recall only for the set of positive cycles. From the cycles that are positive, which percentage has been predicted as such.

- AUC-ROC score: *Area Under the Receiver Operating Characteristic Curve*, is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied. The ROC curve is created by plotting the true positive rate (TPR or recall) against the false positive rate (FPR or fall-out) at various threshold settings. The true-positive rate is the number of embryos classified as positives, over the total number of positives (class 1). The false-positive rate is the probability of false alarm, the number of embryos mis-classified as positive, over the total number of negatives (class 0)[10].

  - When AUC = 1, then the classifier is able to perfectly distinguish between all the Positive and the Negative class points correctly. If, however, the AUC had been 0, then the classifier would be predicting all Negatives as Positives, and all Positives as Negatives.
  - When $0.5 < \text{AUC} < 1$, there is a high chance that the classifier will be able to distinguish the positive class values from the negative class values. This is so because the classifier is able to detect more numbers of True positives and True negatives than False negatives and False positives.
  - When AUC = 0.5, then the classifier is not able to distinguish between Positive and Negative class points. Meaning either the classifier is predicting random class or constant class for all the data points.

5.3.1 Reference measures

There are two measures that are not used to test the performance of a probabilistic classifier, but to provide more information about the behaviour of the model using such classifier.

- Proportion of Predicted Positive (*ppr*): Applied to cycles as well as embryos. It indicates how often the classifier is predicting the positive class. This is not a real measure, but we can use it to see the behaviour of the model.

- theta1: probability of an embryo to implant in perfect conditions. This is an hyper-parameter for the model.

5.4 Experiment 1: Best model in EM strategy

The aim for this experiment is to test our model with different probabilistic classifiers.

The datasets of cycles and embryos have been split into 5 folds using 5K-Fold cross-validation to avoid possible problems like over-fitting or selection bias are avoided[2].

To initialize the EM algorithm all the instances of the embryos, as well as cycles, are randomly assigned to a class that, eventually, will define if an embryo is willing to implant or if an uterus is willing to accept embryos for her cycle (class 1), or not (class 0). Some general variables (see Section 3) have been created to keep track of the relations between cycles and embryos.

For each of the probabilistic classifiers, and each of the Fold combinations, the EM algorithm (which is run for a maximum of 100 iterations or until convergence) is run for 20 executions. After those, the best model for this fold and probabilistic classifier can be found. This would be the one with lower level proportion loss (lp_loss), meaning that the number of embryos predicted and implanted are the same as shown in real data.

The probabilistic classifiers that have been tested with our EM model are the ones described in Section 5.2, from which some performance measures have been computed (see section 5.5).

In the process of learning the best model, the weights and probabilities for cycles and embryos stabilize through the iterations of the EM algorithm.

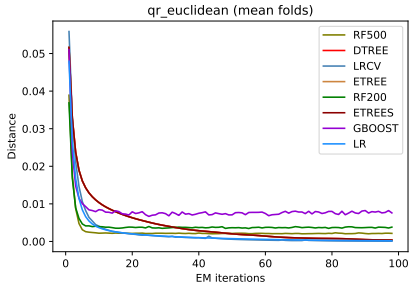
- weights: cycle's and embryo's viability estimations
- probabilities: cycle's and embryo's probabilities according to the respective probabilistic classifier

To show how the model converges as iterations go on, Euclidean Distance and Relative Entropy have been computed for each of the weights and probabilities' matrices along the EM steps.

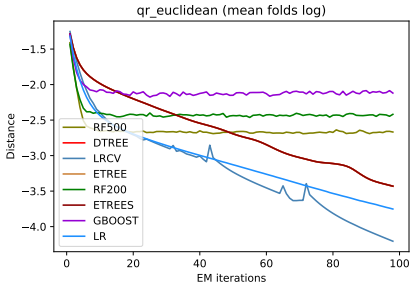
Figure 4 and Figure 5 show the plots for the mean of the folds, for all the probabilistic classifiers that have been tested. Each of the figures show a different behaviour, and the plots are shown in normal-scale and logaritmic-scale to help the reader comprehend them.

The SVC probabilistic classifier (Support Vector Classifier) was also tested, although its results are not competitive and it was very time consuming, so they are not shown.

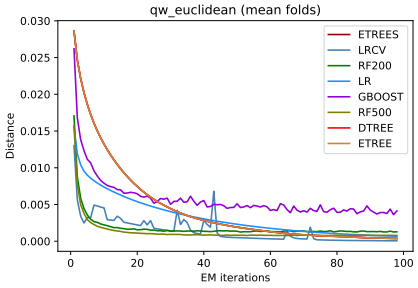
As seen in the following figures, Euclidean Distance, as well as Relative Entropy, show that the EM model converges faster with the LRCV probabilistic classifier.



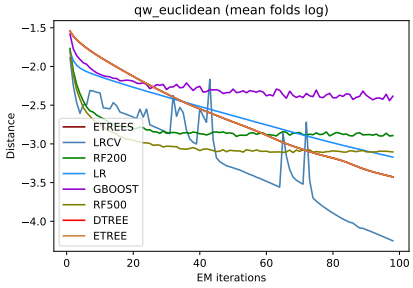
(a) q_r: cycle's viability estimation



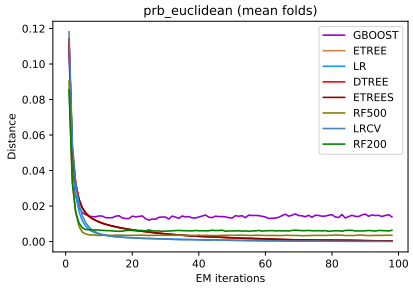
(b) $\log q_r$: cycle's viability estimation in log-scale



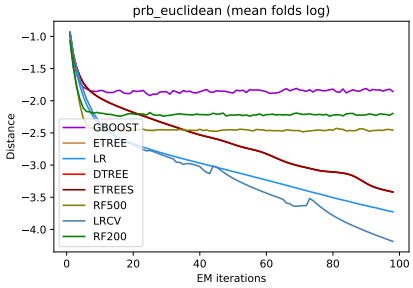
(c) q_w: embryo's viability estimation



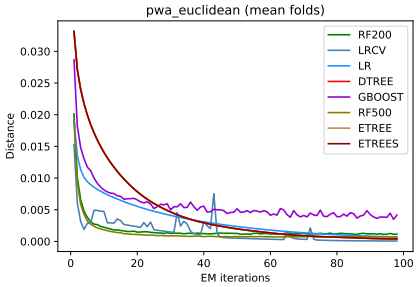
(d) $\log q_w$: embryo's viability estimation in log-scale



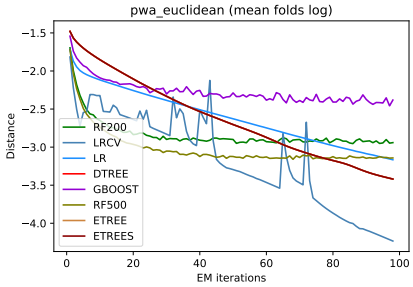
(e) p_r: cycle's probability according to the respective probabilistic classifier



(f) $\log p_r$: cycle's probability according to the respective probabilistic classifier in log-scale

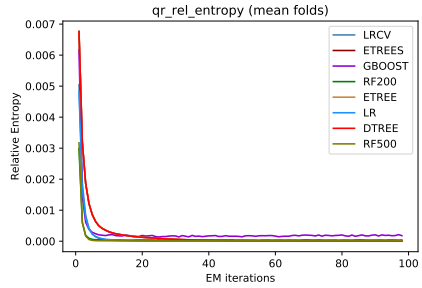


(g) p_w: embryo's probability according to the respective probabilistic classifier

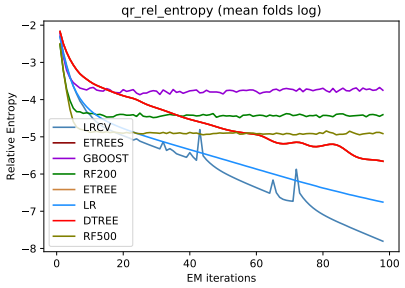


(h) $\log p_w$: embryo's probability according to the respective probabilistic classifier in log-scale

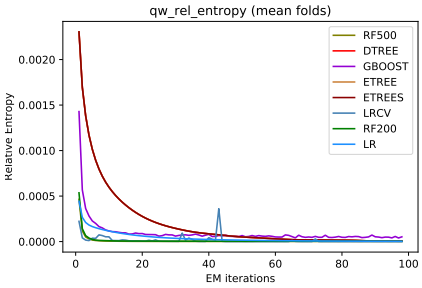
Figure 4: Euclidean Distances measured between the consecutive iterations of the EM algorithm (mean of the k-folds), for each of the estimations and probability's matrices.



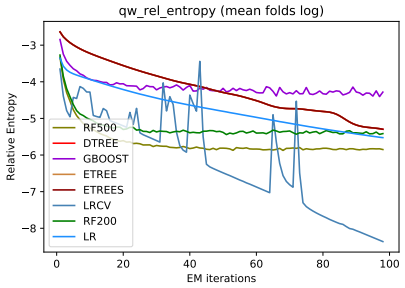
(a) q_r: cycle's viability estimation



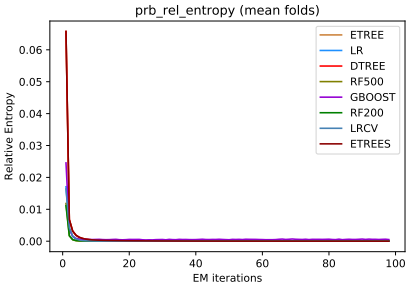
(b) $\log q_r$: cycle's viability estimation in log-scale



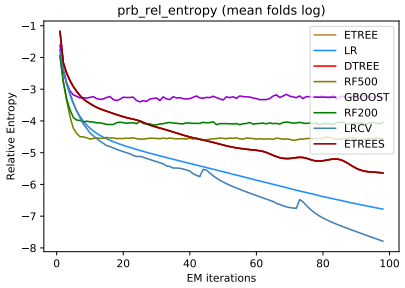
(c) q_w: embryo's viability estimation



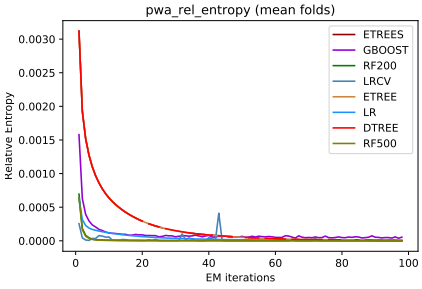
(d) $\log q_w$: embryo's viability estimation in log-scale



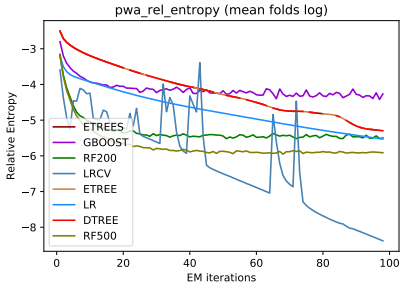
(e) p_r: cycle's probability according to the respective probabilistic classifier



(f) $\log p_r$: cycle's probability according to the respective probabilistic classifier in log-scale



(g) p_w: embryo's probability according to the respective probabilistic classifier



(h) $\log p_w$: embryo's probability according to the respective probabilistic classifier in log-scale

Figure 5: Relative Entropy measured between the consecutive iterations of the EM algorithm (mean of the k-folds), for each of the estimations and probability's matrices.

As in EM strategy there are some latent values, it is not possible to compute some measures such as accuracy. The following are the measures used to see the performance of each of the probabilistic classifiers applied to our EM model (see Table 9):

Pr.Classifier	**lp_loss**	**ppr cycles**	**ps_recall**	**ppr embryos**	**theta1**
DTREE	**0.6301** ±0.0502	0.6324 ±0.0371	0.4209 ±0.0423	0.4475 ±0.0455	0.7356 ±0.0313
ETREE	**0.6096** ±0.0528	0.5994 ±0.0585	0.4223 ±0.0459	0.4436 ±0.0518	0.7356 ±0.0313
ETREES	0.9651 ±0.0374	0.9636 ±0.0237	0.4069 ±0.0394	0.4244 ±0.0470	0.7356 ±0.0313
GBOOST	0.9784 ±0.0191	0.9884 ±0.0084	0.4851 ±0.0646	0.4137 ±0.0793	0.5198 ±0.0054
LRCV	1.0 ±0.0	1.0 ±0.0	0.2768 ±0.1309	0.0869 ±0.1738	0.4825 ±0.0057
LR	0.9684 ±0.0286	0.9751 ±0.0156	0.4040 ±0.0435	0.3079 ±0.0309	0.5307 ±0.0065
RF200	1.0 ±0.0	1.0 ±0.0	**0.6661** ±0.1745	0.6982 ±0.1728	0.4836 ±0.0001
RF500	1.0 ±0.0	1.0 ±0.0	**0.6846** ±0.1685	0.7231 ±0.2748	0.4832 ±0.0056

Table 9: Performance measures for the model using EM strategy

The lower results in lp_loss are for ETREE classifier and DTREE classifier, which are the lower values. The higher results in ps_recall is for the RF500 and RF200 classifiers.

5.4.1 Discussion

Euclidean Distance show smaller distances between iterations than Relative Entropy. The TREE models show the same behaviour, which is a convergence worse than LRCV and LR but does not present any jump in the plots.

For the cycle's and embryo's probability to the respective probabilistic classifier, LRCV present some jump between states that seem of no importance in the final convergence. On the contrary, there are probabilistic classifiers like GBOOST, RF200 or RF500 that, in the first iterations start to converge as fast as the other classifiers, but there is a point when they remain in an almost flat plot jumping between two states not being able to converge so properly.

To find the best probabilistic classifier for the EM model, a balance between lp_loss and ps_recall has to be found. Although RF500 and RF200 present the higher values for ps_recall, they have a very high value for lp_loss (1.0) which invalidate these results. DTREE and ETREE, on the other hand, have a lower value of ps_recall, but the lower values in lp_loss.

There are two references in the same table of results, ppr for cycles and ppr for embryos, that indicate how often the classifier predict a positive class, for cycles as well as for embryos. Our best interest is that those values are close to 0.5. We can see again that for RF500 and RF200, the ppr for cycles is 1.0 which means that this classifier is always predicting positive classes, which invalidates the ps_recall result.

Thus, the best combination in performance measures and references is for ETREE and DTREE probabilistic classifiers, which balance the lp_loss (0.6301 and 0.6096) and ps_recall (0.4209 and 0.4223) with the ppr for cycles 0.6324 and 0.5994) and embryos (0.4475 and 0.4436) close to 0.5 for both cases.

Given the plots generated in 4 and 5, ETREE, DTREE are two of the probabilistic classifiers that converge continuously, although slow. On the other hand, RF200 and RF500 are two of the probabilistic classifiers that do not converge.

Moreover, the other reference to take into account is theta1, which tells us the percentage of uncertainty that the model tolerates. The higher the value the better its tolerance, which means that for a well-configured cycle, the probability of implantation for an embryo selected for transfer will be high. In this example, ETREE and DTREE have also a high value (0.7356) for theta1.

5.5 Experiment 2: EM strategy vs Baseline classifier

The aim for this experiment is to compare the model with a baseline classifier, which uses non-EM strategy at all.

The Baseline classifier is a probabilistic classifier used to test the same dataset as the EM model. The difference in how the data is treated in this case is that the embryos that belong to a cycle where the number of transferred embryos is different than the number of implanted embryos, are considered as failed embryos. Those embryos can be used for training the baseline classifier but, **only** the embryos belonging to cycles where all the embryos were implanted or all the embryos failed implanting, are used to test it.

For each of the probabilistic classifiers described in Section 5.2, we compute the AUC-ROC score from predicted embryo classes, which allows us to compare our EM model with the baseline model. The higher the AUC value for a classifier, the better its ability to distinguish between positive and negative classes.

The results for the mean for all folds for AUC-ROC score are shown in Table 10:

Prob.Classifier	EM model	Baseline classifier
DTREE	0.6606	0.7107
ETREE	0.6618	0.7137
ETREES	0.6567	0.7089
GBOOST	0.7290	0.7406
LRVC	0.7111	0.7373
LR	0.6792	0.7224
RF200	**0.7350**	0.7423
RF500	0.7338	**0.7435**

Table 10: AUC-ROC score comparison for EM model vs Baseline classifier

The results shows that the baseline classifier is better than the EM model for all the probabilistic classifiers tested. RF500 is the classifier with higher AUC-ROC score for baseline (0.7435) and RF200 is the classifier with higher result for the EM model (0.7350).

5.5.1 Discussion

Although the results for the baseline classifier are better than the results for the EM model, there is a small difference in results. Four of the eight probabilistic classifiers (RF200: 0.7350, RF500: 0.7338, GBOOST: 0.7290, LRVC: 0.7111), which is half of the classifiers, perform better with EM model than the worse baseline classifier (ETREES:0.7089).

Although the best results for AUC-ROC score in EM model are for RF200, RF500, GBOOST and LRCV, we have seen in Table 9 that these classifiers are the ones with better ps_recall but, at the same time, have the higher values of ppr per cycles, which invalidate its performance.

Additionally, the better performance measures for the model using EM strategy, coincide with three of the lower results for AUC-ROC score using EM model.

6 Conclusions

In this work, we propose a novel probabilistic graphical model to deal with the embryo selection problem in assisted reproduction. This is a complex problem where, not only the characteristics of each cycle and its belonging embryos take part in the problem. The proposed model is, to the extent of our knowledge, the first one that takes into account three different possible sources of uncertainty and, accounts for the unknown factors that can cause that an embryo in perfect conditions, selected to transfer, fail to implant. We also derived completely a method based on the EM strategy to learn from the type of data characteristic of this problem.

The results show that the approximation presented to the embryo selection problem using an EM algorithm to create a model, present worse performance results than a Baseline probabilistic classifier, which makes this model worse at selecting embryos for transfer than a Baseline Model.

We have also noticed that ETREE, DTREE, two of the classifiers with worse results for the EM model, are two of the probabilistic classifiers that converge continuously as EM iterations go on. However, RF200 and RF500, which are the classifiers withe best AUC-ROC results for Baseline classifier, are two of the probabilistic classifiers that do not converge, and that have worse lp_loss and ppr for cycles in the performance measures using EM strategy, contrary to ETREE and DTREE.

The main success of this work is that we have been able to compute an estimation for $\boldsymbol{\theta_1}$, something novel, that provides a very valuable information for physicians, as it helps model reality. As explained in previous sections, θ_1 is an hyper-parameter that estimates the probability of an embryo to implant in perfect conditions. This means the probability that for a cycle that has been properly configured, a high quality embryo, selected for transfer, is implanted.

The novelty of the estimation for the hyper-parameter θ_1, provides a new modelling of the system, that can assist clinicians in decision making and improve ARTs.

6.1 Future Work

After finishing this work, many issues have been left open. First of all, the empirical validation of the method by means of enlarged experimental setting is still possible. Secondly, although we managed to obtain and use real data, note that it only involves a single hospital. In order to evaluate more fairly our approach, data from different sources should be considered. Moreover, our approach for embryo selection was based on extracting the probabilistic classifier for $p(w|x;\alpha)$ and using it to make the predictions. An alternative that should be studied would be the whole model for making predictions. Finally, the most challenging idea for future work would be to try to validate, in collaboration with clinicians, the value for θ_1 obtained by our model and it relationship with the proportion of properly configured cycles that fail to implant.

List of Figures

List of Tables

References

[1] Andrew E Bradley. "The use of the Area Under the ROC curve in the evaluation of Machine Learning Algorithms". In: ().

[2] Cross-Validation. In: (). DOI: https://en.wikipedia.org/wiki/Cross-validation_(statistics).

[3] Pedro Larrañaga Dinora A. Moralesa Endika Bengoetxea. "Selection of human embryos for transfer by Bayesian classifiers". In: ().

[4] Euclidean Distance. In: (). DOI: https://docs.scipy.org/doc/scipy/reference/generated/scipy.spatial.distance.euclidean.html.

[5] M.Sc. Elnur Babayev M.D. "Man versus machine in IVF—can artificial intelligence replace physicians?" In: ().

[6] Relative Entropy. In: (). DOI: https://docs.scipy.org/doc/scipy/reference/generated/scipy.special.rel_entr.html#scipy.special.rel_entr.

[7] C. Moscatellia G. Patrizia C. Mannab and L. Nieddua. "Pattern recognition methods in human-assisted reproduction". In: ().

[8] Ph.D. Nikica Zaninovic and M.D. Zev Rosenwaks. "Artificial intelligence in human in vitro fertilization and embryology". In: ().

[9] Asociación para el estudio de la biologia de reproducción. "Cuadernos de embriología clínica". In: ().

[10] ROC AUC score. In: (). DOI: https://en.wikipedia.org/wiki/Receiver_operating_characteristic.

[11] "Tasas de éxito de los tratamientos de reproducción asistida". In: ().

More References

[rif02] Should zygote intrafallopian transfer be offered to all patients with unexplained repeated in-vitro fertilization cycle failures?.Itai Gat, Jacob Levron, Gil Yerushalmi, Jehoshua Dor, Masha Brengauz, Raoul Orvieto

[rif03] What exactly do we mean by recurrent implantation failure? A systematic review and opinion. Lukasz T.Polanski, Miriam N.Baumgarten, Siobhan Quenby, Jan Brosens, Bruce K. Campbell, Nicholas J.Raine-Fenning

[rif04] Natural killer cells in female infertility and recurrent miscarriage: a systematic review and meta-analysis. Srividya Seshadri, Sesh Kamal Sunkara

[paperdeu] Estimating the chance of success in IVF treatment using a ranking algorithm. H. Altay Güvenir, Gizem Misirli, Serdar Dilbaz, Ozlem Ozdegirmenci, Berfu Demir, Berna Dilbaz

[papernou] Artificial Intelligence in IVF: A Need. Charalampos Siristatidis, Abraham Pouliakis, Charalampos Chrelias, Dimitrios Kassanos.

[papercinc] Artificial intelligence in the IVF laboratory: overview through the application of different types of algorithms for the classification of reproductive data. Eleonora Inácio Fernandez, Andre Satoshi Ferreira, Matheus Henrique Miquelao Cecilio, Doris Spinosa Cheles, Rebeca Colauto Milanezi de Souza, Marcelo Fabio Gouveia Nogueira, Jose Celso Rocha.

[papercatorze] Development of an artificial intelligence-based assessment model for prediction of embryo viability using static images captured by optical light microscopy during IVF. M. VerMilyea, J.M.M. Hall, S.M. Diakiw, A. Johnston, T. Nguyen, D. Perugini, A. Miller, A. Picou, A.P. Murphy, and M. Perugini

[paperset] Artificial intelligence techniques for embryo and oocyte classification. Claudio Manna, Loris Nanni, Alessandra Lumini, Sebastiana Pappalardo.

[paperquatre] Automatic grading of human blastocysts from time-lapse imaging. Mikkel F. Kragh, Jens Rimestad, Jørgen Berntsen, Henrik Karstoft.

[paperdos] A Bayesian network model for predicting pregnancy after in vitro fertilization. G. Corani, C. Magli, A. Giusti, L. Gianaroli, L.M. Gambardella

[papersis] Bayesian classification for the selection of in vitro human embryos using morphological and clinical data. Dinora Araceli Moralesa, Endika Bengoetxea, Pedro Larran, Miguel García, Yosu Franco, Mónica Fresnada, Marisa Merinod.

[papertres] Fitting the data from embryo implantation prediction: Learning from label proportions. Jerónimo Hernández-González, Iñaki Inza, Lorena Crisol-Ortíz, María A Guembe, María J Iñarra and Jose A Lozano.

[papervintiquatre] Novel and conventional embryo parameters as input data for artificial neural networks: an artificial intelligence model applied for prediction of the implantation potential. Lorena Bori, M.Sc., Elena Paya, M.Sc., Lucia Alegre, M.Sc., Thamara Alexandra Viloria, Ph.D., Jose Alejandro Remohi, M.D., Ph.D., Valery Naranjo, Ph.D., Marcos Meseguer, Ph.D.