Bayesian Network Modeling of IVF Blastocyst Score Prediction

Aslı UYAR, *Okan University, Computer Engineering, Assistant Prof., asliuyar@gmail.com,*
0000-0002-7913-1083

Yasemin ATILGAN ŞENGÜL, *Doğuş University, Industrial Engineering, Assistant Prof.,*
yatilgan@dogus.edu.tr, *0000-0002-5109-2262*

ABSTRACT

Embryo transfer may be performed at cleavage stage (on day 2-3) or at blastocyst stage (on day 5) in In-Vitro Fertilization (IVF) treatment. Elective single embryo transfer at blastocyst stage increases the pregnancy probability and reduces the number of multiple pregnancies. However, the extended culture of embryos in the laboratory may result in transfer cancelation if no high quality blastocyst develops by day 5. Predicting the blastocyst score of individual embryos may help physicians to decide whether or not to further culture the embryos in the laboratory.

In this paper, we use Bayesian networks for predicting the blastocyst score by modeling the morphological evolution of IVF embryos. We propose a weighted nearest neighbor approach to adjust the frequency estimates in the conditional probability table. Experimental results show that the proposed method significantly increases the accuracy and reduces false positive rates in IVF data in comparison to the frequency estimate method. Our proposed model can also predict low quality blastocyst development with a 77.3% True Negative rate. Using this model can help preventing developmental failures of embryos during IVF treatment.

Keywords : In-Vitro Fertilization, Predicting Blastocyst Development, Bayesian Networks, Parameter Learning, Frequency Estimates

Bayes Ağları ile Tüp Bebek Tedavi Sürecinde Blastosist Skoru Tahmini

ÖZ

Tüp bebek tedavisinde embriyo transferi bölünme aşamasında (gün 2-3) veya blastosist aşamasında (gün 5) gerçekleştirilebilir. Transfer öncesi tek embriyo seçimi ve transferi gebelik olasılığını arttırırken çoklu gebelik sayısını da düşürür. Diğer taraftan, laboratuvar ortamında uzayan embriyo kültürleme zamanı beşinci güne kadar yüksek kaliteli blastosist gelişmediği takdirde transferin iptal olmasına sebep olabilir. Blastosist skorlarının tahminlenmesi klinisyenlere her bir embriyonun laboratuvar ortamında kültürlenmeye devam edilip edilmeyeceği konusunda destek sağlayabilir.

Bu çalışmada Bayes Ağları kullanarak, tüp bebek tedavi sürecinde embriyo morfolojik gelişim değerleri modellenerek blastosist skorları tahminlenmiştir. Çalışmada koşullu olasılık tablosundaki frekans tahminlerini ayarlamak için ağırlıklı en yakın

komşu yaklaşımı önerilmiştir. Sonuçlar önerilen modelin tüp bebek tedavisinde doğruluğu önemli ölçüde artırırken yanlış pozitif oranının frekans tahmini yöntemine göre düşük olduğunu göstermektedir. Bunun yanında model düşük kaliteli blastosist gelişimini %77.3 oranıyla doğru negatif tahmin etmektedir. Bu da modelin kullanılmasının tüp bebek tedavisinde embriyo gelişimsel başarısızlığını ciddi ölçüde önlemeye yardımcı olacağını göstermektedir.

Anahtar Kelimeler	**:**	*Tüp Bebek Tedavisi, Blastosist Gelişim Tahminlemesi, Bayes Ağlar, Parametre Öğrenimi, Frekans Tahminleri*

1. INTRODUCTION

In-vitro fertilization (IVF) has been a common infertility treatment method since 1978 (Steptoe & Edwards, 1978). In the IVF process, female germ cells (oocytes) are inseminated by sperm in IVF laboratories and embryos are cultured during a period of 2 to 6 days. Embryonic growth is observed and recorded by embryologists. Finally, selected embryo(s) is (are) transferred into the woman's womb. IVF embryos may be transferred either at the cleavage stage (day 2-3) or at the blastocyst stage (day 5-6).

Extended culture until the blastocyst stage allows for the self-selection of the most viable embryos since not all embryos can reach this stage in in-vitro conditions. Delaying the transfer until day 5 increases the implantation probability. On the other hand, it also increases the risk of developmental failure. Consequently, the prediction of blastocyst development is an important research question in the IVF domain.

In this research, we use Bayesian networks for modeling the morphological evolution of IVF embryos and predicting blastocyst development. We aim to encode statistical relations between the variables of interest throughout the stages of embryonic growth.

Learning a Bayesian network from data involves two subtasks, structure learning, which is required to identify the topology of the network, and parameter learning, which identifies the statistical parameters (conditional probabilities) for a given network topology.

Here we construct the topology of the network using a mutual-information-based preconditioning, and we propose a nearest-neighbor-based approach for adjusting the frequency estimates in the parameter learning stage. Experimental results show that the proposed approach significantly improves the classification performance in our IVF dataset.

Such a model can be used as part of the clinical procedure in order to prevent the wasting of embryos due to a possible developmental failure when they are further cultured in the laboratory. If embryos are predicted to result in low quality blastocysts on day 5, clinicians may decide to transfer or freeze them earlier on day 3.

2. PROBLEM STATEMENT

Obtaining many embryos is possible at each cycle of the IVF treatment; however, generally, the three highest quality embryos are transferred to the woman's uterus. Multiple embryo transfers increase the pregnancy probability, but they also increase potential complications of multiple pregnancies (Gerris & De Neubourg, 2005; Irmawati et al., 2019; Martikainen et al., 2004; Thurin et al., 2004; Veleva et al., 2006) Thus, it is aimed to maximize success rates with single embryo transfers by improving embryo selection (Zhan et al., 2020).

Elective single embryo transfer (eSET) has been favored as a solution to the IVF multiple pregnancy problem. Clinicians perform eSET at blastocyst stage in a much safer way because extended culture until the blastocyst stage allows the self-selection of the most viable embryos since not all embryos can reach this stage in in-vitro conditions.

2.1. Blastocyst Stage Transfer

The transfer of blastocyst-stage embryos on day 5 is thought to yield embryos with high implantation potential, increasing implantation and pregnancy rates in IVF treatment. When equal number of embryos are transferred, it is suggested that the probability of live birth is significantly higher after blastocyst-stage embryo transfer on Day 5 as compared to cleavage- stage embryo transfer on Day 2 or Day 3 (Papanikolaou et al., 2008). It is also recommended that in patients with a top-scoring blastocyst, the transfer of a single blastocyst should be considered (Gardner et al., 2000) preventing possible complications of multiple pregnancies. However, an extended culture of IVF embryos may result in transfer cancelation if no blastocyst develops.

2.2.Prediction of Blastocyst Score

When a further culture of embryos until Day 5 with the expectation of good quality blastocyst development is considered, a tradeoff exists between the higher probability of implantation success and the risk of transfer cancelation. If the development of blastocysts is predicted, the risk of transfer cancelation can be minimized. Different scoring systems for blastocysts-stage variables are developed for the selection of the best embryo in the development stage(Blank et al., 2020).

A cycle based model has been applied to predict blastocyst transfer cancelation (Dessolle et al., 2010). In a cohort of at least 5 good quality embryos, the authors have proposed a model to predict if any blastocyst would develop or not. This model is useful in preventing transfer cancelation; however, there are limitations related to the requirements of the model since it can be applied to specific cycles only.

Clinicians need reliable models to predict blastocyst development for individual embryos considering the tradeoff between increasing pregnancy rate and the possibility of transfer cancelation. It is necessary to model the entire embryo growth process in order to determine the relationships between the daily morphological variations of embryos.

2.3. Embryo Growth Process

Figure 1 represents the developmental stages of IVF embryos day by day. The initial state is considered to be the ICSI insemination process. A fertilization check is performed at 16-18 hours after ICSI process. Early cleavage morphology is observed on Day 1. The number of cells, nucleus characteristics, the fragmentation rate, the equality of blastomeres, and the appearance of the cytoplasm are graded on Day 2 and Day 3. Finally, if the embryo is decided to be cultured until Day 5, the morphology of the blastocyst is evaluated by using the Gardner scoring system (Gardner et al., 2004).

3. PROPOSED SOLUTION

The researchers are still investigating the statistical properties of the morphological evolutions of embryos and the interdependency of embryo development and patient characteristics. The literature presents conflicting results concerning predictive factors and their correlations. Therefore, as a starting point, we need to construct a model to analyze all available features and their statistical relations to blastocyst morphology.

A Bayesian Network is a graphical model that encodes conditional dependencies among variables of interest (Heckerman, 2020). In this study, we use Bayesian networks in analyzing the statistical relationships between the sequential observations of embryo morphology and predicting the blastocyst score. We consider the prediction of the IVF blastocyst score as a binary supervised classification problem to discriminate blastocysts into two classes as high quality (having a Gardner's score ≥ 3AA) and low quality ones. We construct an embryo-based dataset including daily morphological observations and patient and cycle characteristics.
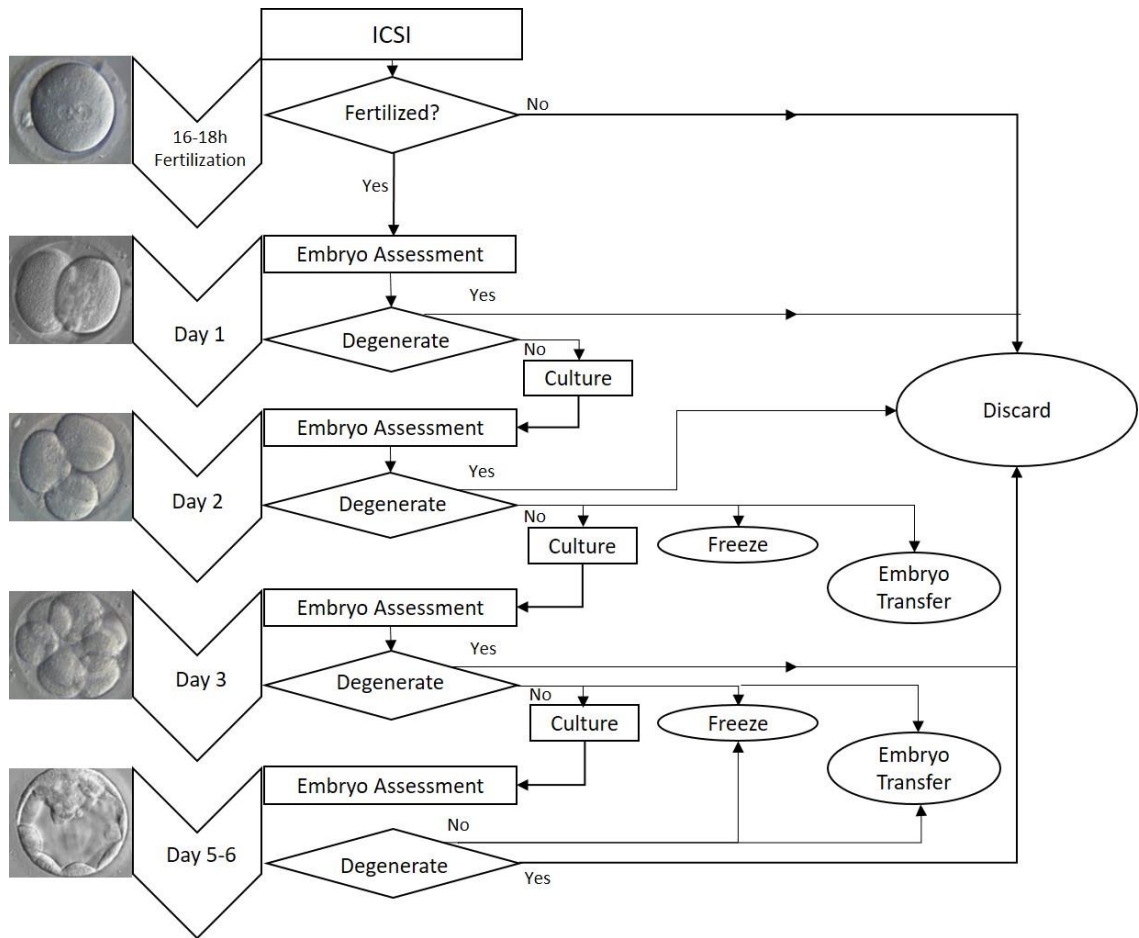
Figure 1: Demonstration of embryo growth process together with related embryo morphological variables. Time stamps correspond to the routine morphological observations performed in the IVF laboratory.

4. DATASET

Due to social, ethical and financial reasons some legislative rules have been defined for assisted reproduction process in every country. The restrictions usually apply to donation, embryo manipulation, the number of embryos to be transferred in each cycle etc.

Along with the legal procedures effect in different countries, each IVF clinic, including those in the same country, applies different technologies and methodologies. Because of this variety, each clinic has distinctive IVF databases. In this research, we analyze the dataset that has been used in a previous study (Uyar et al., 2010).
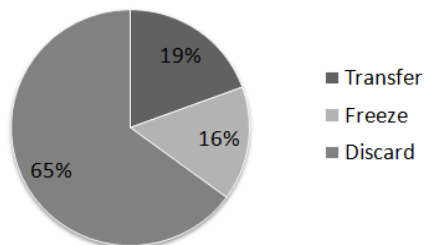
Distribution of fertilized oocytes

Figure 2: Distribution of transferred frozen and discarded embryos.

The raw dataset includes a total of 81371 oocytes. Among 62800 fertilized oocytes, 12185 embryos have been transferred and 9858 embryos have been frozen (Figure 2). The remaining 40757 embryos, which constitute 64.9% of the fertilized oocytes, have been discarded due to developmental failure. This rate can be reduced by using accurate prediction models supporting the decision concerning the extended culture of embryos although the degeneration of the embryos cannot be totally prevented.

Table 1: Selected dataset features for each blastocyst feature vector

Dataset Features	Data Type
Patient and Cycle Characteristics	
Woman age	Continuous
Gravidity	Categorical
Infertility factor	Categorical
Treatment protocol	Categorical
Duration of stimulation	Continuous
Follicular stimulating hormone dosage	Continuous
Peak Estradiol level	Continuous
Endometrium thickness	Continuous
Sperm quality	Categorical
Embryo Related Data	
Early cleavage morphology	Categorical
Early cleavage inspection time	Continuous
Number of cells (day 2-3)	Categorical
Nucleus characteristics (day 2-3)	Categorical
Fragmentation (day 2-3)	Categorical
Blastomeres (day 2-3)	Categorical
Appearance of cytoplasm (day 2-3)	Categorical

A total of 9043 embryos have been cultured until the blastocyst stage. We have eliminated the records including missing values. Finally, a total of 7735 blastocysts have been

analyzed where 1779 blastocysts have been developed with a Gardner's score ≥3AA (23.0%) (Gardner et al., 2004).

We have included the available features based on the literature and expert judgement. The list of features is provided in Table 1.

5. METHODOLOGY

In this section, we briefly summarize the methods we used in our experiments.

5.1. Bayesian Networks

The structure of the Bayesian network is used to characterize a probability distribution for each node depending on its parents and posterior probabilities are computed in the form of local conditional distributions. A brief definition of Bayesian networks and Bayesian network classifiers (Friedman et al., 1997) is given below:

A Bayesian network is represented by $B = (G, \Theta)$, where G is a directed acyclic graph. The nodes of the graph correspond to the random variables $X_1, ...X_n$ which are the dataset features and edges represent the direct dependencies between the associated variables. The graph G encodes the independence assumption where each variable X_i is independent of its non-descendants given its parents Π_{X_i} in G. The second component Θ represents the conditional probability distribution that quantifies the dependency between the nodes.

A Bayesian network defines a unique joint probability distribution over the set of random variables X_i in the network given by:

$$P(X_1, ... , X_n) = \prod_{i=1}^{n} P(X_i | \prod X_i)$$

(1)

where Π_{X_i} denotes the set of parents of X_i in the network.

In practice, the components of the Bayesian networks are generally unknown and must be inferred from the data. Learning a Bayesian network from data involves two subtasks: structure learning, which is required to identify the topology of the network, and parameter learning, which identifies the statistical parameters (conditional probabilities) for a given network topology.

Most studies concentrate on structure learning which is a complex procedure when there are lots of input features (Cheng et al., 2002; Csató & Reiz, 2008; Meloni et al., 2009). Learning the parameters in conditional probability tables is recognized as a trivial task based on frequency counts of data points when the observed frequencies are optimal in a sufficiently large database (Cheng et al., 2002). Here, we review the main approaches for construction of

the network structure and estimation of parameters when learning Bayesian networks from data.

5.1.1. Structure Learning

Structure learning is a search for encoding appropriate dependencies between the features of a given a dataset. It has been argued that Bayesian network structure learners are computationally expensive and require an exponential number of conditional independence tests (Cheng et al., 2002). There are two main approaches to learning the network structure from data efficiently through reducing the search space: constraint-based methods and methods that maximize a selected score.

The simple learning algorithm (SLA) and three-phase dependency analysis (TPDA) are examples of constraint based methods that make use of the information theory concept in order to reduce the computational complexity of the structure learning procedure (Cheng et al., 2002). Csató and Reiz also propose a mutual information based approach where direct causal relations encoded by the Bayesian network are interpreted as the maximal conditional mutual information between nodes (Csató & Reiz, 2008).

The algorithms that maximize the selected score search for the optimum structure by evaluating how well a given network matches the data. Meloni *et al.* propose a variation of the standard search-and-score approach that computes a square matrix containing the mutual information among all pairs of variables (Meloni et al., 2009). The matrix is binarized to find out which relationships must be suppressed in order to prevent the inference of too many connections.

Furthermore, Naive Bayesian network, which assumes mutual independence of the feature variables given the class variable, and the Tree Augmented Network (TAN), which represents a tree-like dependency structure over the feature variables, are well-known Bayesian network structures (Lucas, 2004).

In our experiments, we construct a constraint-based Naive Bayesian network structure using the mutual information between nodes.

Information Gain Feature Weighting: Information Gain represents the average amount of information about the class value C contained in the feature value F (Mladenić & Grobelnik, 2003). Information Gain is also known as mutual information between F and C.

$$InfoGain(F) = I(C, F) = H(C) - H(C|F)$$

$$(2)$$

where

$$H(C) = \sum_i P(C_i) \log_2 P(C_i)$$

$$(3)$$

is the Shannon's entropy.

Higher Information Gain means a higher predictor effect of the feature individually. The Information Gain values of features provide reasonable knowledge required to reduce the search space for feature subset selection.

The features with an Information Gain value less than a pre-defined threshold are selected as the input parameters in the structure learning phase. For example, the threshold can be defined as the average of the Information Gain of all of the features, $\mu_{IG}(F)$. Then,

add F_i to S if $InfoGain(F_i) < \mu_{IG}(F)$

5.1.2. Parameter Learning

Parameter learning in Bayesian networks is often based on Frequency Estimate (FE) which determines the conditional probabilities by computing the frequencies of instances from the data. The FE method is efficient since it counts each data point in the training set only once. The parameters estimated using the FE method maximize the likelihood of the model given the data and thus FE is known as a generative learning method (Su et al., 2008).

The relative frequencies in the conditional probability table (CPT) are obtained as follows:

$$\hat{P}(X_i = x | \prod X_i = \vec{u}) = \frac{count\ (X_i = x | \prod X_i = \vec{u})}{count(\prod X_i = \vec{u})}$$

$$(4)$$

In our case, X_i denotes the class label as the child node that is the blastocyst score and $\vec{u}$ denotes a vector of parent nodes $\prod x_i$ representing the predictor factors affecting the blastocyst score.

The classification capability of FE method is debated because of its generative property. Grainer and Zhou proposed a gradient descent based discriminative parameter learning method, which significantly outperforms the FE method with a high computational cost(Greiner et al., 2005).

A Discriminative Frequency Estimate (DFE) is proposed to maximize the generalization accuracy of classification rather than likelihood (Su et al., 2008). The authors compared the DFE and FE methods based on the Naive Bayesian network structure and showed that the DFE significantly improved the performance of classification in terms of accuracy. However, it has been widely accepted that accuracy is not an appropriate performance measure especially for imbalanced datasets. On the other hand, the training time of the DFE method is significantly higher than the FE method. Consequently, an efficient and effective method for parameter learning in Bayesian networks is still an open question.

We propose a method for parameter learning from data taking advantage of the efficient FE method and handling the insufficiencies in the data.

5.1.3. Proposed Nearest Neighbor Based Approach for Adjusting Frequency Estimates

When the frequencies of each possible combination of feature values are computed, we can identify the samples that occur less than a predefined threshold of the sample size. Then, finding the nearest neighbors of those samples constitutes a cluster in the neighborhood of the infrequent sample. In this case, rather than computing the conditional probabilities for each feature vector, we can compute a common conditional probability entry for the cluster of feature value combinations.

The idea behind this approach is as follows: any combination of feature values may be represented insufficiently in the training data. This fact may shadow the real statistical properties of the nodes in the Bayesian Network. By clustering the less frequent samples up to a certain level, it may be possible to obtain more accurate conditional probabilities. However, it is crucial to avoid the uniformity of conditional probabilities that would lead to information loss. Therefore, there are two critical hyper-parameters in the proposed approach:

t_u: which represents the level of insufficiency in terms of the frequency of feature vectors, and

t_l: which represents the sufficient number of samples in the neighborhood of less frequent samples.

The thresholds should be determined in training phase using a grid search method that utilizes a pre-defined set of values for each threshold parameter. The search space depends on the estimated frequencies in the conditional probability table.

When computing the distance between two instances in the nearest- neighbor approaches, all the features may not have equal impact on the similarity measure. Therefore,

identification of relative effects of the features on the distance can improve a nearest neighbor learning process (Kohavi et al., 1997; Vivencio et al., 2007).

Feature weighting algorithms may be used to identify the relative effects of features on the outcome. We use Information Gain feature weighting algorithm to rank the features of the dataset, and the ranked list of features is then used to define a feature weighting vector embedded in the Euclidean distance metric.

In this research, the nearest-neighbor approach is used for finding the most similar cases to samples which were represented less frequently in the training dataset. The weighted Euclidean distance between the instances x_i and x_j, $d_w(i, j)$ is:

$$d_w(i,j) = sqrt\left(\sum_{k=1}^{n} \left(1/w_k\right) * \left(x_{jk} - x_{ik}\right)^2\right)$$

(5)

where, n is the number of features and w_k is the pre-evaluated Information Gain ranking of the k_{th} feature.

When the cluster of the nearest neighbors that includes the sufficient size of samples is obtained, the conditional probabilities that average the probabilities of the samples in the cluster are computed.

The pseudocode given in Algorithm 1 outlines the structure learning strategy that we used in network construction and our proposed approach for the parameter learning.

Algorithm 1: Pseudocode for adjusted CPT entries

1: F = [Set of input features]

2: C = class variable

3: %Subset selection for Naive Bayesian network structure.

4: S =ø

5: **for all** f in F **do**

6:       compute *IG* (*f*) = *InfoGain*(*f, C*)

7:       **if** *IG*(*f*) ≥ μ_{IG}(*F*) **then**

8:         *S = S* U *f*

9:       **end if**

10: **end for**

11: %Frequency estimates $n(\Pi_C = \vec{u})$ and adjusted frequency estimates $\hat{n}(\Pi_C = \vec{u})$

12: %t_u upper bound for insufficient frequency and t_l lower bound for sufficient number of data points in clustered neighborhood

13: **for all** $\vec{u}$ in S **do**

14:       **if** $n(\Pi_C = \vec{u}) < t_u$ **then**

15:         $\hat{n}(\Pi_C = \vec{u}) = n(\Pi_C = \vec{u})$

16:         **while** $\hat{n}(\Pi_C = \vec{u}) < t_l$ **do**

17:           $\hat{n}(\Pi_C = \vec{u}) = \hat{n}(\Pi_C = \vec{u}) + n(WeightedNearestNeighbors(\vec{u}))$

18:         **end while**

19:       **end if**

20: **end for**

6. EXPERIMENTAL RESULTS

The network is visualized using Netica software (https://www.norsys.com/netica.html). Initially we have used a feature selection based on the Information Gain feature weighting as described in Section 5.1.1. The estimated weights given in Figure 3 have been used to define the network structure and to evaluate the weighted nearest neighbors in the parameter learning stage.

The initial network structure is given in Figure 4 where class 2 represents the high-quality blastocysts.

In the experiments, the CPT entry of the feature vectors that has less than 50 samples (t_u) in the trainset have been accepted as insufficient frequency estimates. The proposed nearest neighbor based approach has been used to cluster the insufficiently represented CPT feature vectors to constitute a cluster of at least 200 samples (t_l) in the trainset. The resulting probabilities are shown in Table 2.

Random two-thirds of the dataset is used for training and the remaining one-third is used for testing. Stratified random splitting of the data into training and test sets is repeated 10 times in order to avoid sampling bias. Stratified random splitting ensures that the proportion of positive and negative instances are the same in training and test sets.
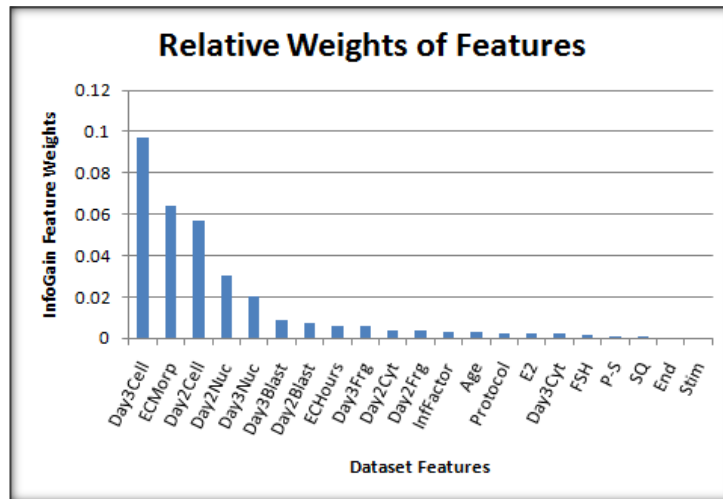

Figure 3: Information Gain feature weights. Number of cells in day 3 has the highest feature weights among all data features.

Day3Cell: Number of cells in day 3, ECMorp: Early cleavage morphology, Day2Cell: Number of cells in day 2, Day2Nuc: Nucleus characteristics on day 2, Day3Nuc: Nucleus characteristics on day 3, Day3Blast: Blastomeres on day 3, Day2Blast: Blastomeres on day 2, ECHours: Early cleavage hours, Day3Frg: Fragmentation on day 3, Day2Cyt: Appearance of cytoplasm on day 2, Day2Frg: Fragmentation on day 2, InffFactor: Infertility factor, E2: Estradiol hormone, Day3Cyt: Appearance of cytoplasm on day 3, FSH: Follicle-

Stimulating Hormone, P-S: Primer/Seconder, SQ: Sperm Quality, End: Endometrium thickness, Stim: Duration of stimulation
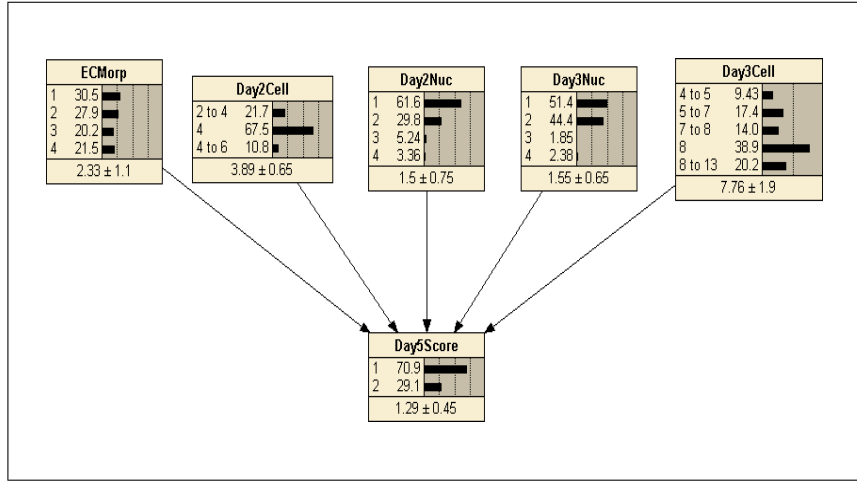

Figure 4: Bayesian Network structure for blastocyst score prediction.

ECMorp: Early cleavage morphology, Day2Cell: Number of cells in day 2, Day2Nuc: Nucleus characteristics on day 2, Day3Nuc: Nucleus characteristics on day 3, Day3Cell: Number of cells in day 3

Table 2: Initial probabilities in the CPT*(Prob.) and the updated probabilities (UProb.)

ECMorp	D2Cell	D2Nuc	D3Nuc	D3Cell	Freq.	C1	C2	Prob	Uprob
1	1	1	1	1	98	98	0	1	1
1	1	1	1	2	8	8	0	1	0,921
1	1	1	1	3	3	3	0	1	0,759
1	1	1	1	4	7	6	1	0,857	0,738
1	1	1	1	5	3	3	0	1	0,705
1	1	1	2	1	39	38	1	0,974	0,947
1	1	1	2	2	48	47	1	0,978	0,924
1	1	1	2	3	13	12	1	0,923	0,87
1	1	1	2	4	2	2	0	1	0,792
1	1	1	2	5	2	2	0	1	0,774
1	1	1	3	1	1	1	0	1	0,919
1	1	1	3	2	1	1	0	1	0,89
1	1	1	3	3	0	0	0	0,5	0,866
1	1	1	3	4	0	0	0	0,5	0,789
1	1	1	3	5	0	0	0	0,5	0,767
1	1	1	4	1	4	4	0	1	0,798
1	1	1	4	2	1	1	0	1	0,796
1	1	1	4	3	2	2	0	1	0,777
1	1	1	4	4	0	0	0	0,5	0,765
1	1	1	4	5	0	0	0	0,5	0,758
1	1	2	1	1	51	51	0	1	0,981
1	1	2	1	2	46	45	1	0,978	0,924

* Conditional Probability Table

Table 3: Comparison of the initial network (Network1) using FE** and the network with updated CPT* (Network2) using our proposed approach for parameter learning

Network	Accuracy (%)	TP Rate (%)	FP Rate (%)
Network 1	69.1 ± 2.9	59.4 ± 7.5	29.4 ± 6.6
Network 2	72.6 ± 1.7	58.7 ± 4.8	22.7 ± 1.4

* Conditional Probability Table

** Frequency Estimate

The results are given in Table 3 in terms of accuracy, true positive (TP) rate (sensitivity) and false positive (FP) rate. Since the dataset represents an imbalanced distribution of the two classes of blastocysts, the decision threshold is optimized to handle the imbalance problem and decided as 0.7, mapping to the point closest to the upper left corner on the ROC curve.

Paired t-tests indicate that the networks produce significantly different results in terms of accuracy and FP rate ($p < 0.05$). Network 2 with updated CPT reduce the false positive predictions as required in clinical procedure. This would result in reducing the number of degenerated embryos at blastocyst stage.

7. CONCLUSION

In this paper we modeled the embryo growth process using Bayesian Networks with the aim of predicting the blastocyst score. The results of the FE method were relatively lower that motivated us to analyze the data and the methods. We recognized that although we have a sufficiently large dataset, the observed frequency estimates are not optimal and we proposed a nearest- neighbor approach to cluster the insufficient data points.

There are two hyper-parameters of the proposed model: threshold-1 that indicates the lower bound for insufficient frequencies and threshold-2 that indicates the upper bound for the sufficient number of training instances in the neighborhood of the infrequently represented data points. The optimum values of these two parameters depend on the distribution of training instances in the conditional probability table and size of the dataset. Adjustment of the thresholds is critical for the success of the proposed model.

The main assumption underlying our proposed model is that infrequent or missing data points in training set can be clustered in a neighborhood to produce a more accurate collective frequency estimate for all of the instances in the associated cluster. The proposed model will work well, but if this assumption does not hold, the prediction performance of the frequency estimate will not change significantly.

Experimental results show that our model can predict a potential low quality blastocyst development at 77.3% True Negative rate. This can be interpreted as follows: If clinicians use such a model in the laboratory, 77.3% percent of the developmental failure of embryos from Day 3 to Day 5 can be prevented.

As a future work, the algorithm represented in this study can be replicated in additional datasets in biomedical field and other fields to improve prediction performance in Bayesian Network classification. Varying distribution of data points in different datasets would help optimizing the clustering approach to better represent conditional probabilities in network construction.

REFERENCES

Blank, C., DeCroo, I., Weyers, B., van Avermaet, L., Tilleman, K., van Rumste, M., de Sutter, P., Mischi, M., & Schoot, B. C. (2020). Improvement instead of stability in embryo quality between day 3-5: A possible extra predictor for blastocyst selection. *European Journal of Obstetrics and Gynecology and Reproductive Biology*, *253*, 198–205. https://doi.org/10.1016/j.ejogrb.2020.08.027

Cheng, J., Greiner, R., Kelly, J., Bell, D., & Liu, W. (2002). Learning Bayesian networks from data: An information-theory based approach. *Artificial Intelligence*, *137*(1–2), 43–90. https://doi.org/10.1016/S0004-3702(02)00191-1

Csató, L., & Reiz, B. (2008). Tree-like Bayesian Network classifiers for surgery survival chance prediction. In *Article in International Journal of Computers: Vol. III*. https://www.researchgate.net/publication/228634935

Dessolle, L., Fréour, T., Barrire, P., Daraï, E., Ravel, C., Jean, M., & Coutant, C. (2010). A cycle-based model to predict blastocyst transfer cancellation. *Human Reproduction*, *25*(3), 598–604. https://doi.org/10.1093/humrep/dep439

Friedman, N., Geiger, D., & Goldszmidt, M. (1997). Bayesian Network Classifiers. *Machine Learning*, *29*(2–3), 131–163. https://doi.org/10.1023/a:1007465528199

Gardner, D. K., Lane, M., Stevens, J., Schlenker, T., & Schoolcraft, W. B. (2000). Blastocyst score affects implantation and pregnancy outcome: towards a single blastocyst transfer. *Fertility and Sterility*, *73*(6), 1155–1158. https://doi.org/10.1016/S0015-0282(00)00518-5

Gardner, D. K., Surrey, E., Minjarez, D., Leitz, A., Stevens, J., & Schoolcraft, W. B. (2004). Single blastocyst transfer: A prospective randomized trial. *Fertility and Sterility*, *81*(3), 551–555. https://doi.org/10.1016/j.fertnstert.2003.07.023

Gerris, J., & De Neubourg, D. (n.d.). *Single embryo transfer after IVF/ICSI: present possibilities and limits*.

Greiner, R., Su, X., Shen, B., & Zhou, W. (2005). Structural extension to logistic regression: Discriminative parameter learning of belief net classifiers. *Machine Learning*, *59*(3), 297–322. https://doi.org/10.1007/s10994-005-0469-0

Heckerman, D. (2020). A Tutorial on Learning With Bayesian Networks. *Studies in Computational Intelligence*, *156*, 33–82. http://arxiv.org/abs/2002.00269

Irmawati, Basari, & Gunawan, D. (2019). Automated Detection of Human Blastocyst Quality Using Convolutional Neural Network and Edge Detector. *2019 1st International Conference on Cybernetics and Intelligent System, ICORIS 2019*, 181–184. https://doi.org/10.1109/ICORIS.2019.8874925

Kohavi, R., Langley, P., & Yun, Y. (n.d.). *The Utility of Feature Weighting in Nearest-Neighbor Algorithms*.

Lucas, P. J. F. (2004). *Restricted Bayesian Network Structure Learning* (pp. 217–234). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-39879-0_12

Martikainen, H., Orava, M., Lakkakorpi, J., & Tuomivaara, L. (2004). Day 2 elective single embryo transfer in clinical practice: Better outcome in ICSI cycles. *Human Reproduction*, *19*(6), 1364–1366. https://doi.org/10.1093/humrep/deh197

Meloni, A., Ripoli, A., Positano, V., & Landini, L. (2009). Mutual information preconditioning improves structure learning of bayesian networks from medical databases. *IEEE Transactions on Information Technology in Biomedicine*, *13*(6), 984–989. https://doi.org/10.1109/TITB.2009.2026273

Mladenić, D., & Grobelnik, M. (2003). Feature selection on hierarchy of web documents. *Decision Support Systems*, *35*(1), 45–87. https://doi.org/10.1016/S0167-9236(02)00097-0

Papanikolaou, E. G., Kolibianakis, E. M., Tournaye, H., Venetis, C. A., Fatemi, H., Tarlatzis, B., & Devroey, P. (2008). Live birth rates after transfer of equal number of blastocysts or cleavage-stage embryos in IVF. A systematic review and meta-analysis. *Human Reproduction*, *23*(1), 91–99. https://doi.org/10.1093/humrep/dem339

Steptoe, P. C., & Edwards, R. G. (1978). Birth after the reimplantation of a human embryo. In *Lancet* (Vol. 2, Issue 8085, p. 366). Lancet. https://doi.org/10.1016/s0140-6736(78)92957-4

Su, J., Zhang, H., Ling, C. X., & Matwin, S. (2008). Discriminative parameter learning for Bayesian networks. *Proceedings of the 25th International Conference on Machine Learning*, 1016–1023. https://doi.org/10.1145/1390156.1390284

Thurin, A., Hausken, J., Hillensjö, T., Jablonowska, B., Pinborg, A., Strandell, A., & Bergh, C. (2004). Elective Single-Embryo Transfer versus Double-Embryo Transfer in in Vitro Fertilization. *New England Journal of Medicine*, *351*(23), 2392–2402. https://doi.org/10.1056/nejmoa041032

Uyar, A., Bener, A., Ciray, H. N., & Bahceci, M. (2010). Bayesian networks for predicting IVF blastocyst development. *Proceedings - International Conference on Pattern Recognition*,

2772–2775. https://doi.org/10.1109/ICPR.2010.679

Veleva, Z., Vilska, S., Hydén-Granskog, C., Tiitinen, A., Tapanainen, J. S., & Martikainen, H. (2006). Elective single embryo transfer in women aged 36-39 years. *Human Reproduction*, *21*(8), 2098–2102. https://doi.org/10.1093/humrep/del137

Vivencio, D. P., Hruschka, E. R., Do Carmo Nicoletti, M., Dos Santos, E. B., & Galvão, S. D. C. O. (2007). Feature-weighted k-nearest neighbor classifier. *Proceedings of the 2007 IEEE Symposium on Foundations of Computational Intelligence, FOCI 2007*, 481–485. https://doi.org/10.1109/FOCI.2007.371516

Zhan, Q., Sierra, E. T., Malmsten, J., Ye, Z., Rosenwaks, Z., & Zaninovic, N. (2020). Blastocyst score, a blastocyst quality ranking tool, is a predictor of blastocyst ploidy and implantation potential. *F&S Reports*, *1*(2), 133–141. https://doi.org/10.1016/j.xfre.2020.05.004