



HAL
open science

Predicting in vitro fertilization from ultrasound measurements using Machine Learning techniques

Zeinab Abbas, Sébastien Ménigot, Jamal Charara, Zein Ibrahim, Chadi Fakih,
Jean-Marc Girault

► **To cite this version:**

Zeinab Abbas, Sébastien Ménigot, Jamal Charara, Zein Ibrahim, Chadi Fakih, et al.. Predicting in vitro fertilization from ultrasound measurements using Machine Learning techniques. ICABME, Oct 2023, Beyrouth, Lebanon. pp.190-194, 10.1109/ICABME59496.2023.10293011 . hal-04243615

HAL Id: hal-04243615

<https://hal.science/hal-04243615>

Submitted on 16 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Predicting in vitro fertilization from ultrasound measurements using Machine Learning techniques

Zeinab Abbas
LAUM
Le Mans University
Le Mans, France
Zeinab.Abbas.Etu@univ-
lemans.fr

Sébastien Ménigot
ESEO (Angers)
Angers, France
Sebastien.MENIGOT@eseo.fr

Jamal Charara
Department of Physics &
Electronics
Faculty of science
Lebanese University
Beirut, Lebanon
jcharara@ul.edu.lb

Zein Ibrahim
Department of Informatics
Faculty of science
Lebanese University
Beirut, Lebanon
ibrahim.zein@gmail.com

Chadi Fakih
Obstetrics & Gynecologist
Al Hadi Center
Beirut, Lebanon
drchadifakih@yahoo.fr

Jean-Marc Girault
line 2: dept. name of organization
ESEO (Angers)
Angers, France
Jean-Marc.Girault@eseo.fr

Abstract— Predicting the implantation outcomes of in vitro fertilization (IVF) is important to the treatment's success. Therefore, machine-learning techniques including Support Vector Machines (SVM), Gradient Boosting (GB), Extreme Gradient Boosting (XGB), and Random Forest (RF) were tested for their ability to predict IVF pregnancy success using a combination of Doppler and clinical parameters. Forecasting ability was evaluated utilizing widely known performance indicators such as accuracy rate, sensitivity, specificity, and AUC. The GB combined with Random Forest Importance Feature selection technique showed the highest performance with a sensitivity of 100%, a specificity of 66%, and an accuracy of 82.3% compared to other techniques that tried to predict the outcome of IVF and Intracytoplasmic sperm injection (ICSI) using machine learning techniques. Ultrasound measurement parameters and especially Doppler parameters are an important factor that affects the outcome of IVF.

Keywords— *In Vitro Fertilization (IVF), Intra Cytoplasmic Sperm Injection (ICSI), Machine Learning, Gradient Boosting (GB), Support Vector Machine (SVM), Random Forest (RF), Extreme Gradient Boosting (XGB), Doppler parameters.*

I. INTRODUCTION

Infertility is defined as a couple's biological inability to have children following at least 12 months of regular, well-timed sexual intercourse without the use of birth control. IVF is a common infertility treatment approach in which sperm inseminates female germ cells (oocytes) in the laboratory [1]. Machine-learning algorithms have been proposed to automate the detailed modeling of the IVF technique for result prediction and patient counseling to assist practitioners in making educated decisions. These models are capable of investigating and analyzing the population pattern in a large dataset and can aid in the extraction of latent knowledge from the dataset. Machine

learning techniques such as Artificial Neural Networks (ANN), Support Vector Machines (SVM), Decision Tree (DT), Random Forest (RF), and many more are commonly utilized for prediction [2]. Uyar et al. [1] investigated a range of classifiers to predict the outcome of embryo implantation in IVF. They examined the performance of the classifiers using the Receiver Operating Characteristics (ROC) curve. The NB and RBF classifiers produced better results. Furthermore, Durairaj and Kumar [3] employed 14 male and female characteristics to predict IVF treatment success rate, resulting in an overall accuracy of 73%. In 2019, Qui et al. [4] compare predictive performance, based on the AUC. SVM, RF, and Extreme Gradient Boosting (XGBoost) outperformed classical logistic regression in tailored predictions of live births prior to the first IVF session. Recently, Zhang et al. [5] demonstrated the benefit of adding new Echo-Doppler parameters to clinical parameters. Nevertheless, these authors were not the only ones to have been interested in the use of ultrasound parameters. Indeed, as early as 2012, ultrasound assessment has been used and found to be useful in predicting outcomes in assisted reproduction patients [6] [7]. Ultrasound examination of endometrial receptivity can detect a variety of combinations of imaging parameters. The prediction models used a variety of ultrasound features, including endometrial volume and endometrial vascularization index [5]). Only one of the studies mentioned above used endometrial and subendometrial blood flow as measures to demonstrate the correlation with IVF. However, Myometrial receptivity has not been addressed in [5]. Indeed, Zhang et al. [5] combined clinical and Doppler parameters to predict the outcomes in embryo transfer cycles using flow index (FI), resistivity index (RI), and pulsatility index (PI), as Doppler parameters. The AUC was 0.698 throughout the training procedure. Furthermore, none of the preceding investigations used machine learning methodologies with endometrial and

myometrium Doppler information. Therefore, the purpose of this study was to (i) develop a machine learning-based system to better estimate the ideal endometrial implantation period to predict the outcome of IVF and (ii) use a combination of the most influential IVF clinical attributes and ultrasound imaging characteristics to obtain complementary benefits. The rest of the paper is structured as follows. The collection of data used in this study is first described in section 2. We give a brief explanation of the method in section 3. In section 4, we give the findings demonstrating the classifiers' ability to predict ICSI outcomes, their analyses, and finally a discussion and conclusion for the study.

II. MATERIAL AND METHODS:

A. Data Collection

Data was acquired from all patients receiving ICSI treatments at ALHADI IVF Center (Center of Treatment by Intra Cytoplasmic Sperm Injection in Lebanon, Beirut, Lebanon) between January 2018 and March 2018. All medical acts were performed conventionally, with no additional or unusual diagnostic or treatment processes. As a result, in this circumstance, the documented analysis is retrospective research with no impact on patient care. The "Track md" database was used to record each ICSI cycle. In this study, 85 ICSI patients were used. Women under the age of 45 with at least one embryo of acceptable morphology were eligible. The clinical and Doppler data of women who had IVF/ICSI cycles at ALHADI IVF Center were analyzed using various exclusion criteria such as poor responders, patients with fibroids with a history of myomectomy, and if the patient has an abnormal uterine cavity anatomy. The ultrasound machine's high resolution, as well as the Pulsed and Color Doppler, enabled us to carefully measure all parameters. We used VOCAL Software to calculate endometrial and myometrium volume, as well as the vascularization index. Mean Gray (MG), Vascularization Index (VI), FI, and Vascularization Flow Index (VFI). Clinical parameters such as age, tentative, embryo and oocyte characteristics such as the number of embryo transfers, oocytes retrieved, oocytes placed in fertilizer, 2PN, utilization rate, embryos sampled and the quality of the embryo, and finally the day of transfer can be retrieved from the "Track md" application, which was filled out by the embryologist at "ALHADI IVF Center." Except for the quality of the embryo, which was a categorical data type that had to be transformed into a numerical one, all of these elements were numerical.

We prepared three files with identical data to compare the findings and demonstrate the importance of Doppler parameters. The first and second files contained Doppler and clinical parameters, respectively. All Doppler and clinical parameters were included in the third file. I followed the steps indicated below for each file.

B. Description of the work

After measuring the Doppler parameters, the raw data were preprocessed in Microsoft Excel using data classification tools before being transmitted to the machine learning program. Normalization is a typical data preparation technique used to help adapt datasets to a standard scale. It essentially facilitates

the conversion of values between 0 and 1 or -1 and +1. It's a good method for enhancing the performance, accuracy, and dependability of machine learning models. In general, machine learning models are built on mathematical formulas. As a result, while the equations only require integers, including categorical data in the equation will cause some issues. We used One Hot Encoding in Scikit-Learn, which is a function that allowed us to encode features-based categorical data into numerical ones, such as the quality of the embryo transfer, which was the dependent variable encoded 0 for embryo with poor quality (grade 1, grade 2, and grade 3) and 1 for embryo with good quality (grade 0). The training and testing data was split 80/20. The remaining 80% of the data was purely for training purposes. To produce a successful training result, stratified k-fold cross-validation was performed. The remaining 20% was just used to evaluate the performance of the final model. This procedure is useful for evaluating a prediction's model performance during training since it ensures nonbiased scoring of prediction outputs [8]. The classifier's performance can be improved by selecting a mixture of relevant features that reflect the greatest divergence between the classes. In this investigation, we used RF Importance with a threshold of 0.02 for each classifier. Then, to select the optimal hyper-parameter values, we used three different types of optimization algorithms and compared them for each classifier (Grid search, Bayesian search, and Random search). To predict pregnancy outcomes following ICSI, we assessed the effectiveness of four prediction models: RF, Extreme Gradient Boosting (XtremeGB), SVM, and Gradient Boosting (GB). Finally, the performance of each model is evaluated using several metrics to identify the best model to predict the outcome of ICSI, such as accuracy, confusion matrix, and AUC-ROC. This study's calculations were completed entirely in Python, which makes use of a variety of existing libraries that may be especially valuable for machine-learning tasks.

III. RESULTS

Table I shows the performance of each of the 4 models. GB gives an accuracy higher than 0.8. The highest results were with GB with 0.823 as an accuracy and 0.75 as an AUC. You can see that AUC for SVM and RF is 0.777 but with lower accuracy (0.647 and 0.705).

TABLE I: ACCURACY AND AUC VALUES FOR ALL MODELS.

Model	Accuracy	AUC
SVM	0.647	0.777
GB	0.823	0.75
RF	0.705	0.68
XGB	0.705	0.736

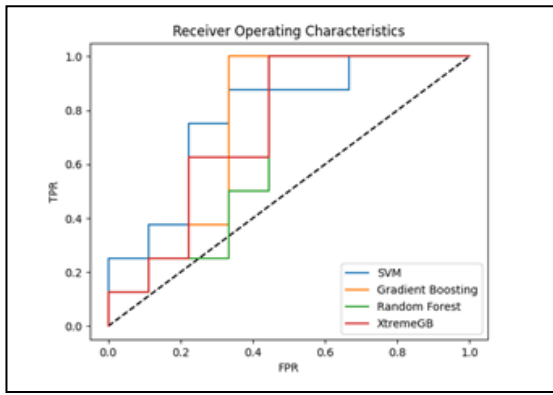


Fig.1: Figure Represents the ROC curve of all models

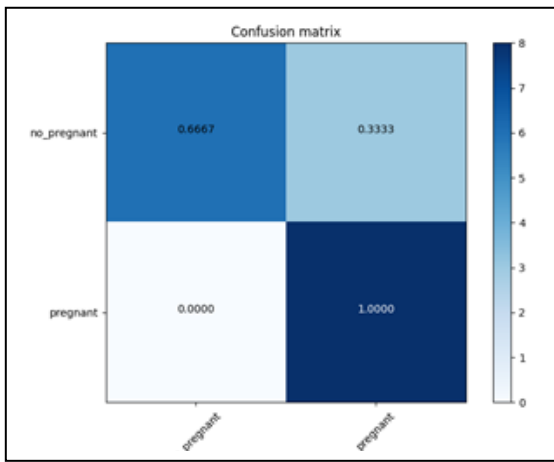


Fig.2: Confusion matrix of the GB model.

Experimental results (Fig.1 & Fig.2) showed that GB Algorithm, with RF importance feature selection technique with 0.02 as a threshold, outperformed all other algorithms in terms of accuracy (0.823) with an AUC of 0.75, a sensitivity of 100%, and a specificity of 66%.

Figures 3 and 4 represent histograms comparing the accuracy and AUC of three files. The blue bars (in both graphs below) show the accuracy and AUC of file 1, which contains Doppler parameters. The orange bars in the figures below indicate the accuracy and AUC of file 2, which provides clinical parameters. The grey bars in the graphs below reflect the accuracy and AUC of file 3, which contains the combination of Doppler and Clinical parameters. Doppler provides the lowest accuracy and AUC values for all models. It is obvious that when Doppler and clinical characteristics are utilized simultaneously, the models perform best in terms of accuracy and AUC.

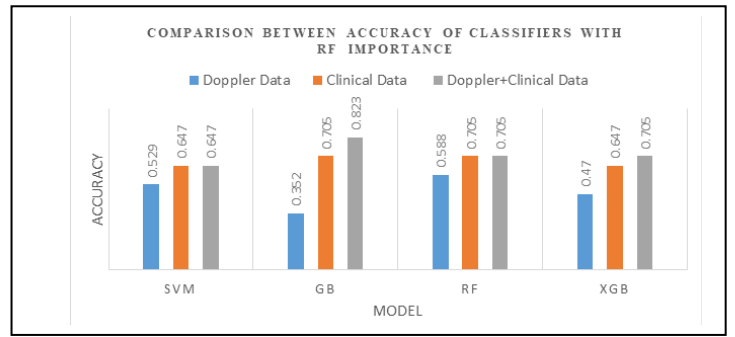


Fig. 3: Histogram representing the comparison of the Accuracy between Doppler, Clinical, and the Combination between Doppler and Clinical.

The accuracy of clinical parameters with RF was the same as with all parameters (0.705), although with a larger AUC for the file combining Doppler and Clinical. The results showed that combining Doppler and clinical parameters produced better outcomes than either Doppler or clinical parameters alone. As a result, the findings of our study could be used to assist clinicians in determining the efficacy of ICSI treatment.

IV. DISCUSSION

This study highlighted the significance of the classifier and the feature selection method based on the Tree learning algorithm. It was demonstrated that when the RF Importance significance feature selection technique was used to locate the relevant features, the outcome of ICSI utilizing the Gradient Boosting classifier could be predicted. In terms of performance and output, this model exceeds all other machine learning methods. Several of physiological and pathological uterine abnormalities can be detected using ultrasound imaging techniques. Infertile patients can benefit from personalized prediction algorithms that have been systematically developed and validated [5]. We tested the ability of simple clinical indicators, ultrasound measurement data, and a combination of these parameters to predict ICSI results. All significant factors must be evaluated to provide useful recommendations about a specific couple's case for pursuing ICSI therapy. One factor that may contribute to such a low success rate is a lack of adequate processes for predicting infertile couples. Using data from ICSI patient records, researchers aimed to construct the best decision-making algorithms. However, Doppler measurements influenced the findings of ICSI prediction and should be taken into account by professionals. The outcomes were highly satisfying and supported our theory that when employing extra features, we were able to predict and enhance the success of ICSI with an accuracy of 82.3%, while Doppler parameters improved the models of prediction. The strength of our study was the way the retrospective data was gathered to create the prediction model, specifically the way the Doppler measurements and clinical characteristics were combined. To validate the use of Doppler technology in conjunction with

clinical factors, it is crucial to expand the number of patients to achieve more precise pregnancy outcomes.

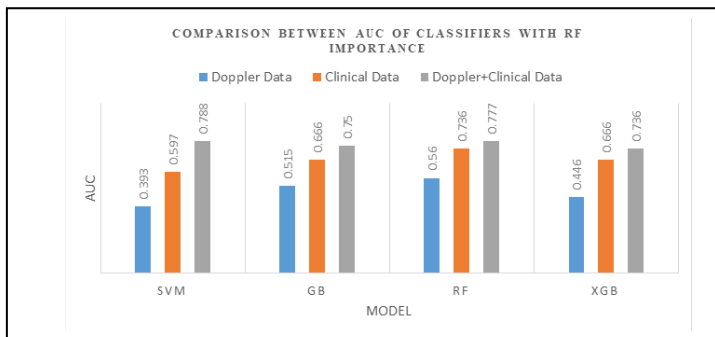


Fig. 4: Histogram representing the comparison of the AUC between Doppler, Clinical and the Combination between Doppler and Clinical.

V. CONCLUSION

The purpose of this study was to use machine learning algorithms to analyze new Doppler factors influencing clinical pregnancy outcomes, which could be useful information for doctors and infertile couples. With an accuracy of 82.3% and an AUC of 0.75, our study revealed that the GB Classifier with RF Importance for feature selection as a pre-processing strategy beat other models when predicting the result of the ICSI treatment. The comparison of Doppler, clinical, and all parameters revealed that the classification with all parameters had higher accuracy and AUC than the classification with Doppler or Clinical parameters. Furthermore, a considerable amount of data is presently being collected to further examine the role of Doppler parameters.

REFERENCES

- [1] Asli Uyar, "Predicting Implantation Outcome from Imbalanced IVF dataset," in *Proceedings of the world congress on engineering and computer science*, San Francisco, USA, 2009.
- [2] M. Babitha, "A Survey on the Machine Learning Techniques used in IVF Treatment to Improve the Success Rate," in *RTESIT*, 2019.
- [3] M. Durairaj and R. N. Kumar, "Data Mining Application on IVF Data For The Selection of Influential Parameters on Fertility," *International Journal of Engineering and Advanced Technology (IJEAT)*, pp. 262-266, 2013.
- [4] J. Qiu, P. Li, M. Dong, X. Xin and J. Tan, "Personalized prediction of live birth prior to the first in vitro fertilization treatment: a machine learning method.," *Journal of Translation Medicine*, p. 17, 2019.
- [5] Q. Zhang, X. Wang, Y. Zhang, H. Lu and Y. Yu, "Nomogram prediction for the prediction of clinical pregnancy in Freeze-thawed Embryo Transfer," *BMC Pregnancy and Childbirth*, 2022.
- [6] M. Ivanovski, "The Role of Ultrasound in the Evaluation of Endometrial Receptivity Following Assisted Reproductive Treatments," in *In Vitro Fertilization - Innovative Clinical and Laboratory Aspects*, Macedonia, www.Intechopen.com, 2012, pp. 31-68.
- [7] J. Zhao, Q. Zhang and Y. Li, "The effect of endometrial thickness and pattern measured by ultrasonography on pregnancy outcomes during IVF-ET cycles," *Reproductive Biology and Endocrinology*, 2012.
- [8] N. Handayania, C. M. Louisa, A. Erwin, T. Aprilliana, A. A. Polim, B. Sirait, A. Boediono and I. Sino, "Machine Learning Approach to Predict Clinical Pregnancy Potential in Women Undergoing IVF Program," *Fertility and Reproduction*, pp. 77 - 87, 2022.
- [9] P. Dr. M.Durairaj, "Application of Artificial Neural Network for IVF data analysis and prediction," *Journal of Engineering, Computers & Applied Sciences*, pp. 11-15, 2014.