

Prediction of Complicated Ovarian Hyperstimulation Syndrome in Assisted Reproductive Treatment Through Artificial Intelligence

Short Title : AI Prediction of OHSS in ART

Arash Ziaee¹, Hamed Khosravi², Tahereh Sadeghi³, Imtiaz Ahmed², Maliheh
Mahmoudinia^{4*}

1. Medical Student, Student Research Committee, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran
2. Department of Industrial & Management Systems Engineering, West Virginia University, Morgantown, WV, US
3. Assistant Professor of Nursing, Department of Pediatrics, School of Nursing and Midwifery, Nursing and Midwifery Care Research Center, Akbar Hospital, Mashhad University of Medical Sciences, Mashhad, Iran
4. Assistant Professor, Department of Industrial & Management Systems Engineering, West Virginia University, Morgantown, WV, US
5. Assistant Professor of Obstetrics & Gynecology, Fellowship of Infertility, Supporting the Family and the Youth of Population Research Core, Department of Obstetrics and Gynecology, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran

*Corresponding Author: Maliheh Mahmoudinia, Email: mahmoudiniam@mums.ac.ir

Number of Words : 3968

Number of Tables : 4

Number of Figures : 5

Number of References : 49

ABSTRACT

Background

This study explores the utility of machine learning (ML) models in predicting complicated Ovarian Hyperstimulation Syndrome (OHSS) in patients undergoing infertility treatments, addressing the challenge posed by highly imbalanced datasets.

Objective

This research fills the existing void by introducing a detailed structure for crafting diverse machine learning models and enhancing data augmentation methods to predict complicated OHSS effectively. Importantly, the research also concentrates on pinpointing critical elements that affect OHSS.

Method

This retrospective study employed a ML framework to predict complicated OHSS in patients undergoing infertility treatment. The dataset included various patient characteristics, treatment details, ovarian response variables, oocyte quality indicators, embryonic development metrics, sperm quality assessments, and treatment specifics. The target variable was OHSS, categorized as painless, mild, moderate, or severe. The ML framework incorporated Ray Tune for hyperparameter tuning and SMOTE-variants for addressing data imbalance. Multiple ML models were applied, including Decision Trees, Logistic Regression, SVM, XGBoost, LightGBM, Ridge Regression, KNN, and SGD. The models were integrated into a voting classifier, and the optimization process was conducted. The SHAP package was used to interpret model outcomes and feature contributions.

Results

The best model incorporated IPADE-ID augmentation along with an ensemble of classifiers (SGDClassifier, SVC, RidgeClassifier), reaching a recall of 0.9 for predicting OHSS occurrence and an accuracy of 0.76. SHAP analysis identified key factors: GnRH antagonist use, longer stimulation, female infertility factors, irregular menses, higher weight, hCG triggers, and, notably, higher number of embryos .

Conclusion

This novel study demonstrates ML's potential for predicting complicated OHSS. The optimized model provides insights into contributory factors, challenging certain conventional assumptions. The findings highlight the importance of considering patient-specific factors and treatment details in OHSS risk assessment.

Keywords: Ovarian Hyperstimulation Syndrome; Machine Learning; Assisted Reproductive Technology; In Vitro Fertilization; Data Augmentation

INTRODUCTION

Ovarian hyperstimulation syndrome (OHSS) represents a critical challenge in assisted reproductive technology (ART), marked by a broad spectrum of clinical manifestations. This iatrogenic condition, typically occurring during the early luteal phase or early pregnancy following ovulation induction (OI) or ovarian stimulation (OS), varies in prevalence across different regions. In the United States, the incidence of moderate to severe OHSS in ART treatments was reported at about 1.2% in 2006, decreasing to 0.5% in 2014[1]. As per the European In Vitro Fertilization (IVF) -monitoring (EIM) Consortium, European figures show an incidence rate of only 0.3%[2].

OHSS is characterized by bilateral cystic enlargement of highly luteinized ovaries, leading to complications like vascular hyperpermeability and hemorrhagic ovarian cysts. Clinically significant OHSS occurs in 2-3% of cases, with milder forms affecting up to 20-30% of IVF patients[3]. Moderate OHSS can lead to symptoms like abdominal distention and nausea, with about 1.9% of patients needing hospitalization for severe effects, including hepatorenal failure and thromboembolism[4, 5]

OHSS typically occurs after ovarian stimulation with gonadotropins, causing an excessive ovarian reaction, including multiple follicle growth, high estradiol levels, and ovarian swelling. Exogenous hCG is crucial for oocyte maturation's final steps. Its pathophysiology involves granulosa/luteal cells releasing vasoactive substances, leading to increased vascular permeability. Pregnancy following ovarian stimulation worsens OHSS as placental hCG boosts ovarian VEGF secretion, aggravating symptoms.[6-8].

Amid these challenges, artificial intelligence (AI) and machine learning (ML) have emerged as promising tools for improving OHSS management. Clinicians have traditionally relied on their expertise to prescribe the appropriate FSH dosage for follicular stimulation. The field of ML in OHSS management has seen significant advancements, such as optimizing the initial FSH dose in controlled ovarian hyperstimulation[9] and enhancing IVF protocols by predicting the effectiveness of each day of controlled ovarian stimulation (COS) monitoring[10]. These developments focus on predictive accuracy and prioritize patient safety and treatment efficacy. These models use historical data and predictive analytics to

improve treatment outcomes, personalize therapies, and minimize the incidence of OHSS, representing notable progress in personalized medicine.

However, the challenge of imbalanced data in this field presents a significant obstacle in developing ML models specifically tailored to predict complicated OHSS occurrence and identify its essential contributing factors. This study addresses this gap by proposing a comprehensive framework for developing various ML models and data augmentation techniques to tackle this challenge. Notably, this study also focuses on identifying key factors influencing OHSS, thereby contributing valuable insights to the field and aiding in more effective complicated OHSS prediction and management.

METHODOLOGY

Methodological Framework

This investigation adopted a retrospective analytical framework, leveraging advanced ML algorithms to forecast the incidence of complicated OHSS among patients undergoing infertility treatments. The dataset for this study was meticulously composed of a wide array of patient characteristics and treatment variables as shown in Table 1.

Table 1: Comprehensive Overview of Parameters Used in the Study

Category	Characteristics
Physiological Information	Age, Weight, Height
Infertility and Treatment Details	Duration of Infertility, FSH, LH, Number of Consumed Drugs, Duration of Stimulation, Previous pregnancy [Negative, Positive], Menstruation [regular, irregular], Infertility [Primary, Secondary]
Ovarian Response Variables	Number of follicles in right ovary, Number of follicles in left ovary, Number of oocytes
Oocyte Quality Indicators and Cell Types	Metaphase one oocyte, Metaphase two oocyte, Necrotic oocyte, Low-quality oocyte, GV (germinal vesicle) oocyte, Macro-oocyte, Parthenogenetic oocyte
Embryonic Development Metrics	Number of Embryos, Grade Embryo [Grade 1, Grade 2, Grade 3]
Sperm Quality Assessments	Count Spermogram, Motility Spermogram, Morphology Spermogram
Treatment Specifics	Type cycle [GnRH Agonist – GnRH Antagonist], Source of Infertility [Female, Male, Female plus Male, Unexpected], Type of consumption drugs, Type of trigger [GnRH agonist, hCG, GnRH plus hCG], Type Drug [Cabergoline, Cabergoline plus Calcium]
Target Variable	OHSS [Painless, Mild, Moderate, Severe]

OHSS categories span from painless, with no significant symptoms or lab changes, to mild, moderate, severe, and critical levels, each marked by escalating severity of clinical and biochemical features. Mild OHSS manifests as abdominal discomfort and nausea without lab abnormalities, indicating its mild nature. Moderate OHSS includes these symptoms plus ascites on ultrasound, necessitating increased monitoring. Severe OHSS adds to the prior symptoms with tense ascites, severe pain, rapid weight gain, pleural effusion, and critical lab changes like elevated hematocrit and serum creatinine, indicating a significant systemic impact. Critical OHSS severely compromises vital organ functions, with potential outcomes including acute renal failure, cardiac arrhythmia, respiratory failure, thrombosis, massive hydrothorax, sepsis, and ARDS, highlighting its life-threatening nature. Our dataset did not contain critical patients.

Development of a Machine Learning Framework

A sophisticated ML framework was engineered, incorporating Ray Tune for the nuanced tuning of hyperparameters[11]. This framework was designed to comprehensively refine the ML pipeline, covering aspects from data preprocessing and feature selection to model optimization and establishing a voting classifier system. The aim was to conduct exhaustive iterations to ascertain the most practical combination of preprocessing methodologies, algorithmic models, and their corresponding parameters.

Ethics Statement

This retrospective study analyzed anonymized medical records with approval from the Mashhad University of Medical Sciences ethics committee (IR.MUMS.REC.1395.326). Data were accessed on July 11, 2023, and contained no identifying variables. Verbal consent was obtained from all patients for the use of their anonymized data, in accordance with ethical guidelines and data protection regulations.

Data Preprocessing

The study aimed to forecast the occurrence of complicated OHSS. To achieve this, the target variable was transformed into a binary format: 'painless' and 'mild' cases were assigned a

value of 0, while 'moderate' and 'severe' cases were labeled as 1. This binary encoding was based on the rationale that moderate and severe cases necessitate medical intervention.

To address the challenge of missing data, our preprocessing strategy employed Random Forest imputation for continuous variables and mean imputation for categorical variables, ensuring the integrity and completeness of our dataset. Additionally, we incorporated a flexible selection among various scaling options (Robust, Standard, Min-Max scalers, and a Passthrough option) and integrated the SMOTE-variants package[12, 13]. This package allowed the algorithm to choose from 60 out of 85 SMOTE variants, including oversampling and undersampling methods, to address data imbalance effectively. We ensured that extrapolated data remained within realistic boundaries by storing and applying min-max values pre-imputation. Categorical variables were processed using a 0.5 threshold for binary conversion.

Initial Analysis

In the initial phase, multiple machine learning models were applied directly to the original dataset, which was used without any modifications for class balancing or parameter tuning. This step focused on assessing the models' performance using all available features, considering a range of metrics. This process identified the crucial roles of data augmentation and tuning in enhancing model effectiveness.

Dynamic Feature Selection

The feature selection process was engineered to be dynamic, allowing the search algorithm to enable or disable each variable independently. This adaptability facilitated delineating the most critical features for each model, thereby augmenting the predictive accuracy of the ML framework.

Model Selection and Hyperparameter Optimization

The selection regime encompassed an array of models, including Decision Trees, Logistic Regression, SVM, XGBoost, LightGBM, Ridge Regression, KNN, and SGD. An extensive

parameter set was provided for algorithmic determination of the most efficacious configurations for each model. The search algorithm then chose the model it saw fit and selected the required parameters. These models were subsequently integrated into a voting classifier, utilizing a soft voting mechanism to include models lacking support for predict_proba functionality.

Objective and Evaluation Metrics

Our primary goal was to improve the recall of the positive class while ensuring that the recall for the negative class remained robust. This approach was comprehensive, considering all metrics, including recall for class 0, F1 scores, and overall accuracy. This balanced strategy was explicitly developed to navigate the challenges arising from the lack of multi-objective optimization capabilities in Ray Tune, ensuring a balanced enhancement of the model's predictive performance.

A confusion matrix in binary classification organizes outcomes into true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), helping distinguish correct from incorrect predictions. It facilitates the calculation of key metrics: precision (accuracy of positive predictions), recall (identification accuracy within the positive class), overall accuracy, and the F1 score, harmonizing precision and recall [14, 15].

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{Overall Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$\text{F1 Score} = \frac{\text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}} \quad (4)$$

Iterative Optimization

The optimization endeavor was structured into two distinct phases, each comprising 15,000 trials. The initial phase's outcomes were scrutinized using the Weights and Biases package,

facilitating the exclusion of underperforming models and SMOTE variants. The subsequent phase concentrated on the refined selection of models, thereby enhancing the ML framework's efficiency and effectiveness.

Interpretation of Models and Outcomes

The interpretative analysis of model outcomes and optimization processes was significantly bolstered by the employment of the SHAP (SHapley Additive exPlanations) package [16]. Utilizing SHAP values, which are rooted in cooperative game theory, allowed for quantifying each feature's contribution to the predictive outcome.

RESULT

In this section, we initiate an analysis of various developed ML models on the original data without augmentation and tuning of the models. Following this, we examine the best-performing approach, encompassing both the most effective data augmentation technique and the top-performing ML model identified in our study. Subsequently, a SHAP analysis is conducted to identify and understand the key factors contributing to the model's predictive capabilities.

ML models' performance on the original data

Utilizing the methodology outlined, the ML models were initially applied to the original dataset to predict the occurrence of OHSS. Table 2 presents the results for the models, where 'Class 1' specifically denotes instances where the target occurs.

Table 2: The performance of different ML models on the real data for prediction of OHSS

Model/Metric	Recall	Recall	F1 Score	F1 Score	Precision	Precision	Accuracy
	Class 0	Class 1	Class 0	Class 1	Class 0	Class 1	
RandomForestClassifier	1.00	0.0	0.95	0.0	0.90	0.0	0.90
DecisionTreeClassifier	0.96	0.1	0.93	0.1	0.91	0.2	0.87
LogisticRegression	1.00	0.0	0.95	0.0	0.90	0.0	0.90

GaussianNB	0.08	0.8	0.14	0.15	0.78	0.09	0.15
KNeighborsClassifier	1.00	0.0	0.95	0.0	0.90	0.0	0.90
SVC	1.00	0.0	0.95	0.0	0.90	0.0	0.90
XGBClassifier	0.98	0.1	0.94	0.15	0.91	0.33	0.89

As presented in Table 2, we explored the performance of various machine learning models in predicting complicated OHSS. These models, including RandomForestClassifier, DecisionTreeClassifier, LogisticRegression, GaussianNB, KNeighborsClassifier, SVC, and XGBClassifier, generally achieved high accuracy rates (over 85% in most cases). However, a detailed analysis uncovered a significant issue often called the 'accuracy paradox' in machine learning [17]. This paradox occurs when models, despite high overall accuracy, fail to effectively predict certain classes, especially the minority class, in imbalanced datasets. Specifically, while these models showed high accuracy, their performance in terms of recall and precision for class 1 (OHSS occurrence) was suboptimal. For the majority of the models, recall for class 1 was notably low (around 0 or 0.1), indicating a poor ability to identify positive OHSS cases correctly. Conversely, the GaussianNB model, although having a higher recall for class 1 (0.8), exhibited a low recall for class 0 (0.075), reflecting a skewed prediction ability. This imbalance in model performance is visually represented in Figure 1, utilizing advanced dimension reduction techniques for more precise data visualization.

The 3D scatter plot shown in Figure 1 represents a Principal Component Analysis (PCA) output, a method commonly employed for dimensionality reduction [18]. In this plot, the data points are color-coded to represent two classes: Class 1 (red) and Class 0 (blue). The plot clearly shows an imbalance in the dataset, with Class 0 being significantly more prevalent than Class 1. Moreover, the plot reveals an overlap between Class 0 and Class 1 data points within the PCA-transformed feature space. This overlap suggests a difficulty for classifiers in distinctly separating the two classes. The classes' proximity in this reduced dimensionality space indicates that they are not linearly separable, which could lead to decreased effectiveness in classification algorithms. This overlap and the lack of clear

separation highlight the complexity of the dataset and the need for advanced or tailored ML approaches to classify such nuanced data accurately.

Best Developed Model

As discussed in the methodology section, we analyzed different data augmentation methods coupled with different ML techniques. This approach created an ensemble learning framework, specifically using voting classifiers, which was further refined through feature tuning and hyperparameter adjustments. As a result of this extensive tuning process, a total of 15,000 distinct models were developed. Our evaluation criteria extended beyond a single metric to identify the best-performing model. We aimed for a model that demonstrated satisfactory performance across various metrics, particularly on recall for class 1. This approach ensures a more balanced and comprehensive assessment of the model's predictive capabilities, especially in accurately identifying the minority class. Figure 2 presents the outcomes of 90 models selected from the pool of 15,000, representing the top, medium, and worst performers.

Figure 2 (a) shows the distribution of key performance metrics—precision, recall, and F1-scores for both classes (0 and 1)—across the selected set of models. The variation in these metrics, especially for class 1, highlights the inherent difficulties in accurately predicting the minority class in the dataset. Figure 2 (b) demonstrates a scatter plot illustrating the trade-off between recall scores for both classes. The color gradient in this plot indicates the accuracy level of each model. Notably, the model that achieved a high recall for class 1 (0.9), which is paramount in identifying true OHSS cases, also maintained a robust recall for class 0 (0.74), thus demonstrating a balanced sensitivity across classes. In Figure 3 (c), a scatter plot features the F1-scores of both classes, with accuracy once again represented by a color gradient. The model highlighted here not only demonstrated a high recall for class 1 but also achieved the best F1 scores for both classes. This indicates an excellent balance between precision and recall, underscoring the model's comprehensive performance capabilities.

In this best-developed model, the Iterative Instance Adjustment for Imbalanced Domains (IPADE-ID) algorithm [19] emerged as the most effective augmentation technique. The implementation of instance-generation techniques like IPADE-ID is increasingly

acknowledged as a practical solution for addressing the challenges posed by highly imbalanced datasets. Figure 3 shows the similarity between the dataset generated using the IPADE-ID technique and the real data, highlighting the technique's efficacy in producing representative and accurate synthetic datasets.

In Figure 3, we present a comprehensive analysis of the data generated by the IPADE-ID, aimed at mirroring the distribution of the real dataset. Figure 3 (a) features a Uniform Manifold Approximation and Projection (UMAP) visualization [20], plotting both real (blue points) and generated (red points) data in a two-dimensional space. The UMAP algorithm captures the underlying structure of high-dimensional data, revealing two distinct clusters. Notably, the synthetic data points are well-integrated with the real data points, signifying that the generative model has accurately learned the complex manifold of the real data. Figure 3 (b) offers a Parallel Coordinates plot [21], which visualizes the values of each feature across the real and generated datasets. This plot highlights the fidelity of the generated data, as shown by the overlapping lines that indicate similar feature distributions between the two datasets. The red lines, representing the generated data, closely match the blue lines of the real data across various features. This parallelism suggests that the generative model has effectively captured the real data's central tendencies, variability, and outliers, confirming the synthetic dataset's validity for further analysis.

Therefore, we utilized the synthetically generated dataset for the ensemble model, which was identified as the most effective. As depicted in Figure 4, the ensemble machine learning models applied to the generated datasets were SGDClassifier, SVC, and RidgeClassifier. It should be noted that when these models were not used in an ensemble configuration, they could not accurately identify any instance of OHSS occurrence, yielding a recall rate of zero for class 1. However, when applied to the data generated using the IPADE-ID technique, the ensemble model exhibited outstanding performance, particularly in predicting class 1, which indicates the occurrence of OHSS. The effectiveness of this approach is detailed in Table 3.

Table 3: The performance of the best-developed ML model

Metric	Score
Recall Class 1	0.9
Recall Class 0	0.74
F2 Score Class 0	0.78
F2 Score Class 1	0.73
Accuracy	0.76

According to Table 2, the ensemble model accurately predicted the occurrence of OHSS 90% of the time. Additionally, the model demonstrated an overall satisfactory performance with an accuracy rate of 0.76. The model also performed well in other key metrics.

SHAP analysis

Building on the results of the best model, the next step involves using this model to perform a SHAP analysis. This analysis helps identify and understand the most influential features in predicting OHSS occurrence, providing deeper insights into the model's decision-making process.

Figure 5 presents a SHAP summary plot, providing a visual representation of how the selected features from the tuning phase influence the model's output. On the X-axis, the SHAP values quantify the impact of each feature, with zero indicating no impact and higher absolute values (positive or negative) indicating a more significant influence on the model's predictions. The Y-axis lists the features used in the model, and the color represents the value of the features, where one side of the color spectrum (blue) shows a low feature value, and the other side (red) shows a high feature value. Key features such as "Typecycle" and "reasoninfertility" stand out in the plot, signifying their crucial effect on determining the model's output. A closer examination reveals those higher values of "reasoninfertility" correlate with a decreased likelihood of predicting OHSS, representing the risk of

complicated OHSS decreases in cases of solely male infertility or combined male and female infertility since this feature mostly aligns with the negative side of the SHAP value spectrum. Conversely, elevated values in "Typecycle" and "mense" [irregular or irregular menstruation

cycles], which denote the use of GnRH agonist and the presence of irregular menstruation, respectively, positively impact the model's output. These factors are associated with an increased probability of accurately predicting OHSS. These findings are extensively analyzed and discussed in the discussion section of our paper, where we delve deeper into how these features collectively influence the model's predictions and the implications of their respective impacts on the risk of OHSS.

DISCUSSION

This study represents a pioneering effort in using ML to predict complicated OHSS, addressing a gap in the existing literature. Confronting the challenge of a highly imbalanced dataset, we employed data augmentation techniques and optimized a voting classifier's hyperparameters, achieving notable results. Our work diverges from traditional infertility research, which often focuses on predicting successful pregnancy or oocyte retrieval outcomes, by specifically addressing the less-explored complications of infertility treatments like OHSS. This approach not only fills a significant gap in the literature but also provides new perspectives on the prevention and management of such infertility treatment complications. Table 4 displays a comprehensive comparison of our study's contributions with similar research in the literature, highlighting how this research advances and enhances

Ref	Study setting	Imbalanced data	Number of ML Models	Best ML model	Result
[22]	Number Of Oocytes Retrieved	-	1	LightGBM	Mean Absolute Percentage Error: 0.52
[23]	IUI Success	■	5	Stack model	AUC: 0.92

the field of ART.

Table 4. Similar Studies That Have Utilized ML in the Field of ART

[24]	Corifollitropin Alfa Protocol	-	1	LightGBM	AUC: 0.89
[25]	Day Of Trigger	-	1	LightGBM	Average Outcome Improvement on 2PNs (Custom Metric): 3.015
[26]	Predicting Clinical Pregnancy	-	1	Random forest	AUC: 0.72
[27]	Live-Birth Occurrence Before In-Vitro Fertilization Treatment	-	8	Random Forests	AUC: 0.87 F1-score: 0.76
[28]	Predict PCOS	-	10	Gaussian Naive Bayes	Accuracy: 1.0
[29]	Success In an IVF Treatment Cycle	-	17	Super Learner and random forest	Accuracy: 0.89 F1 scores: 0.74
[30]	Prediction Of Euploidy	-	5	Random Forests	Accuracy: 0.71 AUC: 0.75
This Study	Prediction of Complicated OHSS	■	11	Ensemble of SGDClassifier, SVC, and RidgeClassifier	Accuracy: 0.76 Recall: 0.90

GnRH Antagonists vs. GnRH Agonists and Stimulation Duration

The results of our machine learning study using SHAP revealed that the use of GnRH antagonists, compared to GnRH agonists, and longer stimulation duration were identified as important factors in predicting the occurrence of complicated OHSS. The existing literature largely supports these findings. A comprehensive Cochrane review provided moderate quality evidence that GnRH antagonists significantly reduce the incidence of OHSS without compromising live birth rates compared to long-course GnRH agonist protocols [30]. Similarly, a clinical trial focusing on PCOS patients found that the GnRH antagonist group had a significantly lower incidence of OHSS while maintaining similar pregnancy rates compared to the GnRH agonist group [31]. A large RCT and a prospective, randomized study also reported a significantly lower incidence of OHSS in the GnRH

antagonist group compared to the GnRH agonist group while maintaining similar ongoing pregnancy rates (OPR) and live birth rates in both protocols [32-34]. These findings, taken together with the results of our machine learning study, suggest that the use of GnRH antagonists is an important factor in predicting and potentially reducing the occurrence of complicated OHSS. The identification of longer stimulation duration as a predictive factor for OHSS also highlights the need for further research to optimize stimulation protocols and minimize the risk of this potentially life-threatening complication.

Reason [Source] of Infertility

Our ML model, using SHAP, found that the source of infertility is a significant predictor for complicated OHSS. The risk of complicated OHSS increases when the female factor is more involved, while the risk decreases in cases of solely male infertility or combined male and female infertility. This finding can be attributed to the fact that female-factor infertility often requires more aggressive ovarian stimulation protocols to overcome underlying issues such as ovulatory disorders or poor ovarian reserve [35]. These aggressive stimulation protocols, involving high doses of gonadotropins, are a crucial risk factor for developing OHSS. Additionally, certain genetic predispositions and biochemical factors, such as young age, lower time of infertility, lower baseline FSH, higher female factor and PCOS phenotype, body mass index, and age, and a high number of follicles and elevated estradiol levels, are associated with a higher risk of developing OHSS [35]. In contrast, when infertility is primarily due to male factors or unexpected or both female and male, the female partner may undergo less aggressive stimulation protocols, especially when using techniques like ICSI, where only a few eggs are needed, thus reducing the risk of OHSS.

Types Of Women Infertility, Irregular vs. Regular Menstrual Cycles, Weight And PCOS

Our model identified irregular menstruation cycles, higher weight, and primary infertility as important predictors of complicated OHSS occurrence, which broadly aligns with the literature. Irregular menstruation can be associated with conditions that cause hormonal imbalances, such as PCOS, a condition known for its increased risk of OHSS[36-38]. However, the relationship between weight and OHSS remains inconclusive, with some studies finding no significant correlation [39] and others suggesting that obesity may decrease the risk of OHSS-related complications[40]. Interestingly, the model assigned lower importance to secondary infertility as a predictor of OHSS, which contradicts previous findings linking conditions that cause secondary female infertility, such as pituitary adenoma, to an increased incidence of OHSS[41]. These discrepancies highlight the need

for further research to clarify the role of weight and secondary infertility in complicated OHSS prediction.

Number of Oocytes, Embryos, Other Cell Types, and Rule of Sperm Morphology

Our machine learning model for predicting complicated OHSS presents novel insights that diverge from conventional wisdom and prior research findings. Contrary to established risk factors such as the number of retrieved oocytes, which studies like those by Verwoerd et al. suggest increase OHSS risk significantly beyond a 15-oocyte threshold[42, 43], our model de-emphasizes the importance of oocyte count. Instead, it identifies a surprising correlation between the number of embryos and higher complicated OHSS risk, assigning significant weight to the presence of 11 or more embryos as a critical predictor. This finding challenges the traditional understanding, as higher oocyte retrieval numbers have been linked to increased OHSS risk without corresponding benefits in live birth rates in fresh autologous IVF cycles. Our analysis also highlights the potential predictive value of other cell types, including pathogenetic oocytes, GV oocytes, and macroocytes for complicated OHSS cases. Furthermore, while no direct studies have linked sperm morphology with OHSS incidence, our observations suggest that improved sperm quality may contribute to better-fertilized oocyte outcomes, indirectly influencing complicated OHSS risk. This aligns with the literature on the predictive value of normal sperm morphology for pregnancy success in IUI and IVF, where a significant improvement in pregnancy rates was noted for strict sperm morphology criteria[44].

Types of Triggers

The utilization of different trigger drugs in ovarian stimulation protocols has been shown to influence the risk of complicated OHSS. Our model, using SHAP analysis, found that the use of hCG or hCG combined with GnRH agonists as trigger drugs significantly increases the risk of complicated OHSS. In contrast, the use of GnRH agonists alone was a predictor of uncomplicated OHSS. These findings align with previous literature, which has shown that GnRH agonists result in a significantly lower incidence of moderate to severe OHSS compared to hCG, albeit with lower live birth and ongoing pregnancy rates[45, 46]. While dual trigger with GnRH agonists and low-dose hCG may be associated with a modest increase

in oocyte yield[47], the risk of OHSS following dual triggering remains unclear[48]. A recent study recommends the use of a single GnRH agonist trigger for high responders treated with the freeze-all strategy to prevent moderate to severe OHSS and obtain satisfactory pregnancy and neonatal outcomes in subsequent frozen embryo transfer cycles[49].

Several key limitations, including the reliance on a small dataset from a single center, can affect the generalizability of our findings. Additionally, the absence of an early prediction mechanism and the study's non-utilization of time series analysis limit our model's ability to preemptively identify OHSS risk and dynamically track its progression over time. These constraints highlight areas for future research to enhance predictive accuracy and clinical utility.

CONCLUSION

In conclusion, our pioneering study demonstrates the potential of machine learning models in predicting complicated OHSS, a previously unexplored area in infertility research. Despite the challenges posed by the highly imbalanced dataset, we successfully implemented various data augmentation techniques and optimized the hyperparameters of a voting classifier to achieve satisfactory results. Our findings not only contribute to the growing body of research on machine learning applications in assisted reproductive technology but also provide valuable insights into the factors influencing the occurrence of complicated OHSS.

AUTHOR CONTRIBUTIONS

A.Z. and H.K. developed the machine learning framework, performed data analysis, and revised the drafted manuscript [Software], [Formal Analysis], [Writing – review & editing]. M.M. was responsible for data acquisition, supervision, and preparation of the final draft [Data curation], [Supervision], [Writing – original draft]. I.A. contributed to the

conceptualization, supervision of the study and final draft preparation [Conceptualization], [Supervision], [Writing – original draft]. T.S. prepared the original draft and assisted in data acquisition [Writing – original draft], [Data curation]. All authors contributed to the evaluation and interpretation of the results [Validation]. All authors approve of the submitted manuscript.

DATA ACQUISITION AND ETHICAL CONSIDERATIONS

Data collection adhered strictly to ethical standards, with approval from the relevant university ethics committee with approval code of IR.MUMS.REC.1395.326 . Before inclusion in the study, consent was obtained from all patients to use their anonymized data for research and analysis purposes.

ACKNOWLEDGMENTS

We want to extend our gratitude to Prof. Mehdi Jabbari Nooghabi, whose guidance and insights have been instrumental in the development of this machine learning framework.

CONFLICT OF INTEREST

The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

FUNDING

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

DATA AVAILABILITY STATEMENT

The dataset supporting the conclusions of this article is confidential and cannot be made publicly available. However, data may be available from the corresponding author upon reasonable request and with permission of the original data providers, subject to compliance with applicable confidentiality agreements and data protection laws.

Declaration of Generative AI and AI-assisted Technologies in the Writing Process

In the development and composition of this document, generative AI was not utilized beyond the scope of fundamental tools for grammar, spelling, and reference verification.

Summary Table

What was already known on the topic:

- Machine learning models have been applied to predict outcomes like oocyte retrieval numbers, IUI success, pregnancy rates, and live birth rates in assisted reproductive technology (ART)
- Ovarian hyperstimulation syndrome (OHSS) is a potentially life-threatening complication of ART, but prediction of complicated OHSS using machine learning has not been explored previously
- Imbalanced datasets are a major challenge for developing effective machine learning models in medical domains

What this study added to our knowledge:

- This is the first study to develop and optimize machine learning models specifically for predicting complicated cases of OHSS, addressing the obstacle of imbalanced data
- The optimized ensemble model provides insights that challenge certain conventional assumptions about risk factors for OHSS, such as de-emphasizing oocyte numbers while highlighting the number of embryos as a predictor
- Novel data augmentation techniques like IPADE-ID were effectively applied to tackle the highly imbalanced nature of the OHSS dataset
- The study identified key influential factors like GnRH antagonist use, stimulation duration, female infertility factors, irregular menses, higher weight, hCG trigger usage, and number of embryos through interpretation of the optimized model

REFERENCES

1. Toner JP, Coddington CC, Doody K, Van Voorhis B, Seifer DB, Ball GD, et al. Society for Assisted Reproductive Technology and assisted reproductive technology in the United States: a 2016 update. *Fertil Steril*. 2016;106(3):541-6. Epub 20160611. doi: 10.1016/j.fertnstert.2016.05.026. PubMed PMID: 27301796.
2. De Geyter C, Calhaz-Jorge C, Kupka MS, Wyns C, Mocanu E, Motrenko T, et al. ART in Europe, 2015: results generated from European registries by ESHRE. *Human reproduction open*. 2020;2020(1):hoz038.
3. Papanikolaou EG, Pozzobon C, Kolibianakis EM, Camus M, Tournaye H, Fatemi HM, et al. Incidence and prediction of ovarian hyperstimulation syndrome in women undergoing gonadotropin-releasing hormone antagonist in vitro fertilization cycles. *Fertil Steril*. 2006;85(1):112-20. doi: 10.1016/j.fertnstert.2005.07.1292. PubMed PMID: 16412740.
4. Luke B, Brown MB, Morbeck DE, Hudson SB, Coddington CC, 3rd, Stern JE. Factors associated with ovarian hyperstimulation syndrome (OHSS) and its effect on assisted reproductive technology (ART) treatment and outcome. *Fertil Steril*. 2010;94(4):1399-404. Epub 20090709. doi: 10.1016/j.fertnstert.2009.05.092. PubMed PMID: 19591989.
5. Sun B, Ma Y, Li L, Hu L, Wang F, Zhang Y, et al. Factors Associated with Ovarian Hyperstimulation Syndrome (OHSS) Severity in Women With Polycystic Ovary Syndrome Undergoing IVF/ICSI. *Front Endocrinol (Lausanne)*. 2020;11:615957. Epub 20210119. doi: 10.3389/fendo.2020.615957. PubMed PMID: 33542709; PubMed Central PMCID: PMC7851086.
6. Rizk B, Smitz J. Ovarian hyperstimulation syndrome after superovulation using GnRH agonists for IVF and related procedures. *Hum Reprod*. 1992;7(3):320-7. doi: 10.1093/oxfordjournals.humrep.a137642. PubMed PMID: 1587936.
7. Mozes M, Bogokowsky H, Antebi E, Lunenfeld B, Rabau E, Serr DM, et al. Thromboembolic phenomena after ovarian stimulation with human gonadotrophins. *Lancet*. 1965;2(7424):1213-5. doi: 10.1016/s0140-6736(65)90636-7. PubMed PMID: 4158825.
8. Roberts WG, Palade GE. Neovasculature induced by vascular endothelial growth factor is fenestrated. *Cancer research*. 1997;57(4):765-72.
9. Chambost J, Jacques C, Ferrand T, Hickman C, He C, Reigner A, Freour T. P-682 Predicting the Number of Oocytes Retrieved from Controlled Ovarian Hyperstimulation with Machine Learning. *Human Reproduction*. 2022;37(Supplement_1). doi: 10.1093/humrep/deac107.631.
10. Correa Mañas N, Cerquides J, Arcos JL, Vassena R, Popovic M. O-185 A clinically robust machine learning model for selecting the first FSH dose during controlled ovarian hyperstimulation: incorporating clinical knowledge to the learning process. *Human Reproduction*. 2023;38(Supplement_1). doi: 10.1093/humrep/dead093.226.
11. Liaw R, Liang E, Nishihara R, Moritz P, Gonzalez JE, Stoica I. Tune: A research platform for distributed model selection and training. *arXiv preprint arXiv:180705118*. 2018.
12. Kovács G. An empirical comparison and evaluation of minority oversampling techniques on a large number of imbalanced datasets. *journal*. 2019;83:105662. doi: 10.1016/j.asoc.2019.105662.
13. Kovács G. Smote-variants: A python implementation of 85 minority oversampling techniques. *Neurocomputing*. 2019;366:352-4. doi: 10.1016/j.neucom.2019.06.100.
14. Maxwell AE, Warner TA, Guillen LA. Accuracy assessment in convolutional neural network-based deep learning remote sensing studies—Part 2: Recommendations and best practices. *Remote Sensing*. 2021;13(13):2591.

15. Maxwell AE, Warner TA. Thematic classification accuracy assessment with inherently uncertain boundaries: An argument for center-weighted accuracy assessment metrics. *Remote Sensing*. 2020;12(12):1905.
16. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. *Advances in neural information processing systems*. 2017;30.
17. Valverde-Albacete FJ, Pelaez-Moreno C. 100% classification accuracy considered harmful: the normalized information transfer factor explains the accuracy paradox. *PLoS One*. 2014;9(1):e84217. Epub 20140110. doi: 10.1371/journal.pone.0084217. PubMed PMID: 24427282; PubMed Central PMCID: PMC3888391.
18. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. *Philos Trans A Math Phys Eng Sci*. 2016;374(2065):20150202. doi: 10.1098/rsta.2015.0202. PubMed PMID: 26953178; PubMed Central PMCID: PMC4792409.
19. López V, Triguero I, Carmona CJ, García S, Herrera F. Addressing imbalanced classification with instance generation techniques: IPAD-ED. *Neurocomputing*. 2014;126:15-28. doi: 10.1016/j.neucom.2013.01.050.
20. McInnes L, Healy J, Melville J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:180203426*. 2018.
21. Johansson J, Forsell C. Evaluation of Parallel Coordinates: Overview, Categorization and Guidelines for Future Research. *IEEE Trans Vis Comput Graph*. 2016;22(1):579-88. doi: 10.1109/TVCG.2015.2466992. PubMed PMID: 26529719.
22. Ferrand T, Boulant J, He C, Chambost J, Jacques C, Pena CA, et al. Predicting the number of oocytes retrieved from controlled ovarian hyperstimulation with machine learning. *Hum Reprod*. 2023;38(10):1918-26. doi: 10.1093/humrep/dead163. PubMed PMID: 37581894; PubMed Central PMCID: PMC10546073.
23. Khodabandelu S, Basirat Z, Khaleghi S, Khafri S, Montazery Kordy H, Golsorkhtabaramiri M. Developing machine learning-based models to predict intrauterine insemination (IUI) success by address modeling challenges in imbalanced data and providing modification solutions for them. *BMC Med Inform Decis Mak*. 2022;22(1):228. Epub 20220901. doi: 10.1186/s12911-022-01974-8. PubMed PMID: 36050710; PubMed Central PMCID: PMC9434923.
24. Lai H-H, Kuo EE-S, Li R-S, Chuang T-H, Huang Y-C, Hsieh J-Y, et al. Using an Interpretable Machine Learning Model to Predict Corifollitropin Alfa Protocol. *Fertility & Reproduction*. 2023;05(01):50-60. doi: 10.1142/s2661318223500068.
25. Hariton E, Chi EA, Chi G, Morris JR, Braatz J, Rajpurkar P, Rosen M. A machine learning algorithm can optimize the day of trigger to improve in vitro fertilization outcomes. *Fertil Steril*. 2021;116(5):1227-35. Epub 20210710. doi: 10.1016/j.fertnstert.2021.06.018. PubMed PMID: 34256948.
26. Wang CW, Kuo CY, Chen CH, Hsieh YH, Su EC. Predicting clinical pregnancy using clinical features and machine learning algorithms in in vitro fertilization. *PLoS One*. 2022;17(6):e0267554. Epub 20220608. doi: 10.1371/journal.pone.0267554. PubMed PMID: 35675328; PubMed Central PMCID: PMC9176781.
27. Goyal A, Kuchana M, Ayyagari KPR. Machine learning predicts live-birth occurrence before in-vitro fertilization treatment. *Scientific reports*. 2020;10(1):20925.
28. Nasim S, Almutairi MS, Munir K, Raza A, Younas F. A Novel Approach for Polycystic Ovary Syndrome Prediction Using Machine Learning in Bioinformatics. *IEEE Access*. 2022;10:97610-24. doi: 10.1109/access.2022.3205587.
29. Yigit P, Bener A, Karabulut S. Comparison of machine learning classification techniques to predict implantation success in an IVF treatment cycle. *Reprod Biomed Online*. 2022;45(5):923-34. Epub 20220628. doi: 10.1016/j.rbmo.2022.06.022. PubMed PMID: 36088224.

30. Al-Inany HG, Youssef MA, Ayeleke RO, Brown J, Lam WS, Broekmans FJ. Gonadotrophin-releasing hormone antagonists for assisted reproductive technology. *Cochrane Database Syst Rev*. 2016;4(4):CD001750. Epub 20160429. doi: 10.1002/14651858.CD001750.pub4. PubMed PMID: 27126581; PubMed Central PMCID: PMCPMC8626739.
31. Hosseini MA, Aleyasin A, Saeedi H, Mahdavi A. Comparison of gonadotropin-releasing hormone agonists and antagonists in assisted reproduction cycles of polycystic ovarian syndrome patients. *J Obstet Gynaecol Res*. 2010;36(3):605-10. doi: 10.1111/j.1447-0756.2010.01247.x. PubMed PMID: 20598044.
32. Ludwig M, Felberbaum RE, Devroey P, Albano C, Riethmuller-Winzen H, Schuler A, et al. Significant reduction of the incidence of ovarian hyperstimulation syndrome (OHSS) by using the LHRH antagonist Cetrorelix (Cetrotide) in controlled ovarian stimulation for assisted reproduction. *Arch Gynecol Obstet*. 2000;264(1):29-32. doi: 10.1007/pl00007479. PubMed PMID: 10985616.
33. Toftager M, Bogstad J, Bryndorf T, Lossl K, Roskaer J, Holland T, et al. Risk of severe ovarian hyperstimulation syndrome in GnRH antagonist versus GnRH agonist protocol: RCT including 1050 first IVF/ICSI cycles. *Hum Reprod*. 2016;31(6):1253-64. Epub 20160408. doi: 10.1093/humrep/dew051. PubMed PMID: 27060174.
34. Ragni G, Vegetti W, Riccaboni A, Engl B, Brigante C, Crosignani PG. Comparison of GnRH agonists and antagonists in assisted reproduction cycles of patients at high risk of ovarian hyperstimulation syndrome. *Hum Reprod*. 2005;20(9):2421-5. Epub 20050512. doi: 10.1093/humrep/dei074. PubMed PMID: 15890731.
35. Sousa M, Cunha M, Teixeira da Silva J, Oliveira C, Silva J, Viana P, Barros A. Ovarian hyperstimulation syndrome: a clinical report on 4894 consecutive ART treatment cycles. *Reprod Biol Endocrinol*. 2015;13(1):66. Epub 20150623. doi: 10.1186/s12958-015-0067-3. PubMed PMID: 26100393; PubMed Central PMCID: PMCPMC4477314.
36. West S, Lashen H, Bloigu A, Franks S, Puukka K, Ruokonen A, et al. Irregular menstruation and hyperandrogenaemia in adolescence are associated with polycystic ovary syndrome and infertility in later life: Northern Finland Birth Cohort 1986 study. *Hum Reprod*. 2014;29(10):2339-51. Epub 20140801. doi: 10.1093/humrep/deu200. PubMed PMID: 25085801; PubMed Central PMCID: PMCPMC4164150.
37. Bergh PA, Navot D. Ovarian hyperstimulation syndrome: a review of pathophysiology. *J Assist Reprod Genet*. 1992;9(5):429-38. doi: 10.1007/BF01204048. PubMed PMID: 1482837.
38. Prakash A, Mathur R. Ovarian hyperstimulation syndrome. *The Obstetrician & Gynaecologist*. 2013;15(1):31-5. doi: 10.1111/j.1744-4667.2012.00153.x.
39. Delvigne A, Rozenberg S. Epidemiology and prevention of ovarian hyperstimulation syndrome (OHSS): a review. *Hum Reprod Update*. 2002;8(6):559-77. doi: 10.1093/humupd/8.6.559. PubMed PMID: 12498425.
40. Mandelbaum RS, Bainvoll L, Violette CJ, Smith MB, Matsuzaki S, Klar M, et al. The influence of obesity on incidence of complications in patients hospitalized with ovarian hyperstimulation syndrome. *Arch Gynecol Obstet*. 2022;305(2):483-93. Epub 20210709. doi: 10.1007/s00404-021-06124-5. PubMed PMID: 34241687.
41. Hasegawa H, Nesvick CL, Erickson D, Cohen SC, Yolcu YU, Khan Z, et al. Gonadotroph Pituitary Adenoma Causing Treatable Infertility and Ovarian Hyperstimulation Syndrome in Female Patients: Neurosurgical, Endocrinologic, Gynecologic, and Reproductive Outcomes. *World Neurosurg*. 2021;150:e162-e75. Epub 20210305. doi: 10.1016/j.wneu.2021.02.115. PubMed PMID: 33684575.
42. Verwoerd GR, Mathews T, Brinsden PR. Optimal follicle and oocyte numbers for cryopreservation of all embryos in IVF cycles at risk of OHSS. *Reproductive biomedicine online*. 2008;17(3):312-7.
43. Steward RG, Lan L, Shah AA, Yeh JS, Price TM, Goldfarb JM, Muasher SJ. Oocyte number as a predictor for ovarian hyperstimulation syndrome and live birth: an analysis of 256,381 in vitro

fertilization cycles. *Fertil Steril*. 2014;101(4):967-73. Epub 20140123. doi: 10.1016/j.fertnstert.2013.12.026. PubMed PMID: 24462057.

44. Van Waart J, Kruger TF, Lombard CJ, Ombelet W. Predictive value of normal sperm morphology in intrauterine insemination (IUI): a structured literature review. *Hum Reprod Update*. 2001;7(5):495-500. doi: 10.1093/humupd/7.5.495. PubMed PMID: 11556497.

45. Youssef M, Van der Veen F, Al-Inany HG, Mochtar MH, Griesinger G, Nagi Mohesen M, Aboufoutouh I, vanWely M. Gonadotropin-releasing hormone agonist versus HCG for oocyte triggering in antagonist-assisted reproductive technology. *Cochrane Database Syst Rev*. 2014;(10).

46. Krishna D, Dhoble S, Praneesh G, Rathore S, Upadhaya A, Rao K. Gonadotropin-releasing hormone agonist trigger is a better alternative than human chorionic gonadotropin in PCOS undergoing IVF cycles for an OHSS Free Clinic: A Randomized control trial. *Journal of human reproductive sciences*. 2016;9(3):164.

47. O'Neill KE, Senapati S, Maina I, Gracia C, Dokras A. GnRH agonist with low-dose hCG (dual trigger) is associated with higher risk of severe ovarian hyperstimulation syndrome compared to GnRH agonist alone. *J Assist Reprod Genet*. 2016;33(9):1175-84. Epub 20160627. doi: 10.1007/s10815-016-0755-8. PubMed PMID: 27349252; PubMed Central PMCID: PMC5010813.

48. Chen CH, Tzeng CR, Wang PH, Liu WM, Chang HY, Chen HH, Chen CH. Dual triggering with GnRH agonist plus hCG versus triggering with hCG alone for IVF/ICSI outcome in GnRH antagonist cycles: a systematic review and meta-analysis. *Arch Gynecol Obstet*. 2018;298(1):17-26. Epub 20180329. doi: 10.1007/s00404-018-4751-3. PubMed PMID: 29600322.

49. He Y, Tang Y, Chen S, Liu J, Liu H. Effect of GnRH agonist alone or combined with different low-dose hCG on cumulative live birth rate for high responders in GnRH antagonist cycles: a retrospective study. *BMC Pregnancy Childbirth*. 2022;22(1):172. Epub 20220302. doi: 10.1186/s12884-022-04499-0. PubMed PMID: 35236312; PubMed Central PMCID: PMC8892730.

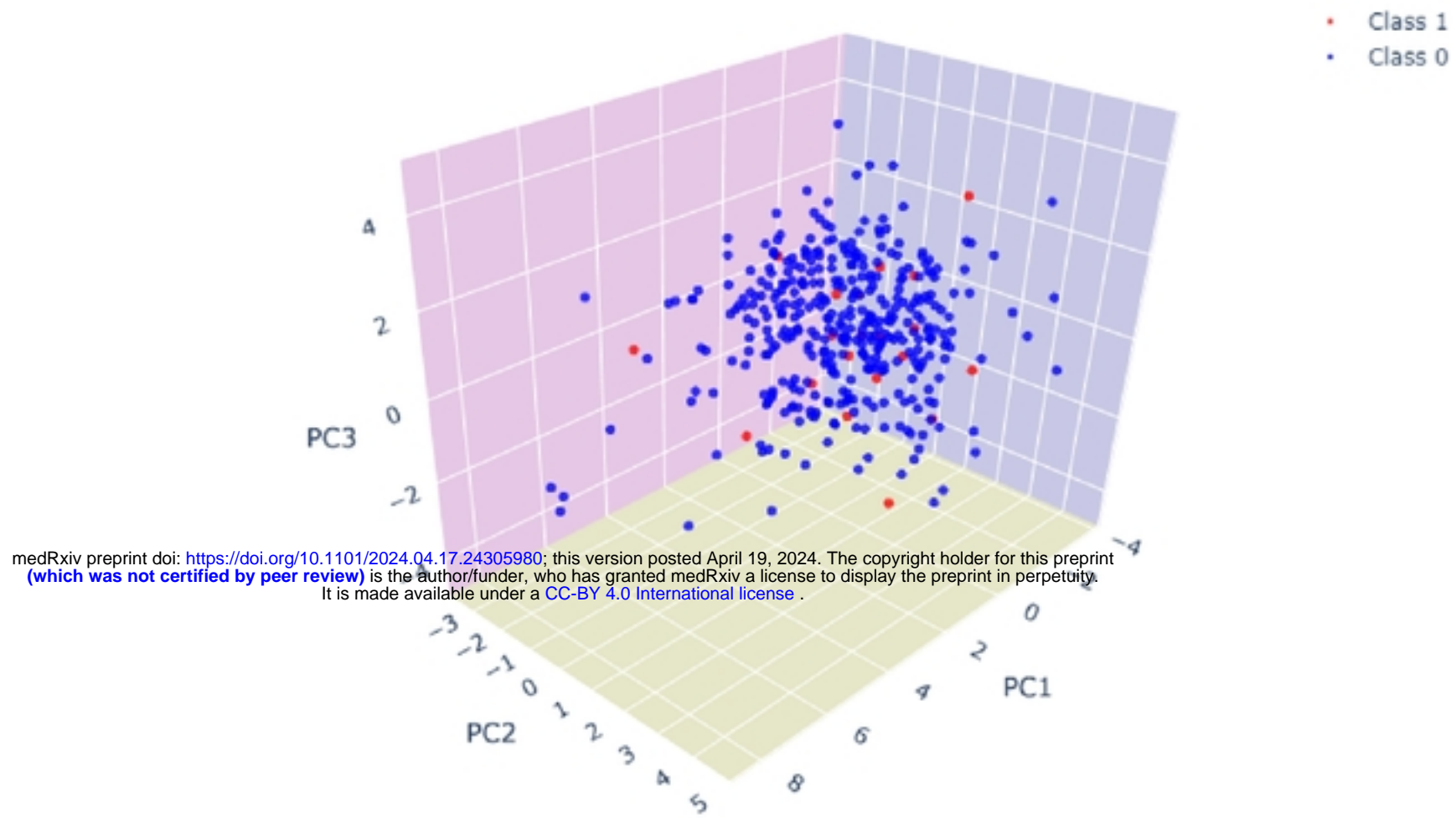
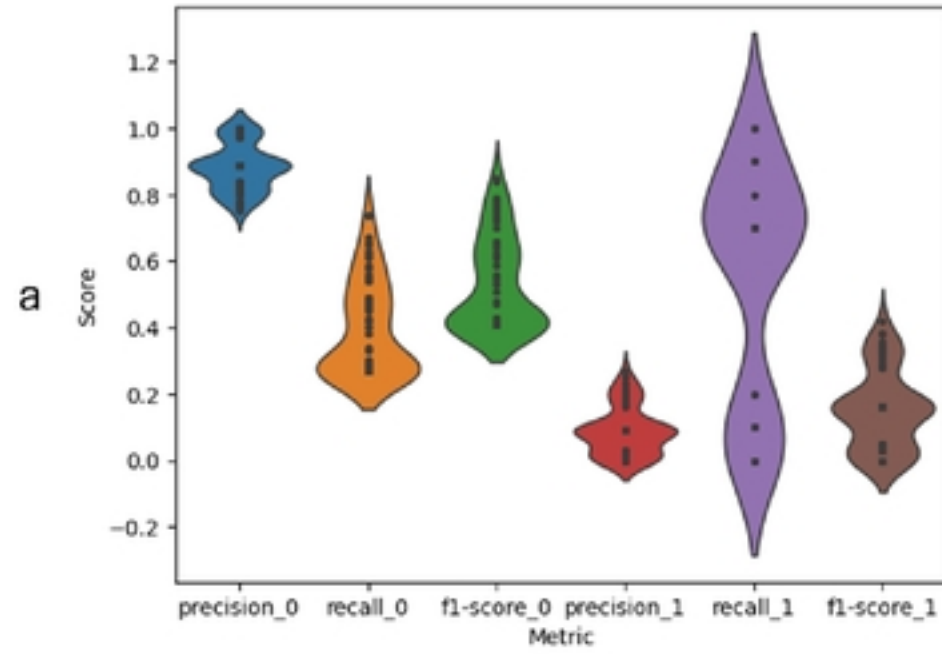


Figure 1: 3D PCA scatter plot showcasing the distribution of two classes within the transformed feature space



medRxiv preprint doi: <https://doi.org/10.1101/2024.04.17.24305980>; this version posted April 19, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

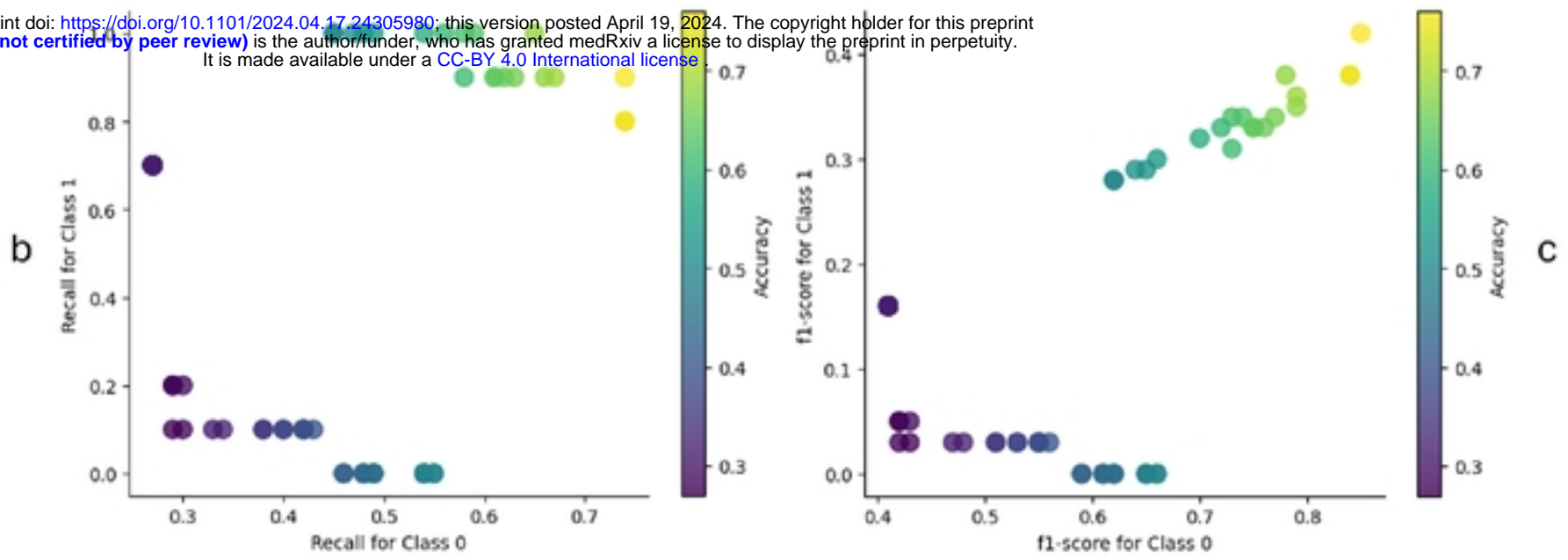


Figure 2: Analysis of the developed ML models: a) metrics distributions, b) trade-off between recall scores, and c) trade-off between f1-scores

medRxiv preprint doi: <https://doi.org/10.1101/2024.04.17.24305980>; this version posted April 19, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

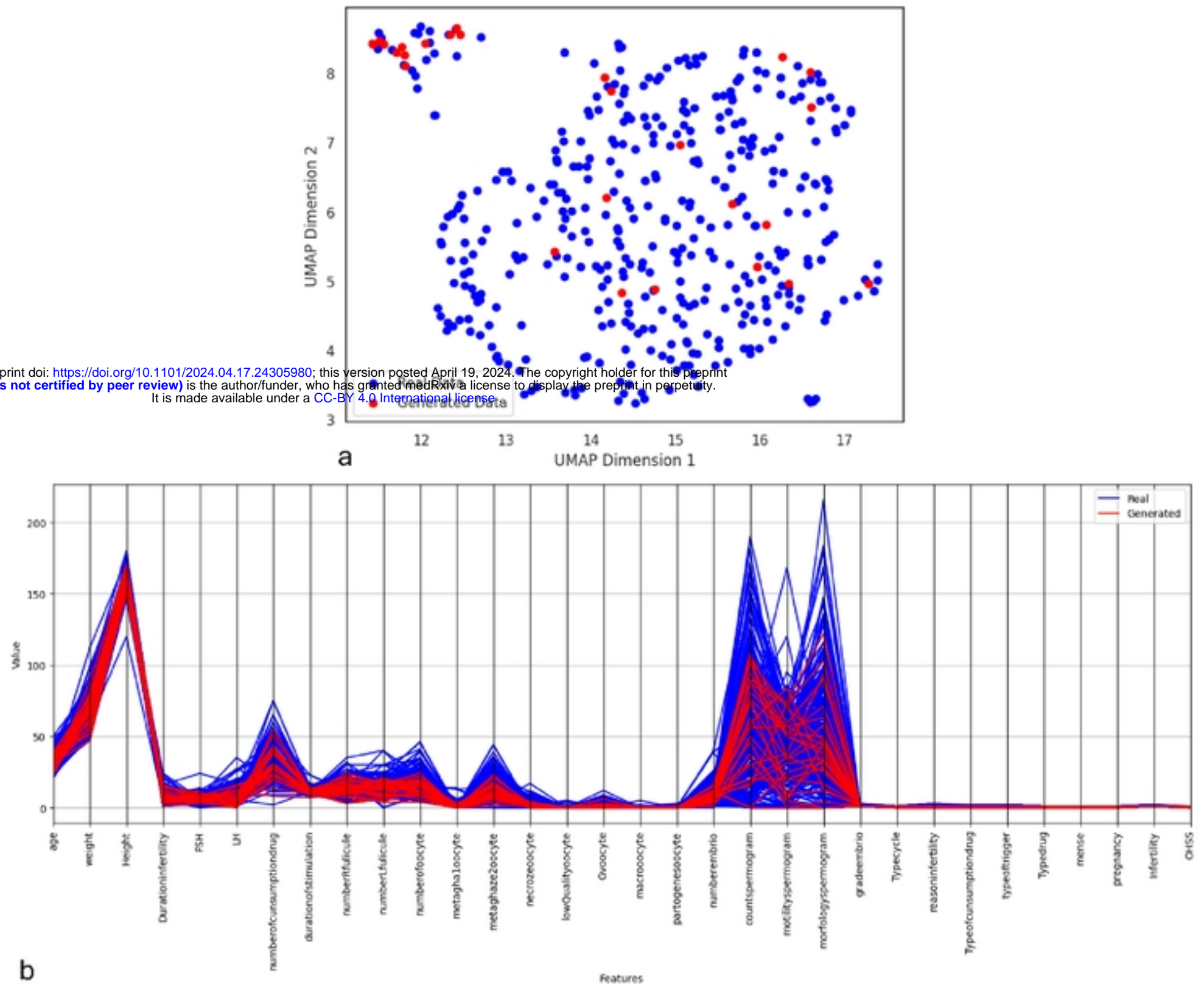
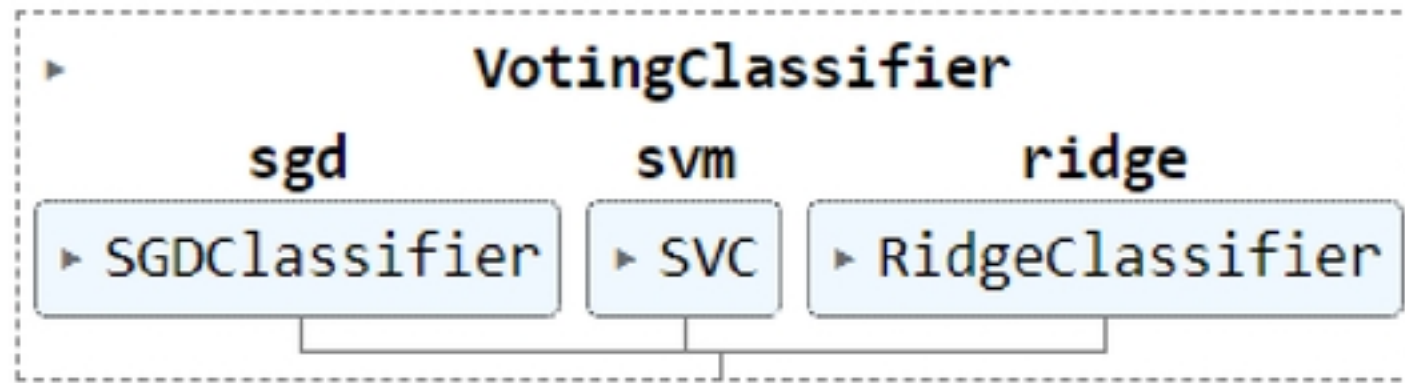


Figure 3: Analysis of the generated data: a) UMAP projection of the real data vs generated data, b) Parallel Coordinates plot of real vs generated Data



medRxiv preprint doi: <https://doi.org/10.1101/2024.04.17.24305980>; this version posted April 19, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Figure 4. Representation of the best ensemble model

medRxiv preprint doi: <https://doi.org/10.1101/2024.04.17.24305980>; this version posted April 19, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

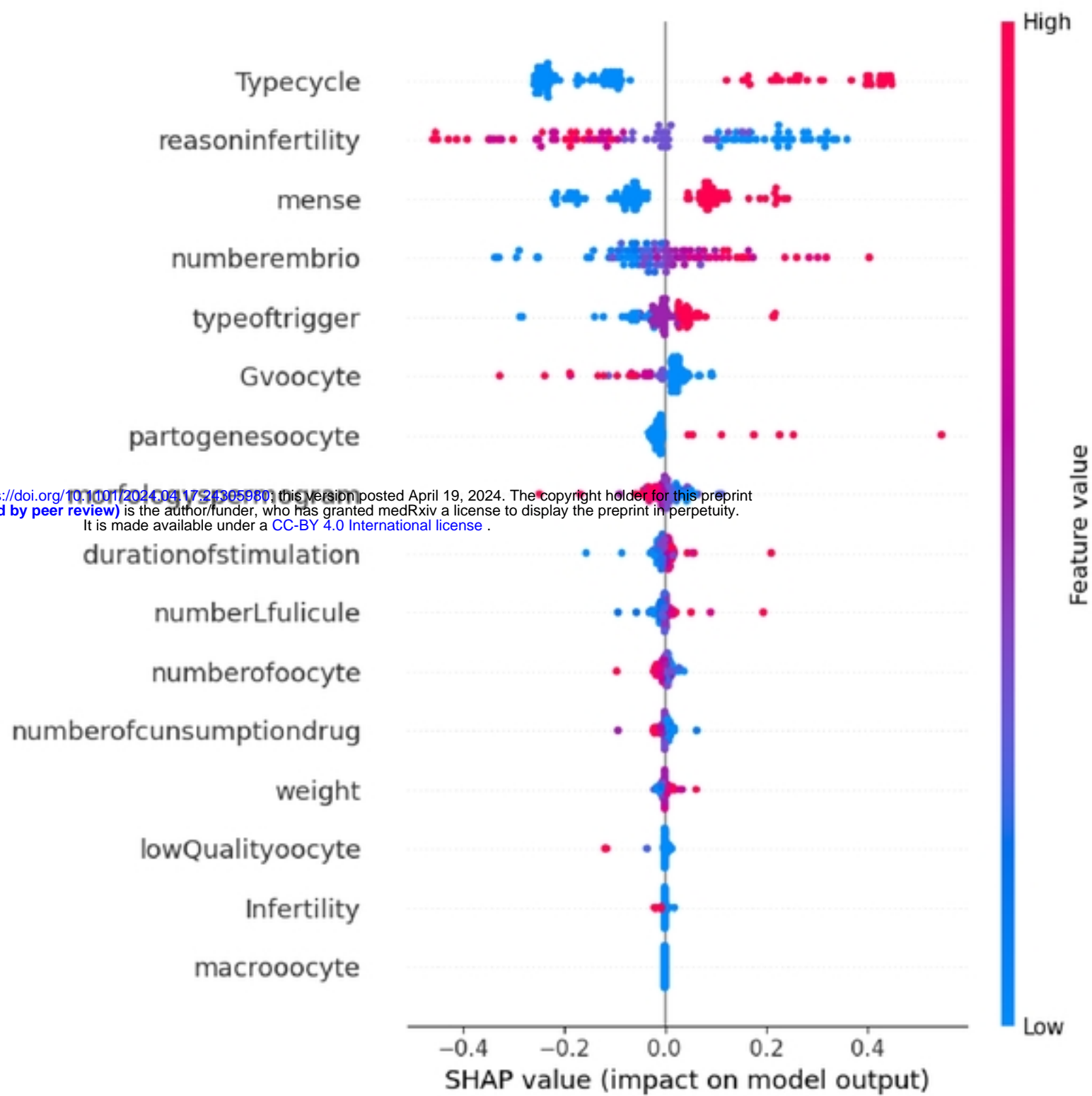


Figure 5: SHAP analysis of the ensemble model for prediction of OHSS