scientific reports

OPEN

Ensemble machine learning models for sperm quality evaluation concerning success rate of clinical pregnancy in assisted reproductive techniques

Ameneh Mehrjerd[1,2], Toktam Dehghani[2], Mahdie Jajroudi[2,4], Saeid Eslami[2,4,5✉], Hassan Rezaei[3] & Nayyereh Khadem Ghaebi[6]

This study aimed to investigate the influence of various sperm quality characteristics, including morphology, motility, and count, on the success rates of clinical pregnancy achieved through assisted reproductive technologies (ART) such as in-vitro fertilization (IVF), intracytoplasmic sperm injection (ICSI), and intrauterine insemination (IUI). The secondary objective was to assess the impact of these sperm parameters on the clinical pregnancy rate that resulted in the detection of a fetal heartbeat during the 11th week of gestation, a crucial milestone in successful ART-derived pregnancies. The researchers employed a retrospective analysis, evaluating data from 734 couples undergoing IVF/ICSI and 1197 couples undergoing IUI across two infertility centers. Exclusion criteria included cases involving donated eggs or sperm, surrogate uteri, and infertile couples with combined male and female factors. Five ensemble machine-learning models were utilized to predict the clinical pregnancy success rates. The Random Forest (RF) model achieved the highest mean accuracy (0.72) and area under the curve (AUC) (0.80), outperforming the other models for both IVF/ICSI and IUI procedures. The Shapley Additive Explanations (SHAP) value analysis revealed that for IUI cycles, all three sperm parameters (morphology, motility, and count) had significant negative impacts on the prediction of clinical pregnancy success. In contrast, for IVF/ICSI cycles, sperm motility had a positive effect, while sperm morphology and count were negative factors. In cycles with 1 to 5 retrieved eggs, sperm motility, and count, they positively affected the clinical pregnancy rate. The study also identified cut-off values for sperm count, with 54 and 35 being the respective thresholds for IVF/ICSI and IUI. Additionally, a significant cut-off point 30 was found for the sperm morphology parameter across all procedures. This study underscores the immense potential of leveraging ensemble machine learning models with traditional sperm quality assessments. This integrated approach can elevate the precision and personalization of clinical decision-making in the field of assisted reproductive technologies, ultimately offering more hope and better outcomes for couples struggling with infertility.

Keywords Infertility, Sperm quality, IVF, ICSI, IUI, Machine learning algorithms

Infertility is an issue that affects couples whose pregnancy fails after 12 months of unprotected sexual activity. The infertility prevalence rate is 15–20%, and almost 40–50% of all cases are caused only by the male factor[1–3]. Assisted Reproduction Technology (ART) treatments are defined as medical procedures aiming to achieve pregnancy, but a myriad of intrinsic and extrinsic factors influence their success. There are numerous ART available, including

[1]Department of Psychiatry and Psychotropic, Faculty of Medicine, University Medicine Greifswald, Greifswald, Germany. [2]Department of Medical Informatics, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran. [3]Department of Mathematics, Statistics and Computer Sciences, University of Sistan and Baluchestan, Zahedan, Iran. [4]Pharmaceutical Sciences Research Center, Institute of Pharmaceutical Technology, Mashhad University of Medical Sciences, Mashhad, Iran. [5]Department of Medical Informatics, Amsterdam UMC (location AMC), University of Amsterdam, Amsterdam, The Netherlands. [6]Department of Obstetrics and Gynecology, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran. ✉email: s.eslami.h@gmail.com; EslamiS@mums.ac.ir

intrauterine insemination (IUI), in vitro fertilization (IVF), and intracytoplasmic sperm injection (ICSI)[4]. IUI is the first-line treatment option and also the most cost-effective strategy for males with unexplained or mild infertility, which is defined as a single abnormal finding of the semen analysis or a total motile sperm count between 10 and 20 $\times$ 10^6 mL[5]. Patients with infertility using the IVF method have a total fertilization failure rate of 5% of IVF cycles[6]. Many variable parameters are considered to predict effective methods and select a suitable treatment for each infertile couple. Female and male factors are both considered. Male factors such as concentration, morphology, motility, volume, total number, and vitality are used to represent abnormal sperm in male infertility[1,7], whereas female parameters such as type and duration of infertility, number of mature follicles, endometrial thickness, tubal factor, and various seminal parameters[5] are caused female infertility.

In this paper, different ensemble models of machine learning were developed to predict the possibility of clinical pregnancy as well as infertility based on the effect of sperm parameters. The primary objective aims to determine how sperm quality influences the initial establishment of a clinical pregnancy, as indicated by a positive pregnancy test or gestational sac visualization in the 5th week of pregnancy. The secondary objective, on the other hand, focuses on the progression of clinical pregnancies to a more advanced stage, where a fetal heart rate can be detected around the 11th week of pregnancy. This objective assesses the impact of sperm quality on the viability and sustainability of the pregnancy beyond the initial implantation and early development stages.

Results

In this study, most-selective ensemble models are developed to predict the success rate of clinical pregnancy. These models are compared from different aspects of two treatments, IVF/ICSI and IUI.

- **Implementation**.

In this research, we utilized Python and various frameworks, including Scikit-learn, Pandas, and NumPy, to develop, evaluate, and visualize all the models. These tasks were performed in the Google Collaboratory web-based IDE.

- **Comparison of Accuracy and AUC of Models**:

In the first step, the results were achieved by each approach, and their criteria in two treatments were compared in terms of accuracy and AUC, as in Fig. 1. Area Under the Curve Operating Characteristic (ROC) curve, which plots the true positive rate (TPR) on the y-axis(AUC), is a performance metric used to evaluate the predictive power of a binary classification model. It is beneficial in situations where the goal is to distinguish between two classes, such as the success rate of clinical pregnancy of treatments, IVF/ICSI, and IUI. The AUC metric is derived from the Receiver and the false positive rate (FPR) on the x-axis as the decision threshold for the classification model is varied. The AUC range is between 0 and 1, with 0.5 representing a random classifier (no predictive power) and 1 representing a perfect classifier (100% accurate). Generally, an AUC value greater than 0.7 indicates a reasonably good model, while an AUC value greater than 0.8 indicates a robust model. Among these models, Bagging and Random Forest accomplished the best mean values of accuracy and AUC for the IVF/ICSI procedure, with Bagging accuracy (0.74) and AUC (0.79) and Random Forest accuracy (0.72) and AUC (0.80). However, for the IUI procedure, the mean accuracy of Bagging and Random Forest is equal to 0.85, and the AUC of Random Forest is higher than that of Bagging; it almost gained better results. Overall, it is observed that Random Forest is a suitable model for two treatments.

- **Comparison SHAP VALUE of Models**:

In the second step, the value of sperm parameters in predicting the success rate of the clinical pregnancy is illustrated based on the best model (Random Forest) from the previous step. In Fig. 2, the impact of each feature in predicting the model and their relationship are shown for each treatment procedure using Shapley Additive Explanations (SHAP) values. The SHAP value impact on a model plot is a powerful visualization that helps understand the contribution of each feature to the model's predictions. This plot shows the SHAP values for each feature, ranked by their importance. The bars' position along the x-axis indicates the direction and magnitude of the feature's impact. Bars to the right (positive SHAP values) suggest that the feature increases the model's output. In contrast, bars to the left (negative SHAP values) indicate that the feature decreases the model's production. Notable, in the IUI procedure, sperm parameters (morphology, motility, and count) had significant negative impacts on clinical pregnancy according to the Random Forest model. However, the coefficient of influence of sperm motility on predicting clinical pregnancy has been reported to be positive for some patients having IVF/ICSI procedures. In contrast, the morphology and count factors had negative impacts.

- **Evaluation of the association between sperm parameters and pregnancy**:

In addition, the statistical results of the association between sperm parameters and pregnancy are shown in Fig. 3.

Figure 3 also includes a box plot that visually represents the distribution of sperm parameters in the successful and unsuccessful pregnancy groups. The box in the figure indicates the interquartile range (IQR), which encompasses the middle 50% of the data, with the line inside the box representing the median. The whiskers extend to the minimum and maximum values within 1.5 times the IQR of the first and third quartiles, respectively. At the same time, any data points outside this range are considered outliers and are plotted individually.
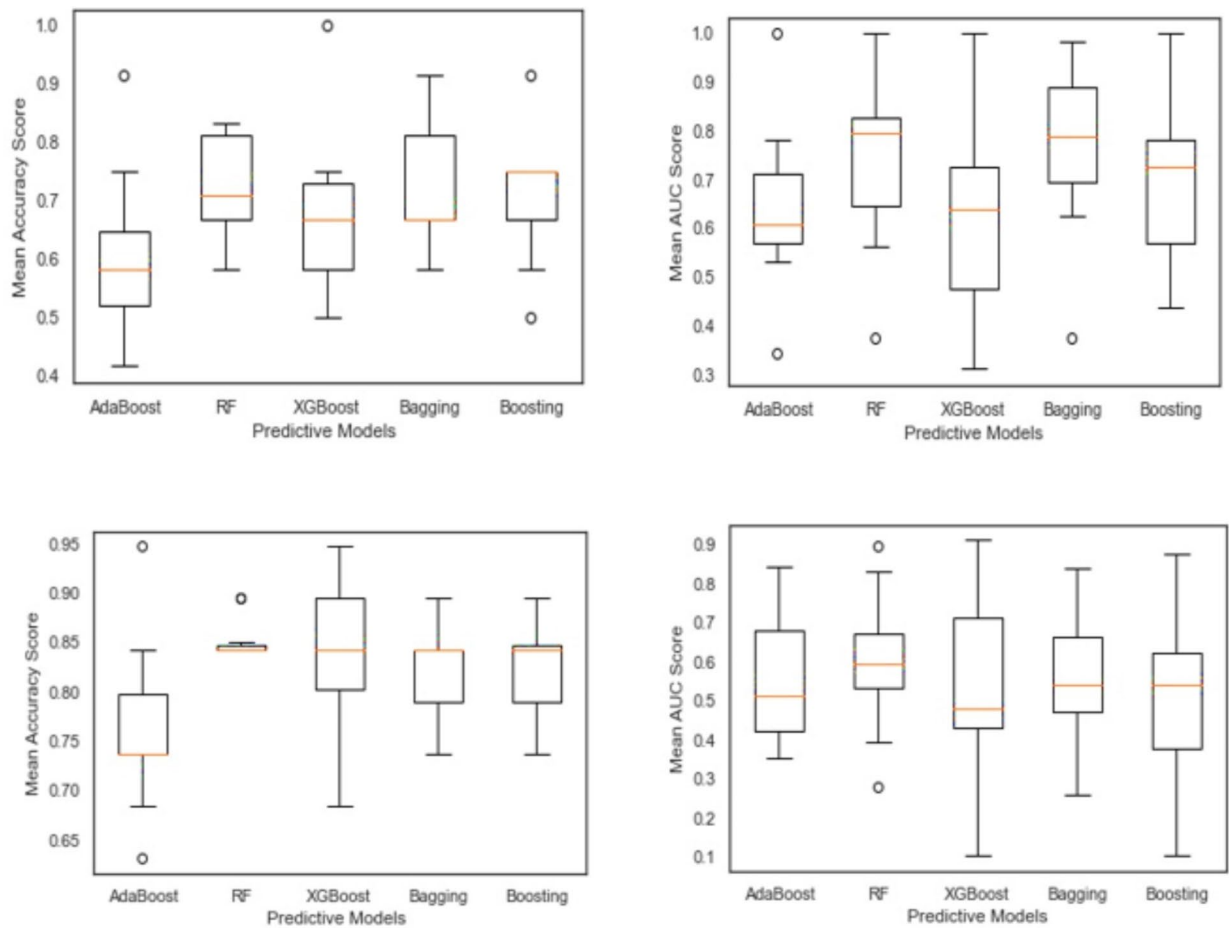
Figure 1. Comparison of accuracy and AUC of models in IVF/ICSI and IUI treatment.

In this figure, the differences between the values of sperm parameters in the successful and unsuccessful pregnancy groups were evaluated using Student's t-test. The results showed that in clinical pregnancies after the IVF/ICSI procedure, there was a significant difference in sperm morphology and number between the successful and unsuccessful groups (p-value < 0.05). Specifically, the morphological parameter significantly differed in both pregnancy outcomes in the IUI procedure. In contrast, the count parameter did not differ substantially between the successful and unsuccessful groups.

These detailed statistical analyses underscore the significant associations between specific sperm parameters and pregnancy outcomes, highlighting the importance of morphology in predicting success in both IVF/ICSI and IUI procedures.

- **Comparison of sperm parameters in treatments based on cycles**:

Determining a cut-off value for the feature can provide valuable information for gynecologists. In this study, for each of these sperm parameters, a cut-off was computed to derive an evidence-based decision rule. Hence, a contingency table evaluated the validity of the results. For CPR purposes, the optimal sperm count parameter cut-offs were 54 million per ml after sperm selection (p-value: 0.02, 95%-CIs: 1.05–2.13) and 35 (p-value: 0.03, 95%-CIs: 1.06–2.86) in IVF/ICSI and IUI procedures, respectively. The cut-off point for the morphological parameter in each of the procedures approach was also found to be 30 million per ml after sperm selection with a p-value of 0.05. However, no significant cut-off for the motility factor was obtained for the methods. The cut-offs were also investigated for FHR, and the results were similar to those of CPR.

Additionally, statistical information about sperm parameters in different courses for two methods is provided based on the WHO recommendation [8]. From the results obtained in Tables 1 and 2, it can be seen that most of the patients in the two procedures were in the first course of the procedure. The results showed the fact that in the IVF/ICSI groups for most patients in the first and second course of the procedure, sperm parameters are below average set by WHO. However, in the IUI group, sperm parameters report a higher-than-average level (Table 2). Furthermore, for patients with sperm parameters above the mean value reported by WHO, the IUI method reached better results, and the two courses of procedure were more successful. Moreover, for patients whose sperm parameters are below average, IVF procedures achieved better results. Especially for the second course of procedure, they achieved higher success rates in CPR and FHR.
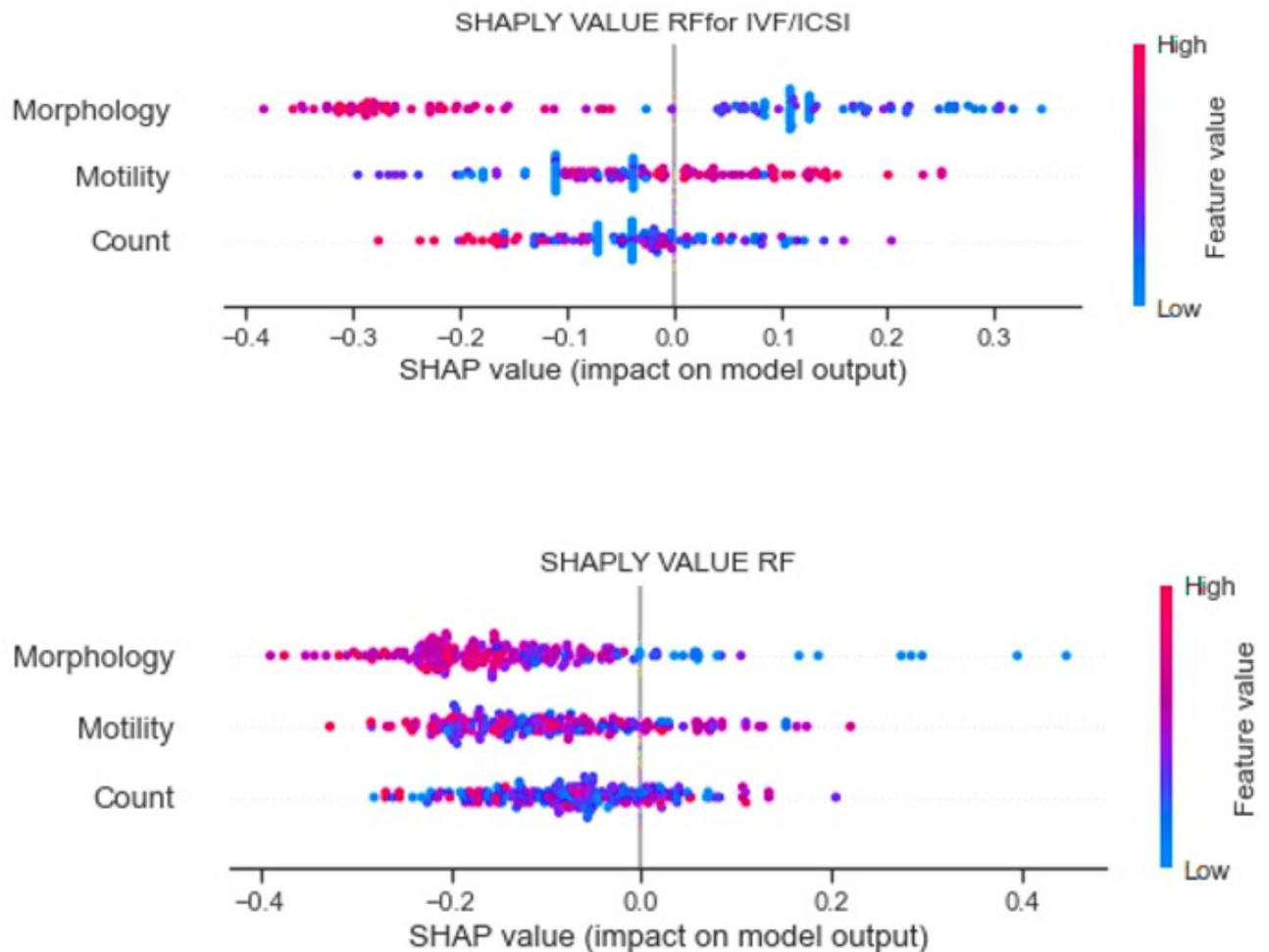
Figure 2. Impact of Sperm parameters on RF model prediction in IVF/ICSI and IUI.procedures.

Discussion

In this study, the most selective ensemble models were developed to predict the success rate of clinical pregnancy. These models were evaluated for two treatments: IVF/ICSI and IUI. Various machine learning models, including Bagging and Random Forest, were employed and compared regarding accuracy and the area under the receiver operating characteristic curve (AUC). Among the five ensemble machine learning models evaluated, the Random Forest model achieved the highest accuracy (0.72) and area under the ROC curve (0.80) in predicting the success rate of clinical pregnancy based on sperm quality parameters (morphology, motility, and count) for both IVF/ICSI and IUI procedures. SHAP value analysis revealed that for IUI, all three sperm parameters had significant adverse impacts. In contrast, for IVF/ICSI, sperm motility had a positive effect, but morphology and count were negative factors.

Statistical analysis showed significant differences in morphology and count between successful and unsuccessful pregnancy groups for IVF/ICSI, and only morphology differed significantly for IUI. Optimal cut-off values were identified as 30 million/mL for morphology in both procedures, 54 million/mL for the count in IVF/ICSI, and 35 million/mL for IUI. Comparing across treatment cycles, IVF/ICSI patients with below-average sperm parameters had improved success rates in the second cycle. At the same time, IUI was more effective for patients with above-average sperm parameters, especially with two cycles.

According to previous studies, there are limitations for standard laboratories in sperm analysis and "normal" or "reference" values as cut-offs[7]. Our identification of specific cut-off values for sperm specification as predictors of successful CPR is consistent with a 2014 systematic review by Ombelet et al.[9], which suggested total motile sperm count and morphology as potential predictive variables for IUI success. One study published in 2019 showed that prewash total motile sperm is a poor predictor of live births in IUI cycles, with no correlation between live births and prewash total motile sperm counts of 2 million sperm[10]. Another study in 2021 revealed that the relation between IUI and prediction variables in patients with unexplained infertility was not statistically significant. Also, the difference between quartiles of total progressive motile sperm count (TPMSC) with live birth rate and clinical pregnancy rate was not statistically significant for the patients older than 40 or the TPMSC less than 10%. Moreover, they reported only one clinical pregnancy and no live birth[11]. In another research
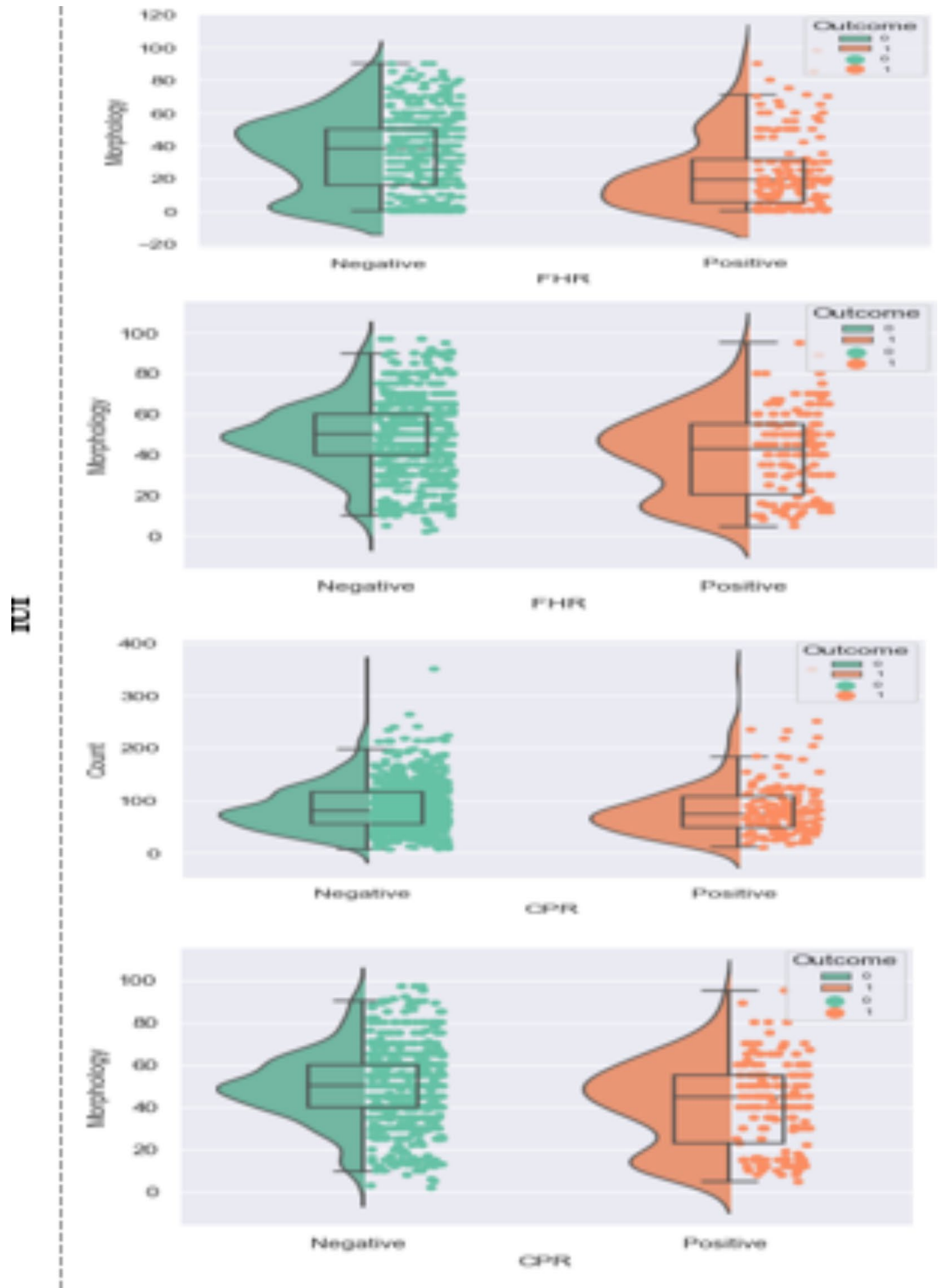
Figure 3. Correlation of sperm parameters with CPR as clinical pregnancy rate and FHR as fetal heart rate for IUI and IVF/ICSI.

in 2020, under the IVF method, for male factors with total sperm count greater than 5 million, there was no correlation between low total testosterone levels and semen parameter changes.[12,13]

In this study, the importance of sperm quality in the efficacy of assisted reproduction IVF higher prewash total sperm count values positively impacted cumulative success rates in cycles with few retrieved oocytes (1 to 5). Still, it did not affect the outcome of cycles with an average (6 to 10) or high (> 10) number of retrieved oocytes[4]. Our findings are consistent with a retrospective analysis of over 22,000 assisted reproductive technology (ART) cycles by Villani et al.[14], highlighting the predictive value of sperm parameters for ART success. They reported that sperm motility and concentration were significant predictors of clinical pregnancy in both IUI and IVF/
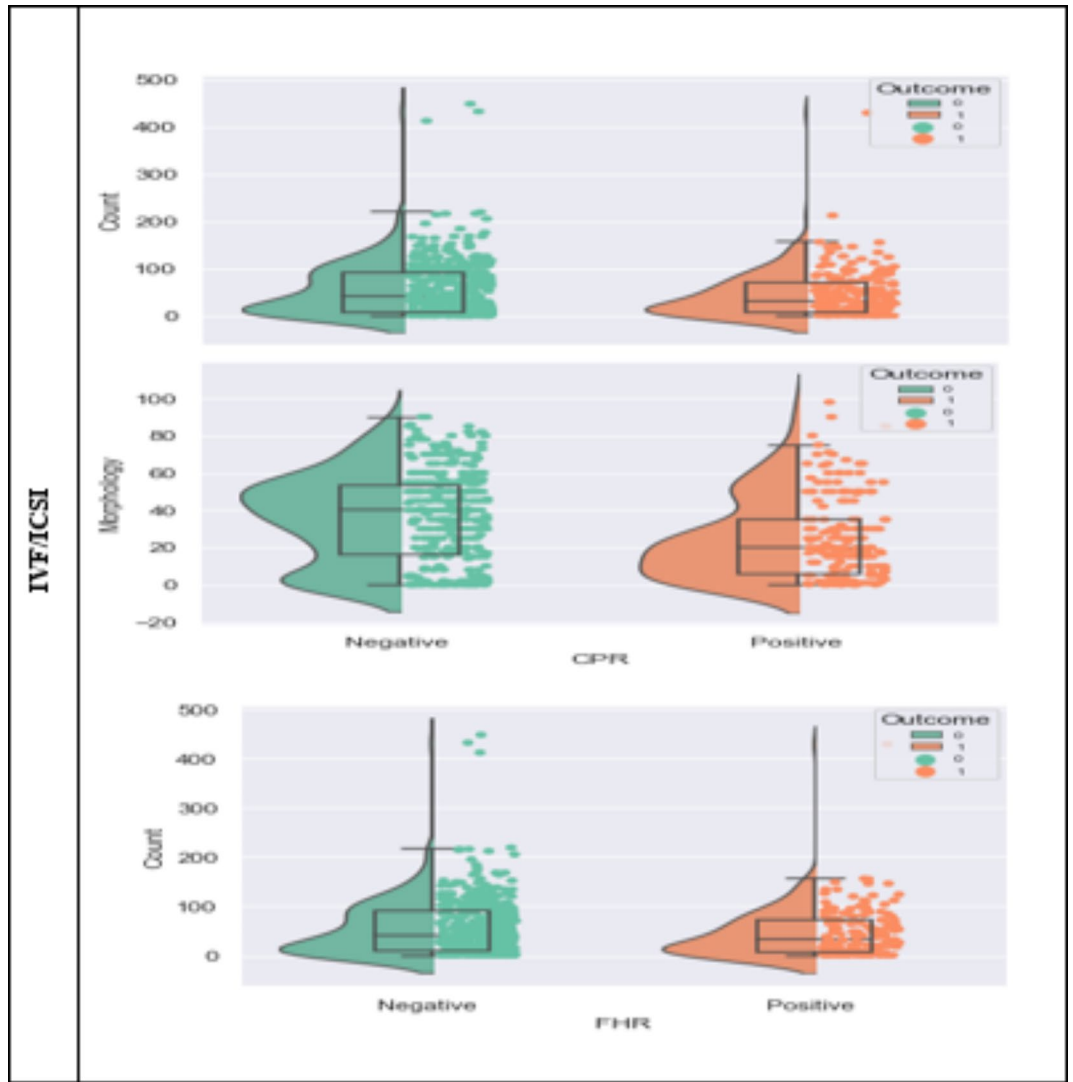
Figure 3. (continued)

Patients Couples (Total Couple No.=599)	IVF/ICSI Cycle 1 (Couples No.=433)	IVF/ICSI Cycle 2 (Couples No.=124)	IVF/ICSI Cycle 3 (Couples No.=42)
Spermiogram below Avg (Pregnant No.=174) (CPR=(38/5%),FHR=(30/4%))	(Pregnant No.= **128)** **(CPR = 37/5%;** **FHR = 20/3%**	(Pregnant No.=**36)** **(CPR = 100%;** **FHR = 97/2%**	(Pregnant No.=10) (CPR = 20%; FHR = 20%
Spermiogram Avg (Pregnant No.=5) (CPR = 0.0%, FHR = 0.0%))	(Pregnant No.=3) (CPR = 0.0%, FHR = 0.0%)	(Pregnant No.=1) (CPR = 0.0%, FHR = 0.0%))	(Pregnant No.=1) (CPR = 0.0%, FHR = 0.0%))
Spermiogram above Avg (Pregnant No.=140) (CPR = 22/1%, FHR = 18/5%)	(Pregnant No.=103) (CPR = 22/3%, FHR = 15/53%)	(Pregnant No.=26) (CPR = 23/7%, FHR = 23.7%)	(Pregnant No.=**11)** **(CPR = 36.3%,** **FHR = 36/3%**

Table 1. Comparison of sperm parameters in IVF/ICSI procedure based on cycle. *Spermiogram Average for Sperm Count (10*6) WHO Standard Count is 33–46; Sperm Motility is 38–42; and Sperm Morphology is 30–40. ** CPR as clinical pregnancy rate and FHR as fetal heart rate.

ICSI procedures. Similarly, our study identified morphology and count as significant predictors, with optimal cut-off values for improved clinical pregnancy.

In addition, the difference between the values of sperm parameters in successful and unsuccessful pregnancy groups in clinical pregnancy illustrated a significant difference in values of the morphology of the two procedures. Furthermore, the difference in the count parameter could be observed between the successful and unsuccessful groups. Moreover, the negative impact of the morphology factor in the two methods must be considered. These

Patients (Couples No.=954)	IUI Cycle 1 (Couples No.=775)	IUI Cycle 2 (Couples No.=150)	IUI Cycle 3 (Couples No.=29)
Spermiogram below Avg (Pregnant No.=9) (CPR = 0.0%, FHR = 0.0%))	(Pregnant No.=8) (CPR = 0.0%, FHR = 0.0%))	(Pregnant No.=1) (CPR = 0.0%, FHR = 0.0%))	(Pregnant No.=0) (CPR = 0.0%, FHR = 0.0%))
Spermiogram Avg (Pregnant No.= 3) (CPR = 1(33/3), FHR = 1(33/3))	(Pregnant No.=2) (CPR = 33/3%, FHR = 33/3%)	(Pregnant No.=1) (CPR = 33/3%, FHR = 33/3%)	(Pregnant No.=0) (CPR = 33/3%, FHR = 33/3%)
Spermiogram above Avg (Pregnant No.= 515) (CPR = 69 (13/82); FHR = 61 (12/2))	(Pregnant No.=409) (CPR = 13/93%; FHR = 12/22%	(Pregnant No.=76) (CPR = 15/78%; FHR = 14/47%)	(Pregnant No.=14) (CPR = 0.0%, FHR = 0.0%))

Table 2. Comparison of sperm parameters in IUI procedure based on the cycle. *Spermiogram Average for Sperm Count (10*6) WHO Standard Count is 33–46; Sperm Motility is 38–42; and Sperm Morphology is 30–40. ** CPR as clinical pregnancy rate and FHR as fetal heart rate.

findings suggest that different sperm parameters may have varying degrees of influence on the success rates of different assisted reproductive techniques, which could explain the contrasting results observed in previous studies.

Despite providing valuable insights into the role of sperm quality parameters in predicting ART success, our study has several limitations. The sample size and diversity may not be sufficient for broad generalization, and the retrospective design can introduce biases. Critical factors like the female partner's age and ovarian reserve were not controlled, potentially confounding results. Although we used a Random Forest model, comparison with other algorithms could enhance understanding. The lack of external validation raises concerns about the model's generalizability. Time-dependent factors influencing ART outcomes and long-term results, such as live birth rates, were not considered. Finally, ethical and privacy issues related to data use are significant challenges for implementing predictive models in clinical practice. Addressing these limitations in future research is essential for enhancing predictive accuracy and clinical applicability.

the proposed ensemble machine learning models, especially the Random Forest model, can effectively predict the success rate of clinical pregnancy and the likelihood of infertility based on sperm parameters. This study highlights the importance of integrating advanced machine learning techniques with traditional sperm quality assessments to enhance predictive accuracy. Such models provide a robust tool for clinical decision-making in assisted reproductive technologies, offering a more personalized and precise approach to evaluating and improving ART outcomes. Future research should aim to refine these models further and validate them across diverse populations and clinical settings to maximize their applicability and reliability.

Materials and methods

- **Data collecting and preprocessing**.

In our study, we analyzed sperm parameters to predict the success rate of clinical pregnancy in two distinct categories: IVF/ICSI and IUI procedures. IVF/ICSI procedures involve more advanced techniques, such as direct injection of sperm into the egg (ICSI) (in vitro procedure). Conversely, in the IUI procedure, where sperm must naturally traverse the female reproductive tract and fertilize the egg. This study collected data from two university centers, an affiliated infertility center, and a private center in Iran. We included all treated patients who had received a maximum of two courses. Additionally, no information was collected on donated eggs, embryos, or surrogates. Female patienOPUts, who are older than 40 years and those whose courses were not completed or who have lost more than 70% of the required clinical factors (missing values) are excluded. Couples with multiple infertility factors were excluded from the study to evaluate the effect of sperm parameters more precisely. The couples (courses) under IVF/ICSI 734 and 1197 under IUI, as shown in Table 3, represent the couples' characteristics. After removing exclusion criteria, 599 couples (courses) and 954 couples (courses) for IVF/ICSI and IUI were selected, respectively. Couples with multiple infertility factors were excluded from the study to evaluate the effect of sperm parameters more precisely.

Sperm stimulation and preparation protocol After 3 to 5 days of abstinence on the day of coitus, semen samples were collected in a sterile oocyte pick-up (OPU). About 15 min after arrival, samples are processed (a maximum of 60 min from ejaculation time). First, 15 cc of sperm were transferred into a sterilized container containing a gradient containing one cc of 45% and 90 cc of 90% silica salt. It was centrifuged at 2500 to 3000 RPM for 15 min. Then, the top layer is removed with a syringe, and 3 to 4 cc of hamsF10 is added to the motile sperm deposition solution and centrifuged again at 2500 to 3000 RPM for 10 min and then, remove the top solution with a syringe. At this stage, if the IUI method is used for treatment, we keep 0.5 cc at the end of the syringe, and in the IVF/ICSI method, the last 250 cc is kept in the syringe. After sperm preparation, sperm analysis parameters such as morphology, motility, and sperm concentration were evaluated using computer-assisted semen analysis (CASA) systems. CASA systems eliminate the influence of subjective factors in manual analysis by utilizing advanced image processing and pattern recognition algorithms to automatically capture and analyze sperm characteristics from digital images or videos.

Couples Characteristic		
	IVF/ICSI	**IUI**
Infertility causes Mix factors Male factors Female factors Unexplainable factors	135 329 79 191	243 398 226 330
Female Characteristic		
AGE (Median)	30	29
Endometrium Thickness (Median)	9	7
Antral Follicle Count (Median)	100	100
Number of Occcytes Collected (Median)	8.5	
Number of Transferred Embryos (percent) 2 3 4	376 of 734 286 of 734 74 of 734	Not mentioned
Quality of Transferred Embryos (GI) (percent) 0 1 2 3 4	60 of 734 215 of 734 113 of 734 113 of 734 25 of 734	Not mentioned
Quality of Transferred Embryos (GII) (percent) 0 1 2 3 4	378 of 734 254 of 734 76 of 734 23 of 734 6 of 734	Not mentioned
Quality of Transferred Embryos (GIII) (percent) 0 1	726 of 734 10 of 734	Not mentioned
Male Characteristic		
Age (Median)	34	32
FSH (Median)	6.9	6.7
Sperm Motility% (Median)	40	50
Sperm Morphology (Median)	30	78
Sperm Count (Median)	39.5	50
Total Gonadotropin Dose (Median)	28	4

Table 3. Couples characteristics under IVF/ICSI and IUI.

- **Ovarian Stimulation Protocol in IUI patients**: In the IUI method, first, a transvaginal ultrasound is performed on the second day of the cycle, and the number of follicles is evaluated. If there is no cyst in the ovaries, the thickness of the endometrium, and the patient's condition are suitable, gonadotropin drugs such as FSH (F. signal) are prescribed. Six days after the initial ultrasound, a follow-up ultrasound was conducted to assess the follicles' growing process. The physician prescribed the HCG ampoule if the follicle was greater than 16 mm. It was carried out around 36 h following egg retrieval.
- **Ovarian Stimulation Protocol in IVF/ICSI patients**: The ovarian stimulation protocol in IVF/ICSI is slightly different from the IUI method. After a primary transvaginal ultrasound on the second day of the cycle, gonadotropin-releasing hormone analogues (GnRH) and steroid hormone inhibitors such as clomiphene citrate, letrozole, and FSH were prescribed. If the treatment regimen includes agonists, these drugs are given in the luteal phase before ovulation is stimulated. If up to 2 dominant follicles with a size of 18 to 20 mm were observed, HCG ampoules were injected. Then, the dominant follicles were retrieved from the ovary by the puncture.

More information about protocols is provided in supplementary data.

- **Methodology**.

Machine learning approaches aim to identify hidden patterns and extract information from data. Machine learning provides a variety of methods and algorithms for predicting the output of specific input predictors that can be used for clinical decision-making.

- **Step 1: Preprocessing method**.

We used ensemble methods to predict clinical pregnancy for predicting the outcome variable; three features were utilized, which included morphology, motility, and count. The examined data were used to train and test prediction models in proportions of 80–20%, respectively. Imbalanced data can present challenges in ML models and lead to biases towards the majority class and inadequate performance on the minority class. To

8

mitigate the effects of the imbalanced dataset, we employed resampling strategies during the preprocessing phase. Resampling techniques can be categorized into oversampling of the minority class, undersampling of the majority class, and hybrid sampling methods[1]. Oversampling involves replicating samples of the minority class, utilizing methods such as Synthetic Minority Oversampling Technique (SMOTE). We utilized SMOTE to balance the data since the data investigated in both treatments is unbalanced regarding the number of samples in the classes[15]. Also, k-fold cross-validation was used to train our ML models. This technique involved multiple splits of the dataset, providing ample data for training and testing. The dataset was divided into several folds (k = 10), and the performance of the ML models was evaluated on each fold.

- **Step 2: Modeling process**.

Here, we capitalized on the advantages of ensemble methods, which combine multiple models with a single type of algorithm to create an optimal prediction model. Bagging, Boosting, Random Forest, Xgboost, and ADABoosting are popular ensemble algorithms. In the following paragraphs, brief descriptions of models are presented:

- **Bagging** is a widely used technique that uses decision trees and significantly improves model stability by increasing accuracy, reducing variance, and removing the challenge of overfitting. Indeed, it assembles the predictions of several weak models to obtain the most accurate predictions.
- **Random forest** is one of the bagging methods. In this method, poor learning methods are combined to build a robust model, and one of its advantages is its robustness toward missing data.
- **Bootstrapping** is a sampling technique using an alternative method to select a sample from a set. Following that, the learning algorithm is applied to the selected samples.
- **AdaBoost** is a boosting technique that identifies and weights unclassified data in each iteration.
- **Gradient boosting (Boost)** strengthens the gradient based on the difference between the predictor and the correct value.

Step 3: Model evaluation.
To assess the performance of the obtained models, we utilized well-known metrics such as accuracy and Area Under the Receiver Operating Characteristic (ROC) Curve (AUC). These metrics for model performance have been used in several studies to assess infertility prediction, thereby facilitating the comparison of the results of this study with those of earlier studies[16].

Data availability
The datasets used and analyzed during the current study are available from the corresponding author upon reasonable request.

References
1. Cannarella, R. et al. FSH dosage effect on conventional sperm parameters: a meta-analysis of randomized controlled studies. *Asian J. Androl.* **22** (3), 309 (2020).
2. Maharlouei, N. et al. Prevalence and pattern of infertility in Iran: a systematic review and meta-analysis study. *Women's Health Bull.*, : pp. 63–71. (2021).
3. Moein, M. R. et al. Prevalence of primary infertility in Iranian men; a systematic review. *Men's Health J.* **5** (1), e12–e12 (2021).
4. Zacà, C. et al. Sperm count affects the cumulative birth rate of assisted reproduction cycles about ovarian response. *J. Assist. Reprod. Genet.* **37** (7), 1653–1659 (2020).
5. Gubert, P. G. et al. Number of motile spermatozoa inseminated and pregnancy outcomes in intrauterine insemination. *Fertility Res. Pract.* **5** (1), 1–9 (2019).
6. Anbari, F. et al. Does sperm DNA fragmentation have a negative impact on embryo morphology and morphokinetics in IVF programme? *Andrologia.* **52** (11), e13798 (2020).
7. Cooper, T. G. et al. World Health Organization reference values for human semen characteristics. *Hum. Reprod. Update.* **16** (3), 231–245 (2010).
8. Findeklee, S. et al. Correlation between total sperm count and sperm motility and pregnancy rate in couples undergoing intrauterine insemination. *Sci. Rep.* **10** (1), 7555. https://doi.org/10.1038/s41598-020-64578-0 (2020).
9. Ombelet, W. et al. Semen quality and prediction of IUI success in male subfertility: a systematic review. *Reprod. Biomed. Online.* **28** (3), 300–309 (2014).
10. Mankus, E. B. et al. Prewash total motile count is a poor predictor of live birth in intrauterine insemination cycles. *Fertil. Steril.* **111** (4), 708–713 (2019).
11. Lin, H. et al. Role of the total progressive motile sperm count (TPMSC) in different infertility factors in IUI: a retrospective cohort study. *BMJ open.* **11** (2), e040563 (2021).
12. Gu, Y. et al. Predicting medication adherence using ensemble learning and deep learning models with large scale healthcare data. *Sci. Rep.* **11** (1), 1–13 (2021).
13. Rosly, R. et al. Comprehensive study on ensemble classification for medical applications. *Int. J. Eng. Technol.* **7** (2.14), 186–190 (2018).
14. Villani, M. T. et al. Are sperm parameters able to predict the success of assisted reproductive technology? A retrospective analysis of over 22,000 assisted reproductive technology cycles. *Andrology* **10** (2), 310–321. https://doi.org/10.1111/andr.13123 (2022).
15. Vaegter, K. K. et al. Which factors are most predictive for live birth after in vitro fertilization and intracytoplasmic sperm injection (IVF/ICSI) treatments? Analysis of 100 prospectively recorded variables in 8,400 IVF/ICSI single-embryo transfers. *Fertil. Steril.* **107** (3), 641–648.e2. https://doi.org/10.1016/j.fertnstert.2016.12.005 (2017).
16. Liu, L. et al. Machine learning algorithms to predict early pregnancy loss after in vitro fertilization-embryo transfer with fetal heart rate as a strong predictor. *Comput. Methods Programs Biomed.* **196**, 105624 (2020).

Declarations

Competing interests
The authors declare no competing interests.

Authors Statements
The authors confirm that all methods carry out relevant guidelines and regulations. Also, we confirm that informed consent was obtained from all subjects and their legal guardian(s).
.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-73326-7.

Correspondence and requests for materials should be addressed to S.E.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.