Modeling Temporal Dynamics of the ICSI Procedure for Fertilization Outcome Prediction

Cristina Castera[1], Nélida Mirabet-Herranz[1 [0000−0002−3372−1192]], Laura Carrión[2 [0009−0009−0394−9733]], Lorena Bori[2 [0000−0003−1495−2646]], Marcos Meseguer[2,3 [0000−0003−4833−2752]], and Valery Naranjo[1 [0000−0002−0181−3412]]

[1] Instituto de Investigación e Innovación en Bioingeniería, HUMAN-tech, Universitat Politècnica de València, Valencia, Spain
ccaslap@alumni.upv.es, nemiher@htech.upv.es, vnaranjo@dcom.upv.es
[2] IVIRMA Global Research Alliance, IVI Foundation, Instituto de Investigación Sanitaria La Fe (IIS La Fe), 46026 Valencia, Spain
Laura.Carrion@ivirma.com, Lorena.Bori@ivirma.com, marcos.meseguer@ivirma.com
[3] IVIRMA Global Research Alliance, IVIRMA Valencia, Plaza de la Policía Local 3, 46015 Valencia, Spain

Abstract. Intracytoplasmic sperm injection (ICSI) is a widely used technique in assisted reproduction (AR), where subtle variations during the microinjection process can influence fertilization success. However, many of these factors are currently assessed subjectively, making systematic and reproducible analysis challenging. In this work, we propose an automatic system for analyzing ICSI procedures from microscopic videos, enabling the extraction of quantitative variables that describe the interaction between the injection pipette and the oocyte. The system integrates temporal and spatial segmentation and feature extraction to obtain static and dynamic biomechanical descriptors, defined as injection velocity and the temporal evolution of the angle signal. To evaluate the relevance of these variables, different experimental configurations are compared, combining clinical features with point-wise and dynamic descriptors. While clinical variables provide a robust baseline for fertilization outcome with an AUC of 72.1%, the addition of dynamic features results in a notable performance gain, resulting in an AUC of 80%. This highlights the importance of modeling the temporal dynamics of the ICSI and its potential to support clinical decision-making in AR. In addition, a SHAP-based interpretability analysis reinforced those conclusions, as temporal descriptors provided complementary and highly informative cues for the prediction, supporting explainable AI in AR healthcare.

Keywords: ICSI · Temporal Modeling · Assisted Reproduction · Computer Vision · Model Interpretability

1 Introduction

Intracytoplasmic Sperm Injection (ICSI) is one of the most widely used techniques in Assisted Reproduction (AR). Since its introduction in 1992 by Palermo

et al. [1], it has revolutionized infertility treatment and enabled the birth of millions of children worldwide. Originally, this technique was indicated for cases of prior fertilization failure or severe male factor infertility. However, its use has progressively expanded, and ICSI is now routinely performed in a wide range of AR cycles, even in the absence of a clear male factor indication [2, 3].

Despite its high fertilization rates, the overall success of the procedure remains limited by multiple biological and technical factors that are not fully understood or standardized. These include both gamete quality and procedural aspects related to the microinjection process, such as the interaction between the pipette and the oocyte, the oolemma rupture pattern, or the injection dynamics [4–6]. Subtle variations in these factors may significantly impact oocyte activation and subsequent embryo development. In clinical practice, many of these factors are assessed subjectively by the embryologist, introducing inter-operator variability and limiting reproducibility. Furthermore, the dynamic nature of the microinjection process limits the analysis using static variables or manual observations, making it difficult to identify objective patterns associated with fertilization success. Similar challenges have been addressed in other medical domains, such as surgical workflow analysis and skill assessment, where the modeling of procedural dynamics and tool–tissue interaction has been shown to be critical for performance evaluation [7–9].

Nowadays, advances in Deep Learning (DL), particularly in computer vision models for medical data [10] have enabled the development of automated tools for biomedical data analysis. In AR, these approaches have been mainly applied to image-based tasks, such as embryo quality assessment using convolutional neural networks (CNNs) [11], as well as to temporal sequence analysis in time-lapse studies [12]. However, their application to the ICSI procedure remains limited and often relies on global image features or black-box models with reduced interpretability [13]. In this context, some studies have explored the use of visual information to predict events during microinjection. In particular, microscopic images prior to microinjection have been shown to contain relevant information about the mature stage of the oocyte [15, 17] as well as mechanical properties of the oocyte, such as its susceptibility to rupture [16]. Nevertheless, these approaches are mainly based on global or texture descriptors, without explicitly modeling geometric variables of the procedure nor systematically evaluating the contribution of temporal information.

In this context, this work proposes the use of geometric and dynamic variables directly derived from the ICSI procedure to analyze whether these variables contain relevant information for predicting fertilization success. In particular, the angle generated by the resistance of the oocyte membrane during pipette insertion is computed at a specific instant of the procedure. Additionally, the temporal evolution of this angle, as well as the injection velocity of the pipette into the oocyte, are analyzed. These variables are studied as a representation of the mechanical behavior of the pipette–oocyte system. Furthermore, the study examines whether incorporating these variables alongside conventional clinical features improves the predictive performance of the models.

The main contributions of this work are threefold: (1) the automatic computation of dynamic variables during the ICSI procedure, including injection velocity and the angle associated with membrane deformation; (2) the improvement of predictive models by combining these computed variables with traditional clinical features, leading to an increase in AUC from 72.1% to 80%; and (3) the development of an end-to-end pipeline that enables embryologists to predict, from an ICSI procedure video, whether the fertilization outcome will be successful.

The rest of the paper is organized as follows. Section 2 presents the state of the art from two perspectives: feature engineering-based and learning-based approaches. Section 3 describes the proposed pipeline for predicting the fertilization outcome from a microscopy video. Section 4 reports the experimental results and Section 5 summarizes the contributions and outlines future work.

2 Related work

This section reviews the most relevant methods for analyzing the ICSI procedure, organized into two categories: hand-crafted approaches and learning-based models. The former rely on explicitly defined variables derived from expert knowledge, while the latter learn representations directly from data, including images and temporal sequences.

2.1 Feature engineering-based approaches

Early studies on ICSI analysis have focused on the evaluation of specific procedural variables, such as the oolemma rupture pattern, injection speed, or sperm deposition site, defined based on clinical expertise and obtained through direct observation [5, 6, 4]. Several works have shown that small variations in these factors can significantly affect fertilization outcomes; however, they are often assessed qualitatively or subjectively, introducing inter-operator variability and limiting reproducibility [5, 4].

Some approaches have attempted to model these relationships using classical machine learning (ML) techniques. Algorithms such as Random Forest (RF), Gradient Boosting (GB), and Extreme Gradient Boosting (XGBoost) [25–27] have been applied in AR for tasks such as predicting live birth outcomes, embryo implantation, and clinical pregnancy from tabular clinical data [11, 18], showing competitive performance on datasets of limited size. Nevertheless, their application to the ICSI procedure remains limited. A key limitation is their reliance on static or point-wise observations, which fail to capture the dynamic nature of the microinjection process and restrict their reproducibility and systematic use in clinical environments.

2.2 Learning-based approaches

With the rise of ML, and especially DL, new approaches have emerged that are able to learn representations directly from biomedical data, reducing the

dependency on manually defined features [14]. In AR, these methods have been mainly applied to predict clinical outcomes from visual data. Most works focus on post-fertilization stages, such as embryo quality assessment using convolutional neural networks (CNNs) [11], as well as automated embryo scoring and ranking systems to predict implantation and live birth outcomes [18–20].

More recently, they have been extended to earlier stages, such as the evaluation of oocyte quality and competence [23]. One of the most promising approaches in this field is Magenta^{TM}, a DL model trained with oocyte images and clinical variables to predict their developmental potential [22]. Several studies with this AI-based tool have found a correlation between Magenta^{TM} scores and fertilization and blastocyst rates [21, 22], reaching an AUC of 0.67 [24]. Despite its innovative nature, such methods still pay limited attention to explicitly modeling the dynamics of the ICSI procedure using interpretable geometric variables.

In particular, the mechanical interaction between the injection pipette and the oocyte remains largely unexplored as a predictive source. To date, only a few studies have addressed specific procedural aspects; for instance, Yagi et al. [16] proposed predicting oolemma rupture in Piezo-ICSI using texture descriptors and Support Vector Machine (SVM), achieving an AUC of 0.90. However, even these specialized models often rely on global image features, failing to model the temporal evolution of geometric variables or providing the interpretability required for clinical decision-making [13].

Consequently, there is a clear lack of research systematically analyzing how quantifiable procedural variables (such as geometric or dynamic features derived directly from the microinjection process) impact fertilization prediction. Moreover, the relative contribution of such variables compared to approaches that do not include them has not been clearly evaluated. This gap motivates the present work, which proposes the use of interpretable geometric variables derived directly from the ICSI procedure specifically, the injection angle and its temporal evolution, as well as the injection velocity to evaluate whether their inclusion on top of clinical variables enhances the predictive performance of fertilization outcome models.

3 Method

This section describes the proposed methodology, including the different stages of the pipeline for analyzing the ICSI procedure. The approach integrates temporal segmentation, spatial segmentation, feature extraction, and predictive modeling.

3.1 Overview of the proposed method

This work aims to develop an automatic system for analyzing the ICSI procedure from microscopic videos, extracting quantitative variables and assessing whether they contain relevant information to predict whether an oocyte will be fertilized or not. The proposed system follows a structured pipeline that combines computer vision and ML techniques. As a starting point, a dataset of long-duration

microscopic videos is available, in which multiple phases of the procedure may appear, including sperm selection, oocyte manipulation, and one or more microinjections. From these videos, the processing workflow is divided into several main stages.

First, a temporal segmentation of the videos is performed to identify and isolate exclusively the intervals corresponding to the microinjection process, that is, from the moment the pipette begins insertion into the oocyte until its subsequent withdrawal. This stage allows discarding the remaining non-relevant information present in the videos. Next, a spatial analysis of the images is carried out, including the segmentation of key structures: the injection pipette and the oocyte. Tracking techniques are then applied to these segmentations in order to capture the temporal evolution of the positions and relationships between both structures throughout the procedure. From this analysis, variables describing the pipette-oocyte interaction are extracted, particularly the angle generated by membrane resistance during pipette insertion (triangle angle) and the injection velocity. Finally, these variables, together with additional clinical information, are used as input to supervised learning models in order to analyze their contribution to predicting the fertilization outcome (non-fertilized / fertilized). Fig. 1 shows an overview of the proposed pipeline.

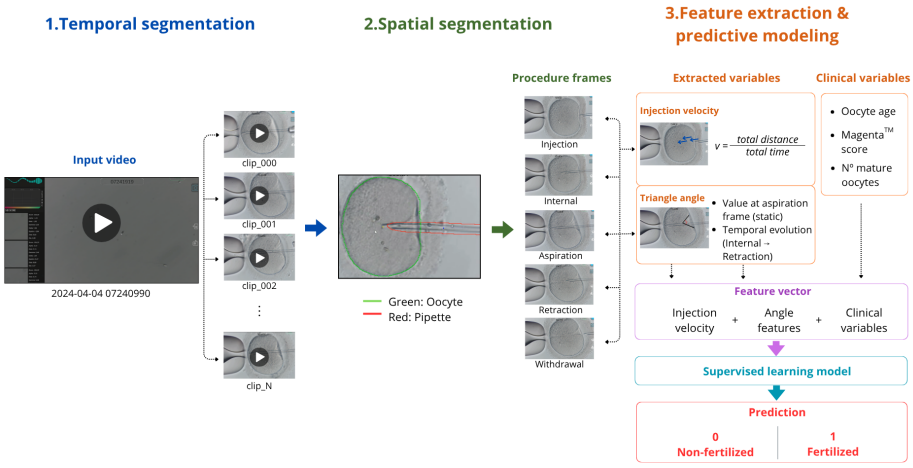

Fig. 1. Pipeline Overview

3.2 Temporal segmentation

To focus the analysis on the relevant stages of the ICSI procedure, a temporal segmentation of the videos is performed to identify intervals corresponding to the microinjection phase, as illustrated in the first block of Fig. 1. To this

end, a ResNet18 architecture [28] pre-trained on ImageNet is adopted, with its final layer replaced by a binary classifier to distinguish between relevant frames (microinjection) and non-relevant frames.

Given an input frame, the model outputs the probability of belonging to the microinjection phase. These frame-level predictions are temporally aggregated by applying high probability threshold to select relevant frames, which are then grouped into continuous segments based on temporal adjacency. To ensure robustness, only segments exceeding a minimum duration are retained, while short or isolated detections are discarded. This process enforces temporal consistency and enables the identification of coherent microinjection events. The resulting intervals are used to extract clips centered on the pipette–oocyte interaction, removing irrelevant information and reducing variability across videos. This preprocessing step enables subsequent stages to focus on the critical phases of the procedure and facilitates robust feature extraction.

3.3 Spatial segmentation

Once the temporal intervals of interest have been identified, a spatial analysis of the sequences is performed to locate and characterize the key structures of the ICSI procedure. In particular, the injection pipette and the oocyte are segmented. To this end, a U-Net architecture [29] is employed and trained in a supervised manner using pixel-wise annotations of both structures, enabling the extraction of binary masks from the input frames. Fig. 2 shows a representation of the model output, where the contours derived from these masks are overlaid on the original images, with the oocyte depicted in green and the pipette in red.

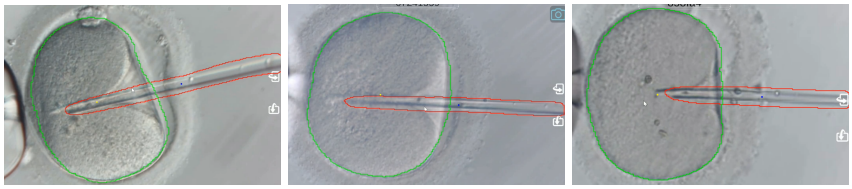

Fig. 2. Examples of segmentation results. The oocyte is shown in green and the injection pipette in red.

From the resulting segmentations, relevant geometric properties are computed. In the case of the oocyte, its contour and centroid are extracted. For the pipette, its main axis is estimated by linear fitting over the segmented mask, as well as the position of its tip, defined as the point closest to the oocyte. Temporal tracking is then performed by directly associating these features between consecutive frames, assuming temporal continuity in the movement. Since only one pipette and one oocyte are present per frame, tracking reduces to selecting the highest-confidence detection at each timestep, with no correspondence ambigui-

ties to resolve. To improve the stability of the trajectories, temporal smoothing is applied to the estimated positions.

Based on the trajectories of the pipette tip and the oocyte contour, their spatial relationships are analyzed to identify key events of the procedure. Specifically, the following events are defined based on geometric and temporal criteria:

- **Injection** is defined as the moment at which the distance between the pipette tip and the oocyte contour reaches its minimum, just before the pipette enters the oocyte.
- The temporal extent of the **internal stage** is delimited by the interval during which the pipette is fully inside the oocyte and remains in a stable position after insertion.
- The **aspiration** moment is set at the moment the spermatozoon is located within the first third of the pipette shaft, as estimated from the longitudinal motion signal projected onto the pipette axis.
- The **retraction** frame is defined by the reversal of pipette motion direction from insertion to withdrawal.
- **Withdrawal** is delimited by the exit of the pipette tip from the oocyte region.

These events define the key frames of the procedure, which are subsequently used for feature extraction. The second block of Fig. 1 shows representative examples, denoted as *procedure frames*, from which quantitative biomechanical features are extracted in the next stage.

3.4 Feature extraction

From the previous stages, a set of quantitative and interpretable variables describing the pipette–oocyte interaction is extracted and used as input to the predictive models. The primary descriptor is the triangle angle, which captures the geometric deformation of the oocyte membrane during pipette insertion. A U-Net model [29] trained on manually annotated frames is used to localize the triangular deformation region, from which the angle is defined by combining the pipette orientation and the local tangent of the oocyte contour at the insertion point. This angle is analyzed from two complementary perspectives. First, its value at the aspiration frame is considered as a static descriptor. Fig. 3 shows representative examples of this measurement.

Fig. 3. Examples of the angle generated by the deformation of the oocyte membrane, extracted at the aspiration frame.

Secondly, the temporal evolution between the internal and retraction frames is analyzed as a continuous signal $\theta(t)$, from which several biomechanical features are derived: amplitude descriptors (range, minimum value), variability measures (standard deviation over different temporal segments), energy-based features, and dynamic properties such as the number of peaks and stability-related metrics. A histogram-based discretization of the signal is also computed.

Additionally, injection velocity is extracted as a complementary biomechanical feature, computed as the total distance traveled by the pipette tip during penetration divided by the total time required for this movement.

Overall, three biomechanical variable groups are extracted: (i) the triangle angle at the aspiration frame, as a static descriptor; (ii) temporal descriptors derived from the angle signal; and (iii) the injection velocity, always included as a complementary feature. It should be noted that groups (i) and (ii) are mutually exclusive within any given model configuration, as each experimental setup uses either the static or the dynamic angle representation, but not both simultaneously.

3.5 Predictive modeling

To predict the outcome of the ICSI procedure, supervised learning models based on decision trees are employed, namely Random Forest (RF) [25], Gradient Boosting (GB) [26], and Extreme Gradient Boosting (XGBoost) [27].

The problem is formulated as a binary classification task, where each sample corresponds to a microinjection and is labeled according to the fertilization outcome (non-fertilized / fertilized), as provided by an expert embryologist from the IVI Valencia clinic. As input variables, both clinical variables (oocyte age, Magenta^{TM} score (AI-based morphological assessment of oocyte quality)[22] and number of mature oocytes) and automatically extracted biomechanical variables (triangle angle, temporal descriptors, and injection velocity) are used as input to the predictive model.

The comparison between the different feature representations, as well as between the evaluated models, is presented in Section 4.

3.6 Model interpretability

Interpretability is a key requirement in medical applications, as it promotes transparency, supports clinical validation, and fosters trust in model-driven decisions. To facilitate the adoption of the predictive model in a clinical setting, SHAP is integrated as a last step of the pipeline. SHAP [30] is a game-theoretic approach that assigns each feature a contribution value for a given prediction, enabling both global and local interpretability. In this work, SHAP values are computed for the best-performing model in order to analyze the impact of each variable on the prediction of the fertilization outcome, as well as to assess whether biomechanical features provide complementary information to clinical variables ultimately contributing to more informed and trustworthy decision-making in assisted reproduction.

4 Experiments

This section describes the experimental setup and presents the results of the evaluation of the proposed pipeline. First, the dataset and protocol are introduced. Then, the performance of each pipeline component is evaluated. Finally, the predictive models are assessed, and interpretability is provided through a SHAP analysis to understand the contribution of the input variables.

4.1 Dataset

The dataset collected for this study consists of real microscopic videos acquired during the ICSI procedure at the IVI Valencia clinic. It is composed of 139 full-length videos, each of which may contain one or more microinjection events, with durations ranging from approximately 10 to 30 minutes.

After temporal segmentation, a total of 804 microinjection sequences were identified, which constitute the units of analysis of the system. Each sequence is considered an independent sample, with an average duration of approximately 20 seconds. Each sample is associated with a label indicating the fertilization outcome (non-fertilized / fertilized), defined by the actual clinical outcome of the procedure. For each sequence, clinical variables are recorded, together with biomechanical features automatically extracted from the videos through the proposed pipeline.

For the training and evaluation of the classification model, a subset of 356 sequences is selected. This selection is performed through a data quality control process, in which samples presenting acquisition inconsistencies or incomplete information are excluded. This subset is composed by 103 non-fertilized cases and 253 fertilized cases. None of these sequences were used in any prior stage of the pipeline.

4.2 Implementation details and experimental setup

To ensure a rigorous evaluation protocol, all models are trained independently using task-specific annotations, with no fertilization outcome labels used in the segmentation stages. The modular design guarantees complete independence between components: intermediate models act solely as feature extractors, and no video used to test the final classifier overlaps with those used in previous stages.

The temporal segmentation model employs a ResNet18 architecture [28] pre-trained on ImageNet, using a transfer learning approach. The backbone parameters are frozen, and the final layer is replaced with a binary classifier. The model is trained on a labeled dataset of 353 invalid and 341 valid frames. Input images are resized to 224×224 pixels and normalized using standard ImageNet statistics. The model is trained for 10 epochs using the Adam optimizer (learning rate (LR) $= 1 \times 10^{-3}$) and Binary Cross-Entropy (BCE) loss, with a batch size of 32 and an 80/20 train-validation split.

The spatial segmentation models use U-Net architectures with a ResNet34 backbone pre-trained on ImageNet and fine-tuned end-to-end on the task-specific

annotated data. The oocyte and pipette segmentation models are trained on 278 annotated images, while the triangular deformation region model uses 276. All models are trained using images resized to 512×512 pixels, normalized with standard ImageNet statistics, and augmented using geometric and photometric transformations. The Adam optimizer is used (LR $= 1 \times 10^{-4}$), with an 80/20 train-validation split. For oocyte and pipette segmentation, a combined loss based on Dice [31] and BCE is used, while for the triangular deformation region only BCE is applied. The models are trained for 40, 50, and 25 epochs, respectively, with batch sizes of 4 (oocyte and triangular region) and 8 (pipette).

For fertilization outcome prediction, supervised tree-based models are used, including RF, GB, and XGBoost. The models are trained using 5-fold stratified cross-validation on the 356-sample subset, where all reported metrics are averaged across folds. The data are preprocessed using median imputation for missing values. To address class imbalance, model-specific weighting strategies are applied during training: RF uses the *class_ weight="balanced"* parameter, which adjusts weights inversely proportional to class frequencies, GB applies sample weights computed via *compute_ sample_ weight* with *class_ weight="balanced"*, passed directly to the *fit* method, and XGBoost uses the *scale_ pos_ weight* parameter, set to the ratio of negative to positive samples computed at each fold.

To analyze the contribution of different sources of information, three experimental configurations are defined based on the type of input variables. The variables are grouped into clinical variables and biomechanical features automatically extracted from the videos. The latter are further divided into (i) point-wise features, such as the pipette angle at the aspiration frame and the injection velocity, and (ii) dynamic features derived from the temporal evolution of the angle during injection. The evaluated configurations are as follows: **Config. 1**, referred to as the clinical configuration, uses only clinical variables; **Config. 2**, the clinical + point-wise configuration, incorporates static geometric features in addition to the clinical variables; and **Config. 3**, the clinical + dynamic configuration, integrates temporal features derived from the angle signal together with the clinical information. Consistent hyperparameter configurations are used across models, with 200 estimators and maximum depths between 3 and 6, and a LR of 0.05 for boosting-based models.

Model performance is evaluated using Accuracy, F1-score, and Area Under the ROC Curve (AUC), with AUC considered the primary metric as it reflects the discriminative ability of the model independently of the decision threshold.

4.3 Experimental Results

Temporal segmentation validation. To assess the quality of temporal segmentation, an expert-based validation was conducted on microinjection sequences automatically extracted by the pipeline. A total of 24 sequences were randomly selected and reviewed by a clinical expert to determine whether they corresponded to the microinjection phase.

The results show that 100% of the analyzed sequences were correctly identified as belonging to the microinjection phase, indicating that the pipeline is able to accurately isolate the relevant intervals of the ICSI procedure.

Spatial segmentation performance. The performance of spatial segmentation was evaluated on the held-out validation set (20% of the annotated images).

The oocyte segmentation model achieved a Dice score of 93.23% and an IoU of 87.78%, indicating high agreement with manual annotations. The pipette and triangular deformation region models achieved Dice scores of 87% and 86.86%, and IoU values of 77.06% and 77.05%, respectively. Although slightly lower than the oocyte model, these results reflect solid performance given the inherent difficulty of segmenting thin structures and localizing the membrane deformation area.

Feature extraction results. To evaluate the automatic extraction of dynamic variables during the ICSI procedure, an expert-based validation was conducted on the same 24 sequences used for temporal segmentation validation. For each sequence, the system automatically identified the injection frame, internal frame, aspiration frame, and the pipette angle at the aspiration frame.

The results show that 87.5% of injection frames and 95.8% of internal frames were correctly identified, indicating that the system reliably detects key procedural events. Since the estimation of dynamic variables, such as injection velocity, directly depends on the accurate detection of these events, these results support the reliability of the extracted dynamic features.

The accuracy for aspiration frame detection was 66.6%, which is attributed to the difficulty of identifying small structures within the pipette, such as the sperm cell. However, when considering a tolerance of ± 3 seconds with respect to the expert annotation, the accuracy rises to 91.6%, which indicates that the erroneous detections correspond to minor temporal deviations rather than systematic failures of the method. In contrast, the estimated angle at the aspiration frame was considered correct in 91.6% of the cases, demonstrating that the system accurately captures the associated geometric parameters.

Predictive modeling results. This set of experiments was designed as an ablation study to assess the contribution of different groups of input variables to prediction performance. Table 1 summarizes the results obtained across the evaluated tree-based models, including RF, GB, and XGBoost. Across all models, a consistent pattern is observed. The model based solely on clinical variables establishes a solid baseline, achieving an AUC of up to 72.1% with RF. The inclusion of static features leads only to marginal performance gains. In contrast, incorporating dynamic features results in systematic improvements across all evaluation metrics, particularly in terms of AUC.

The best performance was obtained using RF with Config. 3, reaching an AUC of 80%, an accuracy of 76.1%, and an F1-score of 84.3%. Similar trends were observed for GB and XGBoost, where dynamic features consistently produced the highest discriminative performance. These results indicate that the use of

dynamic features provides meaningful predictive value for fertilization outcome prediction.

Table 1. Comparison of predictive performance across different models and feature configurations

Model	Configuration	Accuracy	AUC	F1-score
RF	Config. 1	72.4%	72.1%	81.4%
RF	Config. 2	72.1%	74.4%	81%
RF	Config. 3	**76.1%**	**80%**	**84.3%**
GB	Config. 1	70.2%	66.9%	79.2%
GB	Config. 2	71.9%	71%	80%
GB	Config. 3	**75.2%**	**78.1%**	**83.2%**
XGBoost	Config. 1	71.9%	70.1%	80.3%
XGBoost	Config. 2	71.3%	72%	79.6%
XGBoost	Config. 3	**75%**	**80%**	**82.5%**

Model interpretability results. To better understand the contribution of each variable to the model predictions, SHAP analysis was performed on the best-performing model (RF, Config. 3). Results are shown in Fig. 4, where each point represents a sample, color encodes the feature value (red: high, blue: low), and horizontal position indicates the direction of the contribution (positive: towards fertilization, negative: towards non-fertilization).

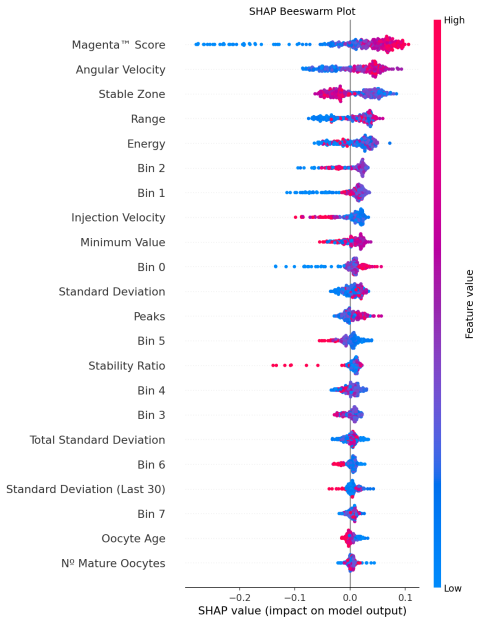

Fig. 4. SHAP beeswarm plot for the best-performing model (RF, Config. 3). Features are ranked by mean absolute SHAP value. Color indicates the feature value (red: high, blue: low).

Magenta^{TM} Score is the most influential feature overall. Low scores (blue) produce strongly negative SHAP values, indicating that a poor oocyte quality assessment is the factor that most strongly drives the prediction towards non-fertilization, consistent with previous studies [21, 22].

Among the biomechanical features, Angular Velocity and Stable Zone rank second and third respectively. Although they show opposite SHAP directions, high Angular Velocity pushes towards fertilization while high Stable Zone pushes towards non-fertilization, both variables reflect the same underlying phenomenon. When the oocyte membrane lacks mechanical resistance, the deformation angle remains obtuse and static throughout the injection: the signal is highly stable (high Stable Zone) but shows little variation (low Angular Velocity). In contrast, a membrane with active resistance produces a more dynamic signal, with greater angular fluctuations (high Angular Velocity) and shorter stable periods (low Stable Zone). Together, these two features provide key information about the biomechanical behavior of the membrane, allowing us to conclude that oocytes exhibiting low membrane resistance during microinjection are more likely to result in non-fertilization.

The histogram-based features (Bin 0-Bin 7) also appear among the most relevant descriptors, with the earlier bins showing greater influence, suggesting that the initial phase of the injection is particularly informative. Injection Velocity shows a modest but consistent negative contribution, indicating that higher injection speeds are associated with lower fertilization probability. Finally, Oocyte Age and Number of Mature Oocytes rank at the bottom of the plot, confirming that the temporal angle signal provides more discriminative information than these clinical variables alone.

5 Conclusion and future works

This work presented an automatic pipeline for analyzing ICSI procedures from microscopic videos, extracting quantitative biomechanical descriptors of the pipette–oocyte interaction to predict fertilization outcomes.

The results consistently confirmed that modeling the temporal evolution of the procedure improves predictive performance across all evaluated models, reaching an improvement of up to 7.9 percentage points in AUC with respect to using only clinical variables. Notably, point-wise features (static angle and injection velocity) yield only marginal gains, whereas dynamic descriptors derived from the temporal evolution of the angle signal provide a consistent and substantial improvement, as further supported by the SHAP analysis.

Future work will focus on extending the dataset by incorporating recordings from multiple clinics and addressing challenging cases such as occlusions, currently excluded by the quality control step, to improve generalization and allow for the study of inter-center variability. The use of time series models applied directly to the full angle signal will also be explored, as they may capture richer temporal patterns. Finally, the development of a clinical interface that allows embryologists to visualize and interact with the extracted variables in real time

could facilitate the practical adoption of the system in assisted reproduction settings.

Acknowledgements This work was supported in part by the Generalitat Valenciana under Grant CIPROM/2022/20 (COMTACTS2)

References

1. Palermo, G., Joris, H., Devroey, P., Van Steirteghem, A.C.: Pregnancies after intracytoplasmic injection of single spermatozoon into an oocyte. Lancet 340, 17–18 (1992)
2. Simopoulou, M., Gkoles, L., Bakas, P., Giannelou, P., Kalampokas, T., Pantos, K., Koutsilieris, M.: Improving ICSI: A review from the spermatozoon perspective. Syst. Biol. Reprod. Med. 62, 359–371 (2016)
3. Vander Borght, M., Wyns, C.: Fertility and infertility: Definition and epidemiology. Clin. Biochem. 62, 2–10 (2018)
4. Mizobe, Y., Oya, N., Iwakiri, R., Yoshida, N., Sato, Y., Onoue, N., Miyoshi, K., Tokunaga, M., Ezono, Y.: Developmental ability of embryos produced from oocytes with fragile oolemma by intracytoplasmic sperm injection. J. Assist. Reprod. Genet. 33, 1685–1692 (2016)
5. Nagy, Z.P., Liu, J., Joris, H., Bocken, G., Desmet, B., Van Ranst, H., Vankelecom, A., Devroey, P., Van Steirteghem, A.C.: The influence of the site of sperm deposition and mode of oolemma breakage at intracytoplasmic sperm injection on fertilization and embryo development rates. Hum. Reprod. 10, 3171–3177 (1995)
6. Palermo, G.D., Alikani, M., Bertoli, M., Colombero, L.T., Moy, F., Cohen, J., Rosenwaks, Z.: Oolemma characteristics in relation to survival and fertilization patterns of oocytes treated by intracytoplasmic sperm injection. Hum. Reprod. 11, 172–176 (1996)
7. Twinanda, A.P., Shehata, S., Mutter, D., Marescaux, J., de Mathelin, M., Padoy, N.: EndoNet: A deep architecture for recognition tasks on laparoscopic videos. IEEE Trans. Med. Imaging 36(1), 86–97 (2017)
8. Zia, A., Sharma, Y., Bettadapura, V., Sarin, E., Clements, M.A., Essa, I., Jannin, P.: Automated surgical skill assessment in RMIS training. Int. J. Comput. Assist. Radiol. Surg. 13(5), 731–739 (2018)
9. Forestier, G., Petitjean, F., Riffaud, L., Jannin, P.: Surgical workflow analysis. Int. J. Comput. Assist. Radiol. Surg. 13(5), 595–605 (2018)
10. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A.W.M., van Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. Med. Image Anal. 42, 60–88 (2017)
11. Khosravi, P., Kazemi, E., Zhan, Q., Malmsten, J.E., Toschi, M., Zisimopoulos, P., Sigaras, A., Lavery, S., Cooper, L.A., Hickman, C., Meseguer, M., Elemento, O.: Deep learning enables robust assessment and selection of human blastocysts after in vitro fertilization. npj Digit. Med. 2, 21 (2019)
12. Tran, D., Cooke, S., Illingworth, P.J., Gardner, D.K.: Deep learning as a predictive tool for fetal heart pregnancy following time-lapse incubation and blastocyst transfer. Hum. Reprod. 34(6), 1011–1018 (2019)
13. Afnan, M.A.M., Rudin, C., Conitzer, V., Savulescu, J., Mishra, A., Liu, Y., Afnan, M.: Interpretable, not black-box, artificial intelligence should be used for embryo selection. Hum. Reprod. Open 2021(4) (2021)

14. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature 521, 436–444 (2015)
15. Targosz, A., Myszor, D., Mrugacz, G.: Human oocytes image classification method based on deep neural networks. Biomed. Eng. Online 22(1) (2023)
16. Yagi, N., Tsuji, H., Morimoto, T., Maekawa, T., Mizuta, S., Ishikawa, T., Hata, Y.: Rupture prediction for microscopic oocyte images of piezo intracytoplasmic sperm injection by principal component analysis. J. Clin. Med. 11(21), 6546 (2022)
17. Letort, G., Eichmüller, A., Da Silva, C., Nikalayevich, E., Crozet, F., Salle, J., et al.: An interpretable and versatile machine learning approach for oocyte phenotyping. Preprint, HAL (2022)
18. Bori, L., Meseguer, F., Valera, M.A., Galan, A., Remohi, J., Meseguer, M.: The higher the score, the better the clinical outcome: retrospective evaluation of automatic embryo grading as a support tool for embryo selection in IVF laboratories. Hum. Reprod. 37(6), 1148–1160 (2022)
19. Bori, L., Toschi, M., Esteve, R., Delgado, A., Pellicer, A., Meseguer, M.: External validation of a fully automated evaluation tool: a retrospective analysis of 68,471 scored embryos. Fertil. Steril. 123(4), 634–643 (2025)
20. Bori, L., Johansen, M.N., Berntsen, J., Parner, E.T., Meseguer, M.: Predicting time to live birth with deep learning embryo ranking: a novel multiple imputation approach. Hum. Reprod. 40(8), 1516–1525 (2025)
21. Cimadomo, D., Badajoz, V., Hebles, M., Mifsud, L., Urda, C., Sánchez, T., Sánchez, A., Ortega, C., Ávila, J., Mariné, C., Mercuri, N., Fjeldstad, J., Krivoi, A., Nayot, D., Rienzi, L.: Artificial intelligence-based donor oocyte quality assessment moderately improves the prediction of blastocyst development. Hum. Reprod. 40(10), 1886–1892 (2025)
22. Fjeldstad, J., Qi, W., Mercuri, N., Siddique, N., Meriano, J., Krivoi, A., et al.: An artificial intelligence tool predicts blastocyst development from static images of fresh mature oocytes. Reprod. Biomed. Online 48(6) (2024)
23. Iannone, A., Carfì, A., Mastrogiovanni, F., Zaccaria, R., Manna, C.: On the role of artificial intelligence in analysing oocytes during in vitro fertilisation procedures. Artif. Intell. Med. 157 (2024)
24. Fjeldstad, J., Qi, W., Siddique, N., Mercuri, N., Nayot, D., Krivoi, A.: Segmentation of mature human oocytes provides interpretable and improved blastocyst outcome predictions by a machine learning model. Sci. Rep. 14(1) (2024)
25. Breiman, L.: Random forests. Mach. Learn. 45, 5–32 (2001)
26. Friedman, J.H.: Greedy function approximation: A gradient boosting machine. Ann. Stat. 29, 1189–1232 (2001)
27. Chen, T., Guestrin, C.: XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794. ACM (2016)
28. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778 (2016)
29. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 234–241 (2015)
30. Lundberg, S.M., Lee, S.I.: A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems (NeurIPS) (2017)
31. Dice, L.R.: Measures of the amount of ecologic association between species. Ecology 26(3), 297–302 (1945)