

Escuela Superior Politécnica del Litoral

Facultad de Ciencias de la Vida

Análisis de variables clínicas de mujeres con hiperrespuesta al estímulo ovárico
controlado mediante Machine Learning

VIDA-465

Proyecto Integrador

Previo la obtención del Título de:

Bióloga

Presentado por:

Andrea de los Ángeles Iglesias Guerrero

Guayaquil - Ecuador

Año: 2025

Dedicatoria

El presente proyecto lo dedico a mi familia, por ser el pilar constante de mi vida. A mis padres, Gina y Pedro, por todos los sacrificios visibles e invisibles que han hecho por mí, por brindarme una vida llena de amor, el apoyo y oportunidades.

A mi hermano Cristopher, la persona con quien comparto el vínculo más profundo y especial. Gracias por ser mi compañero, mi refugio y mi mayor inspiración; nuestra conexión y apoyo mutuo han marcado mi crecimiento.

A mis abuelitas, Germanía y Norma, por su amor incondicional, su compañía y ternura en cada etapa de mi crecimiento.

A Dayanna, quien ocupa un lugar eterno y especial en mi corazón. Su alegría, autenticidad y amor por la ciencia continúan guiándome e inspirándome cada día.

A Sofía, por su apoyo incondicional y su sensibilidad genuina. A Domenica, por su fortaleza y su cariño constante. Ya que han una parte esencial de quien soy hoy, al crecer conmigo y compartir años de aprendizajes y recuerdos.

A Nathalia y Patricia, por su amistad sincera y profunda; a mis amigos de la universidad, por compartir valores, curiosidad y amor por la naturaleza, y por construir una comunidad basada en la ética, la empatía y el deseo de aprender.

A Hunter, por su apoyo y por acompañarme en esta etapa, recordándome que incluso en los días difíciles el camino puede sentirse más liviano.

Finalmente, dedico este logro a Dios, por las personas que puso en mi camino, y por permitirme seguir aprendiendo, buscando siempre actuar con verdad, empatía y amor.

Agradecimientos

Mi más sincero agradecimiento a mi tutora, Fernanda Bertuccez Cordeiro, por su acompañamiento, orientación y apoyo académico a lo largo del desarrollo de este proyecto. Su guía, vocación y ejemplo fueron fundamentales para la consolidación del presente trabajo y de mi formación profesional.

A la clínica NeoVita de reproducción asistida, ubicada en São Paulo, Brasil, por proporcionar la base de datos clínica que hizo posible la realización de este proyecto y por la confianza otorgada para el uso académico de dicha información.

A mis profesores, por transmitir no solo conocimientos científicos, sino también valores, experiencias y enseñanzas que van más allá del aula. Su vocación, pasión y compromiso dejaron en mí una huella profunda y me brindaron herramientas para la vida, no solo para la profesión.

Declaración Expresa

Yo Andrea Iglesias Guerrero acuerdo y reconozco que:

La titularidad de los derechos patrimoniales de autor del proyecto de graduación corresponderá al autor o autores, sin perjuicio de lo cual la ESPOL recibe en este acto una licencia gratuita de plazo indefinido para el uso no comercial y comercial de la obra con facultad de sublicenciar, incluyendo la autorización para su divulgación, así como para la creación y uso de obras derivadas. En el caso de usos comerciales se respetará el porcentaje de participación en beneficios que corresponda a favor del autor o autores.

La titularidad total y exclusiva sobre los derechos patrimoniales de patente de invención, modelo de utilidad, diseño industrial, secreto industrial, software o información no divulgada que corresponda o pueda corresponder respecto de cualquier investigación, desarrollo tecnológico o invención realizada por mí durante el desarrollo del proyecto de graduación, pertenecerán de forma total, exclusiva e indivisible a la ESPOL, sin perjuicio del porcentaje que me corresponda de los beneficios económicos que la ESPOL reciba por la explotación de mi innovación, de ser el caso.

En los casos donde la Oficina de Transferencia de Resultados de Investigación (OTRI) de la ESPOL comunique al autor que existe una innovación potencialmente patentable sobre los resultados del proyecto de graduación, no se realizará publicación o divulgación alguna, sin la autorización expresa y previa de la ESPOL.

Guayaquil, 7 de octubre del 2020.



Andrea de los Angeles

Iglesias Guerrero

Evaluadores

Diego Arturo Gallardo Polit

Profesor de Materia

Fernanda Bertuccez Cordeiro

Tutor de proyecto

Resumen

La hiperrespuesta al estímulo ovárico controlado representa un desafío clínico relevante en fertilización in vitro, debido a la coexistencia de riesgos clínicos y a la dificultad para estimar de manera individualizada la probabilidad de implantación. El presente proyecto tiene como objetivo desarrollar y evaluar un modelo predictivo basado en aprendizaje automático explicable que permita estimar la probabilidad de implantación clínica en pacientes con hiperrespuesta ovárica, integrando variables clínicas, endocrinas y embriológicas. Se plantea como hipótesis que la combinación de modelos no lineales, interpretabilidad clínica y calibración probabilística mejora la utilidad de las predicciones para la toma de decisiones médicas.

Se realizó un estudio observacional retrospectivo utilizando datos clínicos y embriológicos anonimizados proporcionados por la Clínica NeoVita, un centro privado de reproducción asistida ubicado en São Paulo, Brasil. El análisis se realizó sobre una cohorte final de 1.479 pacientes con registros clínicos y embriológicos completos y validados. De estas, 884 correspondieron a pacientes con hiperrespuesta ovárica y 559 al grupo control. En el grupo de hiperrespuesta, 335 pacientes lograron un embarazo y 549 no lograron dicho desenlace, mientras que en el grupo control se registraron 234 casos con embarazo y 325 sin embarazo. El modelo CatBoost alcanzó un desempeño discriminativo sólido con un ROC-AUC en torno a 0.84, con adecuada confiabilidad probabilística e interpretabilidad clínica. Se concluye que el modelo desarrollado constituye una herramienta potencialmente útil para apoyar decisiones clínicas personalizadas en reproducción asistida.

Palabras clave: hiperrespuesta ovárica; implantación clínica; fertilización in vitro; medicina reproductiva personalizada; aprendizaje automático explicable.

Abstract

Hyperresponsiveness to controlled ovarian stimulation presents a significant clinical challenge in in vitro fertilization (IVF) due to the coexistence of clinical risks and the difficulty in individually estimating implantation probability. This project aims to develop and evaluate a predictive model based on explainable machine learning to estimate the clinical implantation probability in patients with ovarian hyperresponsiveness, integrating clinical, endocrine, and embryological variables. The hypothesis is that combining nonlinear models, clinical interpretability, and probabilistic calibration improves the usefulness of predictions for medical decision-making.

A retrospective observational study was conducted using anonymized clinical and embryological data provided by Clínica NeoVita, a private assisted reproduction center located in São Paulo, Brazil. The analysis was performed on a final cohort of 1,479 patients with complete and validated clinical and embryological records. Of these, 884 corresponded to patients with ovarian hyperresponsiveness and 559 to the control group. In the hyperresponsive group, 335 patients achieved pregnancy and 549 did not, while in the control group, 234 cases resulted in pregnancy and 325 did not.

The CatBoost model achieved solid discriminative performance with a ROC-AUC of approximately 0.84, with adequate probabilistic reliability and clinical interpretability. It is concluded that the developed model constitutes a potentially useful tool to support personalized clinical decisions in assisted reproduction.

Keywords: *ovarian hyperresponsiveness; clinical implantation; in vitro fertilization; personalized reproductive medicine; explainable machine learning.*

Índice general

Índice general	III
Simbología	V
Índice de figuras.....	VI
Índice de tablas	VI
Capítulo 1	1
1.1 Descripción del Problema	4
1.2 Justificación del Problema	7
1.3 Objetivos	9
1.3.1 <i>Objetivo general</i>	9
1.3.2 <i>Objetivos específicos</i>	9
1.4 Marco teórico	10
1.4.1 <i>Fisiología de la respuesta ovárica y concepto de hiperrespuesta</i>	11
1.4.2 <i>Microambiente folicular, calidad biológica y competencia embrionaria</i>	11
1.4.3 <i>Medicina personalizada y estratificación del riesgo en reproducción asistida</i>	12
1.4.4 <i>Aprendizaje automático aplicado a fertilización in vitro</i>	13
1.4.5 <i>Interpretabilidad, calibración y uso clínico de modelos predictivos</i>	13
1.4.6 <i>Síntesis teórica y justificación del enfoque adoptado</i>	14
Capítulo 2	15
2. Metodología.	16
2.1. Población de estudio.....	16
2.2. Fuentes de Datos y Estructuración de Variables	17
2.3. Procesamiento de Datos Biológicos	19
2.4 Ingeniería y selección de características.....	20
2.5. Modelado predictivo y validación	20
2.6. Validación Cruzada y Evaluación del Desempeño	21
2.7. Interpretabilidad Biológica y Fase de Despliegue (Deployment)	22

2.8 Consideraciones éticas y limitaciones metodológicas.....	23
Capítulo 3.....	25
3. Resultados y análisis	26
3.1 Desempeño del modelo predictivo (clasificación)	26
3.2 Capacidad de clasificación y matriz de confusión	27
3.3 Análisis del umbral de decisión y balance clínico entre sensibilidad y precisión.....	28
3.4 Calibración del modelo y confiabilidad clínica de las probabilidades	29
3.5 Interpretabilidad del modelo: análisis SHAP global	31
3.6 Interpretabilidad individual (SHAP waterfall)	32
3.7 Relación dosis–respuesta (ovocitos).....	33
3.8 Plataforma clínica basada en inteligencia artificial	34
Capítulo 4.....	36
4. Conclusiones y recomendaciones	37
4.1 Conclusiones	37
4.2 Recomendaciones	38
Referencias.....	40
Apéndice.....	45

Abreviaturas

AFC	Antral Follicle Count (Conteo de folículos antrales)
AMH	Hormona Antimülleriana
ASRM	American Society for Reproductive Medicine
AUC	Área Bajo la Curva
AUPRC	Área Bajo la Curva Precisión–Recall
BMI	Body Mass Index (Índice de Masa Corporal)
ESHRE	European Society of Human Reproduction and Embryology
FIV	Fertilización In Vitro
FSH	Hormona Folículo Estimulante
IMC	Índice de Masa Corporal
IVF	In Vitro Fertilization
MII	Ovocitos en Metafase II
ML	Machine Learning
OHSS	Síndrome de Hiperestimulación Ovárica
PCOS	Síndrome de Ovario Poliquístico
ROC	Receiver Operating Characteristic
SHAP	Shapley Additive Explanations
WHO	World Health Organization

Simbología

%	Porcentaje
mm	Milímetro

Índice de figuras

Figura 1	27
Figura 2	29
Figura 3	30
Figura 4	31
Figura 5	32
Figura 6	33
Figura 7	35

Índice de tablas

Tabla 1	26
---------------	----

Capítulo 1

1. Introducción

La infertilidad es actualmente reconocida como un problema de salud global, con implicaciones médicas, psicológicas y socioeconómicas que afectan de manera directa a millones de personas. De acuerdo con estimaciones recientes de la Organización Mundial de la Salud, la infertilidad afecta aproximadamente a una de cada seis personas a lo largo de su vida reproductiva, lo que ha impulsado un aumento sostenido en la demanda de tecnologías de reproducción asistida (TRA) a nivel mundial (World Health Organization, 2023). Entre estos tratamientos, la fertilización *in vitro* (FIV) se ha consolidado como una de las estrategias terapéuticas más utilizadas; aunque las recomendaciones y reportes recientes de la European Society of Human Reproduction and Embryology (ESHRE) señalan que los resultados de éxito en la fertilización *in vitro* siguen siendo muy variables y difíciles de anticipar de forma individualizada (ESHRE, 2022).

Uno de los principales determinantes del resultado clínico en la fertilización *in vitro* (FIV) es la respuesta ovárica a la estimulación hormonal controlada (La Marca & Sunkara, 2014; Griesinger et al., 2011). En este contexto, la hiperrespuesta ovárica no representa un desafío en términos de la decisión médica inmediata —la cual suele orientarse a la segmentación del ciclo o a la cancelación de la transferencia por razones de seguridad—, sino que plantea interrogantes relevantes sobre la calidad biológica de los ovocitos y embriones obtenidos.

Si bien un mayor número de ovocitos puede incrementar la disponibilidad embrionaria, diversos estudios han demostrado que, más allá de un determinado umbral, el aumento cuantitativo no se traduce en mejoras proporcionales de los desenlaces reproductivos, lo que sugiere la existencia de limitaciones biológicas intrínsecas relacionadas con la competencia ovocitaria y embrionaria (Sunkara et al., 2011; Ji et al., 2013; Steward et al., 2014). Además, la hiperrespuesta se asocia con un mayor riesgo de síndrome de

hiperestimulación ovárica (OHSS), lo que refuerza la necesidad de estrategias clínicas orientadas a la seguridad, sin que ello garantice per se una mayor probabilidad de implantación o embarazo clínico (American Society for Reproductive Medicine Practice Committee, 2016; Papanikolaou et al., 2005; Practice Committee of the ASRM, 2024).

Más allá de los riesgos clínicos inmediatos, la evidencia sugiere que la hiperrespuesta ovárica puede estar vinculada a alteraciones biológicas del microambiente folicular y de la competencia ovocitaria y embrionaria. Estudios metabolómicos y lipidómicos han mostrado que mujeres con hiperrespuesta presentan perfiles foliculares similares a los observados en el síndrome de ovario poliquístico (PCOS), sugiriendo que podría influir negativamente en la calidad embrionaria y en la probabilidad de implantación (Cordeiro et al., 2015; Cordeiro et al., 2018). Estos hallazgos refuerzan la idea, ya discutida en la práctica clínica, de que maximizar la respuesta ovárica no garantiza necesariamente mejores resultados reproductivos, y que la calidad biológica desempeña un rol principal (Cordeiro et al., 2025).

En la práctica diaria, la interpretación del pronóstico reproductivo en FIV se ve condicionada por la naturaleza multifactorial del proceso reproductivo. Factores como la edad materna, la etiología de la infertilidad, la calidad embrionaria y las características endocrinas y metabólicas interactúan de manera compleja, limitando la capacidad de los enfoques tradicionales para ofrecer estimaciones pronósticas individualizadas y biológicamente integradas (La Marca & Sunkara, 2014; Griesinger et al., 2011). En la práctica clínica, estas estimaciones pronósticas suelen apoyarse en la experiencia del médico y en la evaluación aislada de variables, lo que puede generar variabilidad en la estimación del potencial reproductivo, aun cuando las decisiones de seguridad clínica estén claramente establecidas.

Esta problemática adquiere especial relevancia en contextos como Ecuador y otros países de América Latina, donde los tratamientos de FIV implican un elevado gasto directo para las pacientes y un acceso limitado a ciclos repetidos. En estos escenarios, la

optimización de la estimación pronóstica y de la planificación terapéutica no solo tiene implicaciones médicas, sino también económicas y éticas, al influir en la eficiencia del uso de recursos y en la gestión de expectativas de las pacientes (Moreira Cedeño et al., 2025).

En la actualidad, las técnicas de inteligencia artificial y de aprendizaje automático han emergido como herramientas prometedoras para integrar múltiples variables clínicas y embriológicas como apoyo al pronóstico y planificación clínica en medicina reproductiva. Diversos estudios han demostrado el potencial de estos enfoques para predecir resultados como embarazo clínico o nacimiento vivo, aunque persisten desafíos relacionados con la interpretabilidad, la calibración probabilística y la adopción clínica de estos modelos (Jiang & Bormann, 2023; Sadegh-Zadeh et al., 2024; Wang et al., 2022). En este sentido, la incorporación de modelos explicables y clínicamente interpretables se alinea con los principios de la medicina de precisión en reproducción asistida (Gleicher et al., 2017).

En este contexto, el presente estudio propone el desarrollo y evaluación de un modelo para predicción que utiliza aprendizaje automático con el objetivo de aproximar la probabilidad de implantación clínica a nivel individual en pacientes con hiperrespuesta ovárica, integrando desempeño predictivo, interpretabilidad y calibración probabilística. El enfoque adoptado busca no solo mejorar la capacidad de predicción, sino también facilitar su implementación clínica real, contribuyendo a una estimación pronóstica más personalizada, eficiente y costo-efectiva, que sirva de apoyo a la toma de decisiones clínicas ya establecidas en el ámbito de la fertilización in vitro.

1.1 Descripción del Problema

A pesar de los avances técnicos en fertilización in vitro (FIV), la estimación pronóstica individual en pacientes con hiperrespuesta ovárica continúa realizándose con incertidumbre. En la práctica, los clínicos deben integrar múltiples variables clínicas, endocrinas y embriológicas para planificar el manejo clínico, sin disponer de herramientas

objetivas que permitan estimar de manera individualizada la probabilidad de implantación clínica. Esta limitación es especialmente relevante en pacientes hiperrespondedoras, en quienes el número elevado de ovocitos recuperados no se traduce necesariamente en mejores resultados reproductivos, y cuya interpretación clínica sigue siendo un desafío pronóstico y biológico (Sunkara et al., 2011).

En pacientes con hiperrespuesta ovárica, el impacto clínico no se limita al incremento del riesgo de síndrome de hiperestimulación ovárica (OHSS), sino que también involucra cambios biológicos a nivel del microambiente folicular, con potencial repercusión sobre la calidad ovocitaria y el desarrollo embrionario (American Society for Reproductive Medicine Practice Committee, 2016; Cordeiro et al., 2018). Evidencia derivada de estudios metabolómicos indica que el perfil del fluido folicular en mujeres hiperrespondedoras presenta características comparables a las descritas en el síndrome de ovario poliquístico (PCOS), lo que apunta a la existencia de procesos fisiopatológicos compartidos capaces de influir en la probabilidad de implantación embrionaria (Cordeiro et al., 2015; Cordeiro et al., 2018).

Desde una perspectiva clínica, la interpretación del pronóstico reproductivo en pacientes hiperrespondedoras se realiza en un contexto de incertidumbre biológica. Variables como la edad materna, la etiología de la infertilidad, la respuesta endocrina, la calidad embrionaria y las estrategias de laboratorio interactúan de manera no lineal, dificultando la estimación individualizada del pronóstico reproductivo mediante enfoques tradicionales (La Marca & Sunkara, 2014; Griesinger et al., 2011; Cordeiro et al., 2025). En consecuencia, los clínicos suelen apoyarse en la experiencia personal y en indicadores aislados, lo que puede generar variabilidad en la estratificación pronóstica y en la planificación del ciclo, aun cuando las medidas de seguridad clínica estén claramente establecidas.

Esta problemática se ve amplificada en contextos como Ecuador y otros países de América Latina, donde los tratamientos de FIV implican un alto gasto directo para las pacientes. En Ecuador, el costo estimado de un ciclo de FIV oscila entre USD 4.000 y 7.000, a lo que se suman costos adicionales frecuentes como la congelación de embriones (USD 750), la descongelación y transferencia embrionaria (USD 1.000), la biopsia o aspiración (USD 675) y la congelación de muestras (USD 200), según datos reales reportados por el Instituto Quiteño de Infertilidad (IQUI.ec, s/f). En este escenario, estimaciones pronósticas imprecisas y planificaciones subóptimas no solo impactan los resultados reproductivos, sino que también incrementan la carga económica y emocional para las pacientes.

Con el avance de las técnicas de análisis de datos en el ámbito biomédico, han surgido modelos de aprendizaje automático aplicados a la predicción de resultados reproductivos. Sin embargo, muchos de estos enfoques presentan limitaciones relevantes para su adopción clínica, como la falta de interpretabilidad, una calibración probabilística insuficiente o una escasa integración en los flujos de trabajo médicos (Jiang & Bormann, 2023; Wang et al., 2022; Sadegh-Zadeh et al., 2024). En consecuencia, persiste una brecha entre el potencial analítico y su uso efectivo para respaldar matemáticamente el criterio clínico, mejorar la estratificación pronóstica y apoyar la planificación del manejo en el contexto de hiperrespuesta.

El problema abordado en este proyecto se analiza a partir de datos clínicos reales proporcionados por la Clínica NeoVita, un centro privado de reproducción asistida ubicado en São Paulo, Brasil, especializado en tratamientos de fertilización in vitro. La información disponible corresponde a ciclos completos de reproducción asistida y permite identificar distintos fenotipos de respuesta ovárica, así como el desenlace clínico de implantación. Tras procesos de depuración y control de calidad de los registros clínicos, se dispone de una cohorte final de 1.479 pacientes con información clínica y embriológica válida, de las cuales

884 presentan hiperrespuesta ovárica y 559 corresponden a pacientes control. Esta distribución evidencia la coexistencia de grupos con respuestas ováricas diferenciadas y desenlaces clínicos heterogéneos, lo que convierte al problema en un fenómeno susceptible de observación, medición y análisis cuantitativo bajo condiciones reales de práctica médica.

1.2 Justificación del Problema

La necesidad de herramientas predictivas clínicamente útiles en fertilización in vitro responde a la complejidad biológica del proceso reproductivo y a las limitaciones de los enfoques tradicionales para integrar múltiples variables simultáneamente. En el caso específico de la hiperrespuesta ovárica, disponer de estimaciones individualizadas de la probabilidad de implantación clínica permitiría optimizar la planificación terapéutica y reducir intervenciones innecesarias, sin comprometer la seguridad de la paciente (Devroey et al., 2011; Practice Committee of the ASRM, 2024).

Dentro de la cohorte analizada de 1.479 registros, el desenlace de implantación clínica no se distribuye de manera homogénea entre los grupos de respuesta ovárica. En el grupo de hiperrespuesta conformado por 884 casos, 335 pacientes alcanzan implantación clínica mientras que 549 no lo hacen. Mientras que en el grupo control de 559 ciclos, se registran 234 casos con implantación y 325 sin implantación. Esta heterogeneidad y el desbalance entre desenlaces dificultan la estimación individualizada del pronóstico reproductivo mediante enfoques tradicionales, reforzando la necesidad de herramientas capaces de integrar múltiples variables simultáneamente y apoyar de forma objetiva la estratificación pronóstica y el asesoramiento clínico.

El uso de modelos de aprendizaje automático ofrece la posibilidad de integrar, de manera simultánea, información clínica, endocrina y embriológica. No obstante, para que estas herramientas sean adoptadas en la práctica clínica, deben cumplir requisitos clave: desempeño predictivo adecuado, interpretabilidad clínica y confiabilidad probabilística

(Steyerberg, 2010; Topol, 2019; Cordeiro et al., 2025). En este sentido, y desde una perspectiva clínica, la incorporación de métodos de interpretabilidad explicable, como SHAP, permite descomponer las predicciones del modelo en contribuciones individuales de cada variable, facilitando la comprensión del “por qué” detrás de cada estimación y aumentando la confianza del personal médico (Lundberg & Lee, 2017; Ponce-Bobadilla et al., 2024).

Asimismo, la calibración probabilística de los modelos predictivos constituye un elemento central para su utilidad clínica, ya que garantiza que las probabilidades estimadas reflejen de manera adecuada el riesgo real del desenlace. Modelos bien calibrados permiten estratificar a las pacientes según su perfil de riesgo y apoyar el asesoramiento reproductivo y la toma de decisiones compartida, superando la lógica de clasificación binaria (Niculescu-Mizil & Caruana, 2005; Hosmer et al., 2013).

Desde un punto de vista económico y de implementación, el desarrollo de una herramienta predictiva confiable adquiere especial relevancia en el contexto ecuatoriano. Considerando los costos asociados a los ciclos de FIV (hasta 7000 dólares por ciclo) y a los procedimientos complementarios, incluso pequeñas mejoras en la estratificación pronóstica y planificación del tratamiento podrían mejorar la eficiencia en la utilización de recursos clínicos, con un impacto económico favorable para pacientes y clínicas de reproducción (IQUI.ec, s/f). En conjunto, este proyecto se justifica por su potencial para mejorar simultáneamente la seguridad clínica, la eficiencia económica y la calidad de la atención en pacientes con hiperrespuesta ovárica, alineándose con los principios de la medicina reproductiva personalizada (Gleicher et al., 2017).

En este estudio se desarrolla un modelo que combina análisis de datos y teoría de reproducción asistida, orientado a apoyar la estimación de probabilidad, el counseling y la planificación clínica. Se espera la obtención de un modelo predictivo sólido, también el identificar un conjunto de variables clave con mayor impacto sobre la implantación clínica, y

la incorporación de interpretabilidad clínica mediante valores SHAP. Esto permitirá a los médicos comprender el fundamento de cada predicción individual, fortaleciendo la confianza en el modelo y favoreciendo su aplicabilidad clínica (Lundberg & Lee, 2017; Jiang & Bormann, 2023).

Finalmente, esta investigación presenta un valor dual. Por un lado, contribuye a la mejora de la seguridad clínica mediante la prevención del síndrome de hiperestimulación ovárica, al apoyar estrategias más racionales de estimulación y manejo del ciclo. Por otro, optimiza el éxito reproductivo al favorecer decisiones informadas orientadas a la implantación clínica, lo cual resulta altamente relevante tanto para pacientes hiperrespondedoras como para los centros de reproducción asistida (American Society for Reproductive Medicine Practice Committee, 2016; Devroey et al., 2011).

1.3 Objetivos

1.3.1 Objetivo general

Evaluar la relación entre variables clínicas, calidad ovocitaria y calidad embrionaria en mujeres con hiperrespuesta al estímulo ovárico controlado sometidas a tratamientos de reproducción asistida, a través de la aplicación de técnicas de aprendizaje automático explicable, con el fin de estimar la probabilidad de implantación clínica y apoyar la toma de decisiones clínicas personalizadas.

1.3.2 Objetivos específicos

1. Analizar exhaustivamente la base de datos de pacientes hiperrespondedoras mediante técnicas estadísticas y preprocesamiento de datos para identificación de variables potencialmente predictoras de la implantación clínica.
2. Desarrollar y entrenar modelos de Machine Learning (Regresión Logística, CatBoost) integrando variables disponibles hasta el momento de la estimulación ovárica

(edad, AMH, AFC) y del desarrollo embrionario temprano (calidad de embriones, número de sacos gestacionales).

3. Evaluar el rendimiento del modelo predictivo utilizando métricas avanzadas (AUC-ROC, F1-score, sensibilidad) y validación cruzada para análisis de robustez y sobreajuste.

4. Interpretar los modelos mediante técnicas de explicabilidad (como valores SHAP) para analizar qué variables tienen mayor peso en la predicción de implantación para definición de recomendaciones clínicas basadas en los hallazgos.

5. Contrastar los resultados obtenidos a partir de los modelos de aprendizaje automático con la evidencia clínica y biológica disponible, para validación de su coherencia clínica y su utilidad como instrumento de apoyo en el proceso de decisiones terapéuticas en reproducción asistida.

1.4 Marco teórico

La información mostrada a continuación surgió de una revisión dirigida de literatura científica publicada entre 2010 y 2025, orientada a integrar evidencia relevante sobre hiperrespuesta ovárica, calidad biológica ovocitaria y embrionaria, y predicción de resultados en fertilización in vitro. La búsqueda se realizó en bases de datos especializadas en biomedicina y ciencias de la vida, priorizando estudios clínicos y metodológicos con enfoque en biología reproductiva.

La revisión se organizó en torno a tres ejes: la fisiología de la respuesta ovárica y sus implicaciones clínicas; el rol del microambiente folicular y la calidad embrionaria en la implantación; y la aplicación de modelos predictivos interpretables aplicados a reproducción asistida. Para ello, se emplearon combinaciones de términos en inglés relacionados con estimulación ovárica, implantación clínica y aprendizaje automático explicable, permitiendo sustentar el enfoque interdisciplinario adoptado.

1.4.1 Fisiología de la respuesta ovárica y concepto de hiperrespuesta

La estimulación ovárica controlada constituye un componente central de los tratamientos de fertilización in vitro, cuyo objetivo es inducir el desarrollo multifolicular para maximizar la obtención de ovocitos competentes. Sin embargo, la respuesta ovárica presenta una marcada variabilidad interindividual, influenciada por factores como la edad materna, la reserva ovárica, la etiología de la infertilidad y las características endocrinas de la paciente (La Marca & Sunkara, 2014; Griesinger et al., 2011).

La hiperrespuesta ovárica se caracteriza por una producción excesiva de folículos y ovocitos tras la estimulación hormonal, y se asocia a un mayor riesgo de síndrome de hiperestimulación ovárica (OHSS), así como a potenciales alteraciones en la calidad ovocitaria y embrionaria (American Society for Reproductive Medicine Practice Committee, 2016; Practice Committee of the ASRM, 2024). Estudios poblacionales realizados por Sunkara et al. (2011), así como análisis posteriores de Steward et al. (2014) y Ji et al. (2013), han demostrado que, si bien una mayor producción ovocitos incrementa inicialmente las probabilidades de éxito; este beneficio se estabiliza a partir de un determinado umbral, sugiriendo la posterior aparición de rendimientos decrecientes.

Desde una perspectiva clínica, estos hallazgos han impulsado un cambio de paradigma desde la maximización de la respuesta ovárica hacia la individualización de la estimulación, priorizando un balance entre eficacia, seguridad y calidad biológica (Devroey et al., 2011; Griesinger et al., 2011).

1.4.2 Microambiente folicular, calidad biológica y competencia embrionaria

Más allá de la cantidad de ovocitos obtenidos, la competencia ovocitaria y embrionaria depende de la integridad del microambiente folicular y de complejos procesos metabólicos y moleculares. Investigaciones basadas en enfoques metabolómicos y lipidómicos han evidenciado que mujeres con hiperrespuesta ovárica presentan perfiles del

fluido folicular alterados, similares a los observados en síndrome de ovario poliquístico, lo que podría impactar negativamente la calidad ovocitaria y el potencial de implantación (Cordeiro et al., 2015; Cordeiro et al., 2018).

Estos datos muestran que la hiperrespuesta ovárica no puede entenderse exclusivamente desde una perspectiva cuantitativa, ya que involucra procesos biológicos más profundos que influyen en la calidad reproductiva. En este contexto, la calidad embrionaria; evaluada mediante criterios morfológicos, cinéticos o genéticos; adquiere un rol central en la predicción del éxito reproductivo (Nasiri & Eftekhari-Yazdi, 2015; Khosravi et al., 2019; Cordeiro et al., 2025).

La literatura sugiere que indicadores compuestos de calidad, que integran información sobre desarrollo embrionario, morfología y contexto biológico, pueden ofrecer una estimación más precisa del potencial reproductivo que el número bruto de ovocitos recolectados (Sunkara et al., 2011; La Marca & Sunkara, 2014, Cordeiro et al., 2025).

1.4.3 Medicina personalizada y estratificación del riesgo en reproducción asistida

La creciente complejidad en procedimientos de reproducción asistida ha impulsado la adopción de enfoques de medicina personalizada, orientados a adaptar las estrategias terapéuticas al perfil individual de cada paciente. En este marco, la estratificación del riesgo reproductivo emerge como una herramienta clave para optimizar resultados clínicos y minimizar efectos adversos (Gleicher et al., 2017).

La estratificación del riesgo en FIV implica integrar múltiples dimensiones clínicas — edad, reserva ovárica, etiología de la infertilidad, respuesta a la estimulación y calidad embrionaria— en un modelo coherente que permita anticipar el pronóstico reproductivo. Sin embargo, los enfoques tradicionales basados en reglas clínicas simples o análisis univariados resultan insuficientes para capturar la complejidad de estas interacciones (Steyerberg, 2010).

En este contexto, los modelos de predicción clínica multivariable han demostrado ser herramientas útiles para pronóstico y asesoramiento clínico siempre que cumplan criterios de validez, calibración y utilidad clínica (Hosmer et al., 2013; Steyerberg, 2010).

1.4.4 Aprendizaje automático aplicado a fertilización in vitro

El aprendizaje automático ha sido ampliamente explorado en los últimos años como una estrategia para mejorar la predicción de resultados en fertilización in vitro. Diversos estudios han demostrado que algoritmos como Random Forest, Gradient Boosting y redes neuronales pueden superar a los métodos estadísticos tradicionales en la predicción de embarazo clínico o nacimiento vivo, al capturar relaciones no lineales y efectos de interacción complejos (Cordeiro et al., 2025; Breiman, 2001; Wang et al., 2022; Sadegh-Zadeh et al., 2024).

No obstante, la adopción clínica de estos modelos ha sido limitada por factores como la opacidad de los algoritmos, la dificultad para interpretar las predicciones y la falta de calibración probabilística adecuada (Jiang & Bormann, 2023). En respuesta a estas limitaciones, ha surgido un interés creciente en el desarrollo de modelos explicables (explainable AI), capaces de ofrecer no solo predicciones precisas, sino también explicaciones comprensibles para el clínico (Luong et al., 2024).

Específicamente en el campo de la reproducción asistida, la implementación de inteligencia artificial explicable ha mostrado potencial para mejorar la selección embrionaria, la predicción de ploidía y la consulta centrada en el paciente, alineándose con los principios de transparencia y toma de decisiones compartida (Khosravi et al., 2019; Luong et al., 2024).

1.4.5 Interpretabilidad, calibración y uso clínico de modelos predictivos

Para que un modelo predictivo sea clínicamente útil, no basta con alcanzar un alto desempeño discriminativo. Es fundamental que el modelo sea interpretable y que sus

probabilidades estén correctamente calibradas, de modo que reflejen el riesgo real del desenlace (Niculescu-Mizil & Caruana, 2005; Steyerberg, 2010).

Las técnicas basadas en valores SHAP permiten descomponer las predicciones de modelos complejos en contribuciones individuales de cada variable, ofreciendo una explicación intuitiva y consistente tanto a nivel global como individual (Lundberg & Lee, 2017; Ponce-Bobadilla et al., 2024). Este enfoque ha sido recomendado como una buena práctica para facilitar la adopción clínica de modelos de aprendizaje automático en entornos médicos.

La calibración probabilística, por su parte, resulta esencial para el uso clínico de modelos predictivos, ya que decisiones basadas en probabilidades mal calibradas pueden inducir intervenciones innecesarias o generar falsas expectativas (Hosmer et al., 2013; Niculescu-Mizil & Caruana, 2005). En el contexto de la FIV, donde las decisiones tienen implicaciones clínicas, económicas y emocionales, la confiabilidad probabilística adquiere una importancia particular.

1.4.6 Síntesis teórica y justificación del enfoque adoptado

En conjunto, la evidencia teórica y empírica respalda la necesidad de enfoques integradores que prioricen la calidad biológica, la seguridad clínica y la personalización del tratamiento en pacientes con hiperrespuesta ovárica. La combinación de modelos de aprendizaje automático, interpretabilidad explicable y calibración probabilística se presenta como una estrategia coherente para abordar las limitaciones actuales en la predicción del éxito reproductivo.

Este marco sustenta el enfoque del proyecto, el cual está orientado al desarrollo de un modelo predictivo clínicamente interpretable y potencialmente implementable, capaz de apoyar la estratificación pronóstica y planificación clínica en un contexto real y económicamente relevante como el de la fertilización in vitro en Ecuador y América Latina.

Capítulo 2

2. Metodología.

Se propuso un diseño observacional, analítico y retrospectivo, basado en el procesamiento de datos clínicos y embriológicos ("Real World Data"). Los datos analizados provinieron de una base clínica anonimizada proporcionada por la Clínica de Reproducción Asistida NeoVita, correspondiente a ciclos de reproducción asistida completos y de elevada complejidad, los cuales se usaron exclusivamente con fines de investigación académica.

2.1. Población de estudio

La base de datos incluyó inicialmente 6.367 ciclos completos de fertilización in vitro. Posteriormente, se realizó una depuración progresiva del conjunto de datos, excluyendo registros con información incompleta, valores nulos o inconsistencias clínicas, lo que dio lugar a una base intermedia de 3.301 casos con datos estructurados. Finalmente, tras aplicar los criterios de inclusión definidos para el presente estudio, la cohorte final quedó conformada por 1.479 pacientes con información clínica y embriológica válida para el modelado predictivo.

De la cohorte final, 884 pacientes correspondieron al grupo de hiperrespuesta ovárica, definido por el conteo a partir de 16 ovocitos recolectados en adelante, y 559 correspondieron al grupo control que incluyó ciclos de 1 a 15 ovocitos recolectados. En el grupo de hiperrespuesta, 335 pacientes presentaron implantación clínica y 549 no alcanzaron dicho desenlace, mientras que en el grupo control se registraron 234 casos con implantación clínica y 325 sin implantación. Esta distribución evidencia un desbalance de clases inherente al fenómeno clínico estudiado, el cual fue considerado explícitamente durante el entrenamiento, la calibración probabilística y la evaluación del desempeño de los modelos.

La elección de un diseño retrospectivo responde a la necesidad de examinar conjuntos de datos clínicos, extensos y de alta dimensionalidad para identificar patrones latentes no

lineales que rigen el éxito reproductivo, una tarea para la cual los modelos computacionales son idóneos.

Desde el punto de vista metodológico, el análisis se orientó a evaluar el desempeño de un modelo predictivo diseñado para estimar la probabilidad de implantación clínica utilizando únicamente información disponible en etapas tempranas del proceso reproductivo. Con este objetivo, la población de estudio fue organizada mediante una clasificación fenotípica, definida a partir del grado de respuesta ovárica observado tras la estimulación hormonal controlada.

La población objetivo se restringió específicamente al fenotipo de hiperrespuesta ovárica, que se definió como la recuperación de más de 15 ovocitos (complejos cúmulo-ovocito) tras la aspiración folicular controlada. De acuerdo con la literatura vigente, este umbral se ha identificado como un punto de inflexión a partir del cual la relación entre el número de ovocitos recuperados y la tasa de nacido vivo deja de ser lineal. Más allá de este punto, el incremento cuantitativo puede asociarse con un mayor riesgo de síndrome de hiperestimulación ovárica (SHO) y con una posible disminución en la calidad ovocitaria promedio (Sunkara et al., 2011; La Marca & Sunkara, 2014).

2.2. Fuentes de Datos y Estructuración de Variables

Los datos clínicos y de laboratorio fueron extraídos de los expedientes electrónicos y consolidados en una matriz estructurada a nivel de caso (CASE_level). Esta matriz unifica la información pertinente, bajo el criterio de disponibilidad clínica previa a la identificación de la hiperrespuesta ovárica y al desarrollo embrionario temprano. Las variables consideradas incluyeron:

Clínico-Basal:

- **Edad Materna:** Considerada la variable biológica independiente más influyente debido a su correlación directa con la senescencia ovárica, la disfunción mitocondrial y el aumento exponencial de la tasa de aneuploidía en los ovocitos.
- **Índice de Masa Corporal (IMC):** Indicador del entorno metabólico, dado que tanto la obesidad como el bajo peso pueden alterar la farmacocinética de las gonadotropinas y la receptividad endometrial.
- **Hormona Folículo Estimulante (FSH) Basal:** Biomarcador de la sensibilidad folicular y la reserva ovárica funcional.

Asociadas a la estimulación ovárica:

- **Gonadotropinas Totales:** Dosis acumulada utilizada para la estimulación multifolicular.
- **Grosor Endometrial:** Medición ecográfica del espesor de la línea endometrial, un marcador biofísico de la proliferación y receptividad uterina.

Embriológicas y de competencia gamética:

- **Total de Ovocitos Recuperados:** Cuantificación absoluta de la respuesta ovárica.
- **Ovocitos Metafase II (MII):** Subconjunto de ovocitos que han completado la meiosis I y expulsado el primer corpúsculo polar, siendo los únicos biológicamente competentes para la fertilización.
- **Tasa de Blastocisto:** Indicador dinámico de calidad embrionaria, calculado como la proporción de cigotos que logran la activación genómica y la diferenciación celular hasta el estadio de blastocisto (Día 5/6). Esta variable actúa como un proxy de la viabilidad celular y la integridad cromosómica.

Definición del Endpoint o Desenlace (Variable Dependiente):

El desenlace primario del estudio fue la implantación clínica, definida como la confirmación ecográfica de un saco gestacional intrauterino posterior a la transferencia

embrionaria. Esto permite evaluar un evento clínicamente relevante y objetivo, previo a desenlaces más tardíos como aborto o nacimiento vivo.

Se eligió la implantación clínica como endpoint debido a su relevancia como desenlace objetivo temprano y como indicador pronóstico del éxito reproductivo, así como por su disponibilidad en la base de datos analizada. Variables como la concentración sérica de β -hCG, el número de sacos gestacionales, la evolución del embarazo o el nacimiento vivo fueron consideradas desenlaces secundarios o componentes de análisis biológicos posteriores, pero no se utilizaron como variables predictoras del modelo, con el fin de evitar sesgos temporales (data leakage).

2.3. Procesamiento de Datos Biológicos

Dado que los datos reales (RealWorld Data) suelen presentar ruido e inconsistencias, se aplicó un protocolo estricto de curación de datos para asegurar la validez interna del estudio:

- **Tratamiento de Valores Ausentes (Missing Data):** Se utilizó un enfoque de imputación estadística conservadora (SimpleImputer). Para variables biológicas continuas (ej. niveles hormonales), se imputó la mediana de la población, evitando así que valores extremos (outliers) sesgaran la distribución. Para variables categóricas, se utilizó la moda. Este método preserva la estructura de la población sin introducir datos sintéticos que no correspondan a la fisiología observada.
- **Transformación de Variables Categóricas:** Las variables cualitativas (como etiología de la infertilidad) se transformaron mediante One Hot Encoding, convirtiendo categorías biológicas en vectores numéricos binarios interpretables por los algoritmos.
- **Normalización Z-Score:** Para asegurar que todas las variables contribuyeran equitativamente al modelo independientemente de su unidad de medida (donde el

Estradiol se mide en miles de pg/mL y el Endometrio en milímetros), se aplicó una estandarización mediante StandardScaler. Esto asegura que el modelo no asigne pesos artificialmente altos a variables simplemente por tener unidades de medida mayores.

Todo el flujo de preprocesamiento se encapsuló dentro de pipelines computacionales, una práctica esencial para evitar la "fuga de datos" (data leakage) entre los conjuntos de entrenamiento y validación.

2.4 Ingeniería y selección de características

Más allá de las variables crudas, se desarrollaron nuevas características (Feature Engineering) basadas en el razonamiento biológico para capturar fenómenos fisiológicos complejos.

Se destacó el cálculo de la Tasa de Blastulacion definida como la proporción de embriones que alcanzan el estadio de blastocisto respecto a la cantidad de ovocitos fecundados. Esta normaliza la eficiencia del desarrollo embrionario, permitiendo comparar la calidad ovocitaria entre pacientes con distinto número de ovocitos recuperados.

La selección final de características no se basó únicamente en valores p estadísticos tradicionales, sino que priorizó aquellas variables con relevancia clínica probada. Esto asegura que el modelo final no solo sea matemáticamente preciso, sino biológicamente coherente e interpretable por el personal médico (Bagherzadeh-Khiabani et al., 2016).

2.5. Modelado predictivo y validación

Se implementó la metodología analítica partiendo del supuesto de que los procesos biológicos de la reproducción humana exhiben comportamientos no lineales que escapan a la estadística tradicional. Se entrenaron y evaluaron las siguientes arquitecturas:

Regresión Logística (Referencia): Utilizada como modelo de referencia estándar en bioestadística. Modela la probabilidad de embarazo bajo el supuesto de una relación lineal

entre las variables predictoras y el logaritmo de la probabilidad de embarazo, constituyendo una simplificación del proceso analizado.

CatBoost Classifier (Modelo Propuesto): Se implementó este algoritmo de Gradient Boosting basado en árboles de decisión. CatBoost fue seleccionado por su capacidad superior para manejar variables categóricas y para modelar relaciones no lineales, como la saturación en la respuesta ovárica.

A diferencia de modelos lineales tradicionales, la estructura basada en árboles permite segmentar el espacio de datos y reconocer umbrales biológicamente relevantes, como ocurre con el grosor endometrial, cuyo efecto sobre la implantación clínica resulta favorable únicamente dentro de una ventana específica (aproximadamente entre 7 y 14 mm) y desfavorable fuera de ese rango. Asimismo, el algoritmo es capaz de identificar de manera automática interacciones entre variables biológicas y clínicas, sin requerir que dichas interacciones sean definidas previamente, lo que facilita la captura de relaciones complejas propias de los procesos reproductivos.

Calibración Probabilística: En medicina, no basta con clasificar con “SI” o “NO”; es crucial estimar la probabilidad precisa del evento. Se aplicó un proceso de calibración con método sigmoid (Platt) sobre el modelo CatBoost. Esto ajusta la salida bruta del modelo para que se alinee con la frecuencia empírica de embarazos observada, garantizando que un pronóstico sea estadísticamente confiable.

2.6. Validación Cruzada y Evaluación del Desempeño

Para evitar el sobreajuste y asegurar que los resultados sean aplicables a futuras pacientes, se utilizó Validación Cruzada Estratificada (Stratified K-Fold Cross-Validation, $k=5$). Este método divide la cohorte en cinco subgrupos independientes, manteniendo la misma proporción biológica de embarazos/no embarazos en cada iteración. Las métricas de evaluación se seleccionaron por su relevancia clínica:

- **AUC-ROC:** Describe qué tan bien el modelo logra separar los casos favorables de los desfavorables a lo largo de distintos umbrales de decisión. Un $AUC > 0.80$ se considera excelente en biomedicina.
- **F1-Score:** Evalúa el equilibrio entre la precisión y la sensibilidad, penalizando los modelos que simplemente predicen la clase mayoritaria.
- **Exactitud (Accuracy):** Porcentaje global de aciertos diagnósticos.
- **Brier Score:** Mide la distancia entre la probabilidad predicha y el resultado real; esencial para evaluar la fiabilidad del asesoramiento al paciente.

2.7. Interpretabilidad Biológica y Fase de Despliegue (Deployment)

Un modelo clínico no debe operar como una “caja negra”, aun cuando los modelos avanzados de inteligencia artificial tienden naturalmente a presentar estructuras complejas y poco transparentes.

Para garantizar la transparencia y la validación biológica, se integró una capa de explicabilidad mediante SHAP (Shapley Additive Explanations). Esta técnica de teoría de juegos asigna un valor de importancia a cada variable biológica, permitiendo identificar no solo “¿qué?” factores son importantes, sino “¿cómo?” afectan al resultado.

Fase de Despliegue: Plataforma HyperOva AI

Como culminación del proyecto, el modelo validado se integró en una aplicación web de soporte a la decisión clínica (Hyper Ova AI). Esta plataforma fue desarrollada utilizando tecnologías web estándar (HTML5, JavaScript) para ejecutar la lógica del modelo CatBoost en tiempo real.

El diseño conceptual de esta fase se basó en flujos de trabajo clínicos reales observados en la práctica asistencial del centro que proporcionó los datos (Clínica NeoVita), con el objetivo de maximizar la aplicabilidad del modelo en escenarios clínicos cotidianos.

La herramienta incluye:

- **Simulador Dinámico:** Permite ajustar variables clínicas y observar el cambio en la probabilidad.
- **Módulo de Explicabilidad:** Muestra gráficos de contribución factorial personalizados por paciente.
- **Integración de IA Generativa:** Conexión vía API con modelos de lenguaje (LLM) para analizar expedientes clínicos no estructurados (PDF/Texto) y fusionarlos con la predicción numérica, generando reportes integrales.
- **Aprendizaje Activo:** Un sistema de retroalimentación (Feedback Loop) que permite al personal médico ingresar el desenlace real del tratamiento, creando un repositorio de datos para el reentrenamiento y mejora continua del algoritmo.

2.8 Consideraciones éticas y limitaciones metodológicas

El estudio respetó el marco ético internacional propuesto por la Declaración de Helsinki para investigación en seres humanos, sobre confidencialidad y anonimización de datos clínicos (World Medical Association, 2025). La base de datos proporcionada por la Clínica NeoVita está en formato completamente anonimizado, sin identificadores personales, y su uso se limitó estrictamente a los objetivos académicos del presente estudio. Se garantizó la confidencialidad absoluta mediante la anonimización irreversible de los datos clínicos previo al análisis. Al tratarse de un estudio retrospectivo sin intervención directa, no se requirió consentimiento informado individual, de acuerdo con normativas internacionales para investigación secundaria con datos biomédicos.

Entre las limitaciones se incluyen el carácter retrospectivo del estudio, la dependencia de la calidad del registro clínico y la ausencia de validación externa en una cohorte geográfica distinta, los cuales son aspectos ampliamente reconocidos en estudios de ML clínico. (Steyerberg et al., 2010). No obstante, la robustez de la validación cruzada interna mitiga significativamente el riesgo de sobreajuste.

Capítulo 3

3. Resultados y análisis

3.1 Desempeño del modelo predictivo (clasificación)

En el conjunto de prueba, el modelo CatBoost (X_pre) alcanzó un desempeño discriminativo alto, con un ROC–AUC de 0.838 y un AUPRC de 0.796, acompañado de un Brier score de 0.153, lo que sugiere un balance adecuado entre capacidad discriminativa y calidad probabilística. Con un umbral de decisión de 0.5, el modelo obtuvo un Accuracy de 0.766, Precision de 0.712 y Recall de 0.599, con una especificidad de 0.862, evidenciando un comportamiento relativamente conservador para la emisión de predicciones positivas y una baja tasa de falsos positivos.

Tabla 1

Desempeño del modelo CatBoost en el conjunto de prueba

Modelo	ROC–AUC	AUPRC	Brier	Accuracy	Precision	Sensibilidad	Especificidad	F1 score
CatBoost (X_pre)	0.838	0.796	0.153	0.766	0.712	0.599	0.862	0.651
CatBoost + Platt	0.838	0.795	0.182	0.757	0.659	0.688	0.796	0.673

Nota. Comparación del desempeño del modelo sin y con calibración probabilística. La calibración permite ajustar el balance clínico entre sensibilidad y especificidad sin comprometer la capacidad discriminativa.

Si bien el desempeño discriminativo observado de ROC–AUC cercano a 0.84 es consistente con modelos predictivos considerados clínicamente útiles, este valor debe interpretarse en el contexto de la elevada heterogeneidad biológica inherente a la reproducción humana. Estudios metodológicos han señalado que, en escenarios clínicos complejos, incrementos marginales en AUC pueden representar mejoras relevantes en la

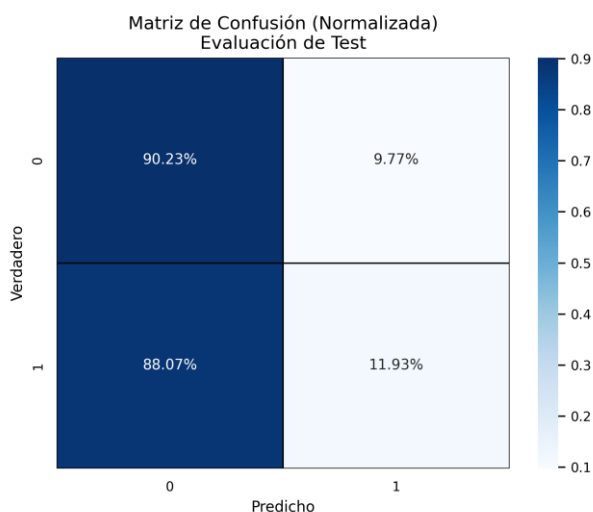
toma de decisiones, sin implicar una capacidad predictiva perfecta ni determinística (Steyerberg, 2010; Sadegh-Zadeh et al., 2024). Por tanto, el modelo propuesto debe considerarse una herramienta complementaria a la evaluación y al juicio clínico y no como un predictor absoluto del desenlace reproductivo.

Al aplicar calibración probabilística mediante Platt scaling, la capacidad discriminativa global se mantuvo estable (ROC–AUC = 0.838); sin embargo, se observó un cambio clínicamente relevante en el punto de operación del modelo. La sensibilidad aumentó a 0.688 y el F1-score a 0.673, a costa de una disminución moderada de la especificidad (0.796). Visto de una manera más práctica, esta versión calibrada resulta más adecuada para escenarios clínicos en los que se prioriza no perder ciclos con potencial reproductivo (estrategias de screening y optimización temprana), aceptando un mayor número de alertas positivas.

3.2 Capacidad de clasificación y matriz de confusión

Figura 1

Matriz de confusión normalizada del modelo CatBoost en el conjunto de prueba (test)



Nota. El modelo muestra una elevada tasa de verdaderos negativos y una sensibilidad moderada para los casos positivos, lo que indica un enfoque conservador clínicamente útil para descartar ciclos con baja probabilidad de implantación.

La matriz de confusión normalizada, generada para el conjunto de prueba (Figura 1) evidencia un desempeño asimétrico del modelo. Se observa una proporción elevada de verdaderos negativos cercana a 90%, lo cual indica un buen desempeño en la identificación de los ciclos sin embarazo. En contraste, la proporción de verdaderos positivos es menor, reflejando una sensibilidad limitada para detectar ciclos exitosos.

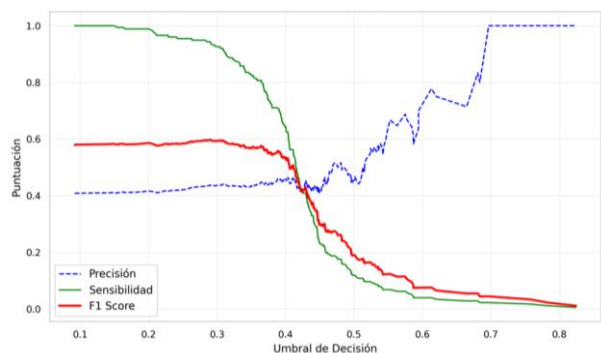
Este patrón es coherente con el desbalance de clases característico de los datos clínicos de FIV y con la naturaleza multifactorial de la implantación embrionaria. Desde una perspectiva clínica, la alta especificidad del modelo adquiere especial relevancia, ya que permite identificar con mayor confianza aquellos ciclos con muy baja probabilidad de éxito, apoyando decisiones como el ajuste de protocolos, la postergación de transferencias o un asesoramiento más realista.

La sensibilidad limitada no debe interpretarse como un fallo metodológico, sino como una consecuencia directa de la variabilidad biológica del proceso reproductivo, donde incluso embriones morfológicamente óptimos pueden no implantar.

3.3 Análisis del umbral de decisión y balance clínico entre sensibilidad y precisión

Figura 2

Optimización del punto de corte del modelo CatBoost (Ajuste del umbral)



Nota. Curvas de precisión, sensibilidad y F1-score en función del umbral de decisión. El máximo del F1-score se observa en un rango intermedio, indicando el balance óptimo entre detección de casos positivos y control de falsos positivos.

El comportamiento del modelo frente a distintos umbrales de decisión se analizó evaluando precisión, sensibilidad y F1-score, simultáneamente (Figura 2). Este análisis es relevante en un contexto clínico, donde el impacto de los falsos positivos y falsos negativos no es equivalente.

A umbrales bajos, el modelo alcanza una sensibilidad elevada, identificando la mayoría de los ciclos con implantación clínica, aunque con una precisión reducida y un aumento de falsos positivos. A medida que el umbral aumenta, la precisión mejora progresivamente, mientras que la sensibilidad disminuye de forma marcada. El F1-score alcanza su máximo en un rango intermedio alrededor de 0.40 a 0.45, lo que representa el mejor compromiso entre ambos tipos de error.

Este resultado confirma que el modelo no debe interpretarse bajo un único punto de corte universal. Su utilidad clínica depende del escenario de aplicación y del objetivo médico específico, reforzando su papel como herramienta de apoyo flexible y adaptable.

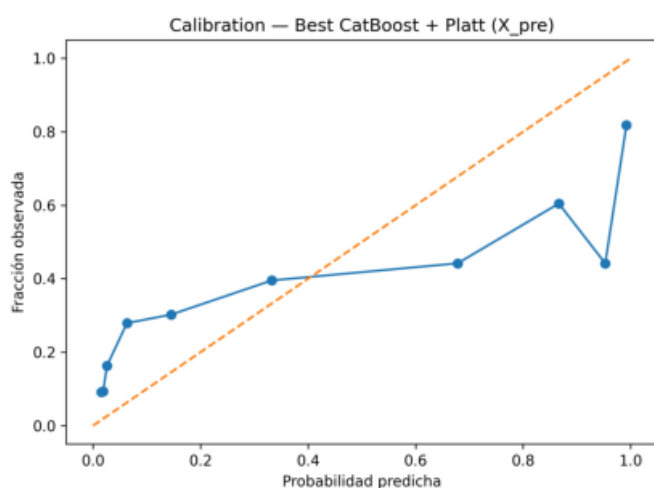
3.4 Calibración del modelo y confiabilidad clínica de las probabilidades

Además de la capacidad discriminativa, se evaluó la calibración del modelo CatBoost mediante Platt scaling, con el objetivo de garantizar que las probabilidades predichas reflejen

adecuadamente el riesgo real de implantación clínica. La calibración es un aspecto crítico para la aplicación clínica de modelos predictivos, ya que decisiones médicas basadas en probabilidades mal calibradas pueden inducir intervenciones innecesarias o generar falsas expectativas en pacientes.

Figura 3

Curva de calibración del modelo CatBoost mediante Platt scaling



Nota. Una buena alineación con la diagonal indica una adecuada confiabilidad probabilística.

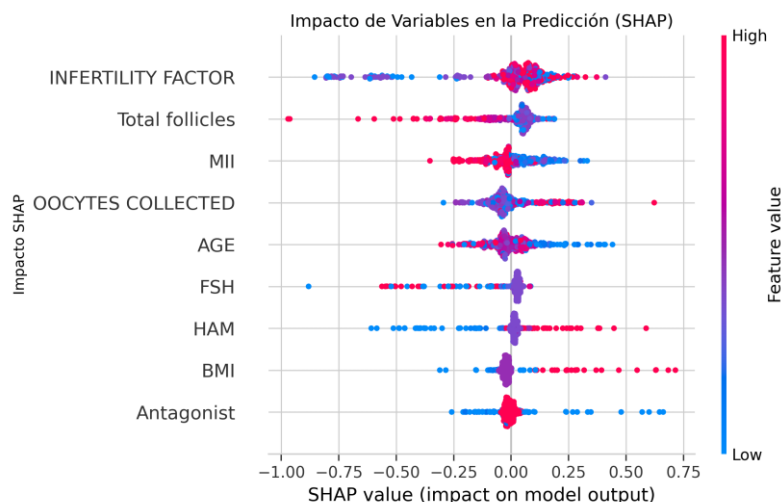
La curva de calibración obtenida (Figura 3) compara las probabilidades predichas por el modelo con la frecuencia observada de implantación clínica. Se muestra una buena alineación con la diagonal lo que indica una adecuada confiabilidad probabilística, especialmente en el rango intermedio de riesgo, que es donde se concentra la mayor parte de los casos clínicos. Esta concordancia indica que el modelo no solo discrimina correctamente entre ciclos con y sin implantación, sino que también proporciona estimaciones probabilísticas clínicamente interpretables.

Esta propiedad permite utilizar el modelo como una herramienta de estratificación de riesgo durante el asesoramiento reproductivo, facilitando la comunicación médico-paciente y apoyando decisiones personalizadas sobre la continuidad del tratamiento, el número de embriones a transferir o la necesidad de estrategias alternativas.

3.5 Interpretabilidad del modelo: análisis SHAP global

Figura 4

Gráfico SHAP tipo beeswarm del modelo CatBoost



Nota. Se muestra la dirección e intensidad del impacto de las principales variables clínicas y embriológicas sobre la predicción del modelo. Los colores representan el valor relativo de cada variable y su efecto positivo/negativo en la probabilidad de implantación.

El análisis de interpretabilidad mediante valores SHAP permitió descomponer la predicción del modelo en contribuciones individuales de cada variable clínica y embriológica. El gráfico tipo beeswarm (Figura 4) muestra que el factor de infertilidad, el total de ovocitos que se recolectados, los ovocitos en MII (metafase II) y la edad materna se encuentran entre las variables con mayor impacto global sobre la predicción.

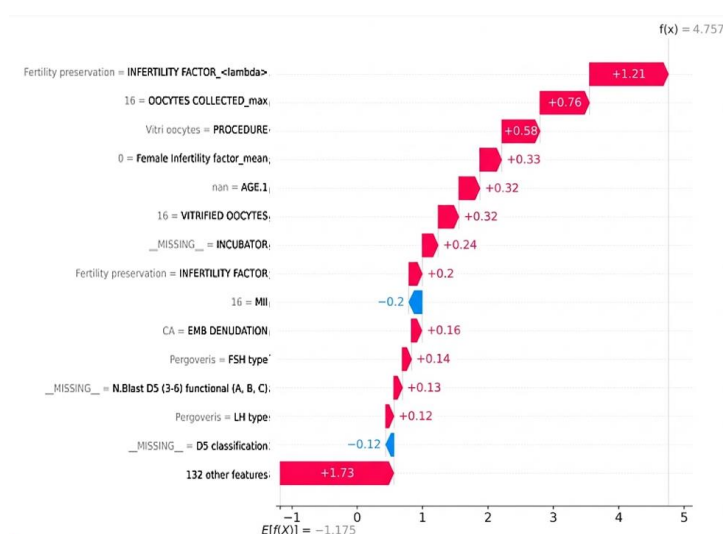
El predominio del factor de infertilidad sugiere que la etiología subyacente continúa siendo un determinante clave del éxito reproductivo, incluso en pacientes con hiperrespuesta ovárica, lo que cuestiona la noción de que una elevada respuesta ovárica pueda compensar completamente factores etiológicos adversos. Las variables endocrinas (FSH, AMH) y metabólicas (IMC) mostraron contribuciones menores pero consistentes, actuando como moduladores del pronóstico reproductivo.

Es importante señalar que los valores SHAP reflejan asociaciones aprendidas por el modelo a partir de los datos observacionales y no implican relaciones causales directas. Su valor radica en la capacidad de transparentar el razonamiento del algoritmo y facilitar la coherencia clínica de las predicciones, más que en establecer mecanismos fisiopatológicos definitivos. Esta distinción es fundamental para evitar interpretaciones deterministas y se alinea con las recomendaciones actuales para el uso responsable de modelos explicables en medicina (Lundberg & Lee, 2017; Ponce-Bobadilla et al., 2024).

3.6 Interpretabilidad individual (SHAP waterfall)

Figura 5

Interpretabilidad del modelo a nivel paciente mediante SHAP waterfall (ejemplo)



Nota. Cada barra indica la contribución de una variable clínica o embriológica al desplazamiento de la predicción final, permitiendo comprender de forma intuitiva el nivel de contribución de los factores que influyen en la predicción de la probabilidad de loimplantación clínica.

Se analizo la interpretabilidad del modelo a nivel de caso individual usando gráficos SHAP tipo *waterfall* (Figura 5). Este enfoque permite descomponer una predicción puntual

en la suma de efectos positivos y negativos de las variables clínicas y embriológicas, ofreciendo una trazabilidad clara del razonamiento del modelo.

En los casos de alto riesgo, el gráfico revela cómo combinaciones desfavorables de variables (como edad materna elevada, determinados factores de infertilidad o perfiles subóptimos de respuesta ovárica) empujan la predicción hacia una menor probabilidad de implantación. De forma complementaria, en casos de bajo riesgo, se observa cómo variables favorables compensan parcialmente otros factores, incrementando la probabilidad estimada.

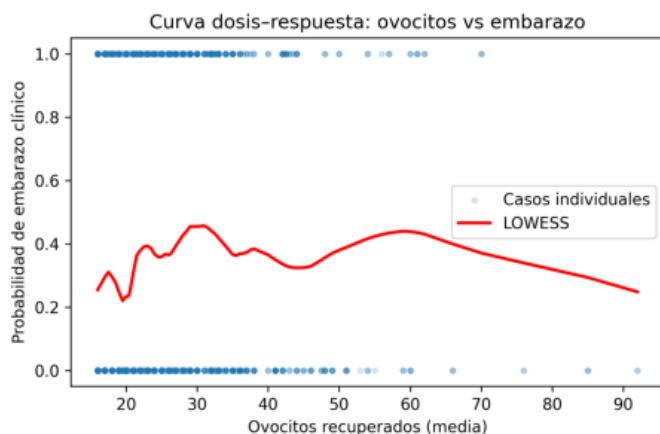
Este nivel de explicación es particularmente relevante para la adopción clínica del modelo, ya que permite al médico responder a una pregunta clave: “¿Por qué el modelo predice este resultado para esta paciente específica?”. De esta manera, el modelo se transforma en una herramienta transparente y comunicable.

3.7 Relación dosis–respuesta (ovocitos)

Las figuras de dosis–respuesta muestran que el efecto de los ovocitos recolectados sobre la probabilidad de embarazo sigue una relación no lineal, con un incremento inicial seguido de una posible meseta y rendimientos decrecientes. Este hallazgo es clínicamente importante, ya que refuerza la idea de que maximizar el número de ovocitos no garantiza un mayor éxito reproductivo.

Figura 6

Curva dosis–respuesta entre ovocitos y probabilidad de implantación.



Nota: Comparación entre la cantidad promedio de ovocitos recuperados por ciclo y la probabilidad de implantación clínica, según las estimaciones del modelo y los resultados reales.

La curva sugiere un rango de mayor utilidad clínica y una meseta a partir de la cual el beneficio marginal disminuye. No obstante, la presencia de una meseta en la relación dosis–respuesta no debe interpretarse como la existencia de un umbral óptimo universal aplicable a todas las pacientes.

La literatura sugiere que el rango de respuesta ovárica asociado a mejores resultados puede variar según la edad, la etiología de la infertilidad y el contexto metabólico de cada paciente (Sunkara et al., 2011; Ji et al., 2013; La Marca & Sunkara, 2014). En este sentido, el valor del modelo radica en su capacidad para contextualizar la respuesta ovárica dentro de un perfil individualizado, más que en proponer puntos de corte rígidos.

Este comportamiento es coherente con evidencia previa que indica que pacientes con hiperrespuesta pueden presentar alteraciones del microambiente folicular y perfiles metabólicos similares a los observados en PCOS, afectando la competencia embrionaria y la implantación (Cordeiro et al., 2018). En este contexto, la identificación de un rango óptimo de respuesta ovárica adquiere un valor práctico significativo.

3.8 Plataforma clínica basada en inteligencia artificial

Como resultado aplicado del modelo, se desarrolló una interfaz clínica interactiva orientada exclusivamente al uso médico. Esta aplicación permite ingresar dinámicamente los datos de la paciente (edad, número de ovocitos, variables embriológicas, factores de infertilidad, condiciones asociadas como PCOS o endometriosis) y observar en tiempo real cómo estos valores modifican la probabilidad de implantación clínica.

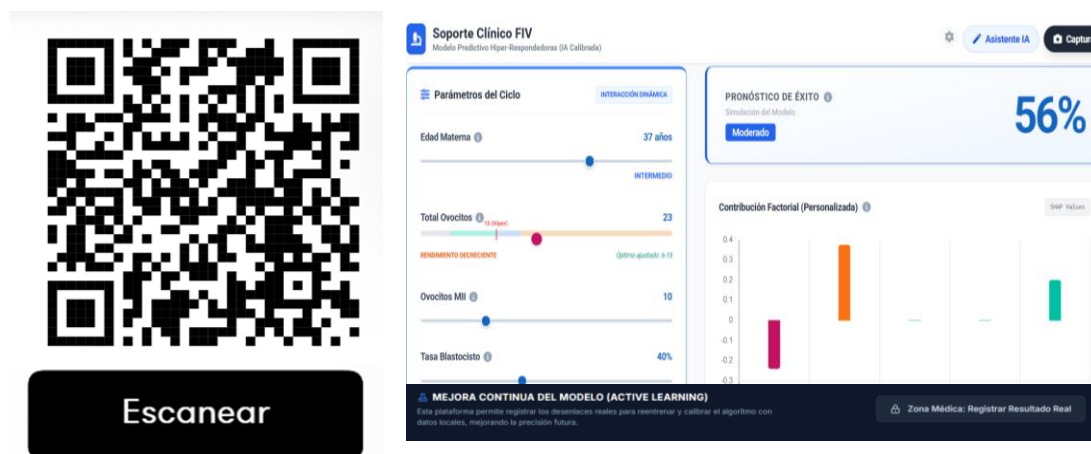
La interfaz incluye advertencias automáticas de riesgo, visualizaciones personalizadas del impacto de cada variable sobre la predicción, y un módulo de inteligencia artificial

generativa que produce reportes clínicos interpretativos con recomendaciones y alertas. Un sistema de acceso restringido mediante clave institucional permite registrar el desenlace real del tratamiento, alimentando continuamente el modelo y habilitando un esquema de retroalimentación y mejora progresiva.

La plataforma desarrollada debe considerarse, en esta etapa, un prototipo funcional orientado a la exploración de la utilidad clínica del modelo y a la facilitación de su interacción con el personal médico. Su implementación no sustituye los procesos de validación prospectiva ni los estudios de impacto clínico requeridos antes de su adopción rutinaria. Este enfoque es coherente con las recomendaciones actuales para la traslación responsable de herramientas de inteligencia artificial desde el ámbito académico hacia la práctica clínica (Jiang & Bormann, 2023; Topol, 2019).

Figura 7

Código QR para la visualización interactiva del modelo (prototipo)



Capítulo 4

4. Conclusiones y recomendaciones

4.1 Conclusiones

En relación con el primer objetivo específico, el análisis integrativo de variables clínicas, endocrinas y embriológicas permitió evidenciar que la hiperrespuesta ovárica constituye un fenotipo biológicamente heterogéneo, en el cual el incremento en el número de ovocitos recuperados no se asocia de manera lineal con la probabilidad de implantación clínica. Este hallazgo refuerza la noción de que la respuesta ovárica elevada no puede interpretarse únicamente desde una perspectiva cuantitativa, sino que refleja la interacción compleja entre calidad ovocitaria, ambiente endocrino y competencia embrionaria.

Respecto al segundo objetivo específico, la aplicación de modelos de aprendizaje automático permitió identificar patrones no lineales entre variables biológicas relevantes y el desenlace reproductivo, destacando la contribución consistente de factores como la edad materna, la etiología de la infertilidad y los parámetros de respuesta ovárica. La relevancia de estas variables, ampliamente descritas en biología reproductiva, respalda la coherencia biológica de los patrones identificados y sugiere que los modelos capturan señales reproductivas con sustento fisiológico.

En relación con el tercer objetivo específico, la evaluación del desempeño predictivo y la calibración probabilística del modelo evidenció que las estimaciones de riesgo obtenidas son consistentes con la frecuencia real de implantación observada. Desde una perspectiva biológica, esta capacidad resulta fundamental para interpretar probabilidades reproductivas de manera realista, evitando sobreestimaciones que podrían derivarse de una interpretación simplificada del número de ovocitos obtenidos.

Adicionalmente, el uso de métodos de interpretabilidad permitió vincular las predicciones del modelo con mecanismos reproductivos conocidos, lo que facilita la traducción de resultados computacionales a un lenguaje comprensible para el análisis

biológico y clínico. Este aspecto resulta clave para garantizar que las herramientas desarrolladas no operen como sistemas opacos, sino como extensiones analíticas del razonamiento biológico en medicina reproductiva.

Finalmente, en relación con el objetivo general, el proyecto demuestra que la integración de herramientas de aprendizaje automático con un enfoque biológico permite abordar de manera más precisa la complejidad de la hiperrespuesta ovárica. Los resultados obtenidos sugieren que este enfoque interdisciplinario contribuye a una comprensión más profunda del equilibrio entre cantidad y calidad ovocitaria, favoreciendo decisiones terapéuticas más informadas y alineadas con los principios de la biología reproductiva.

4.2 Recomendaciones

Arme un párrafo base y breve que indique que tras culminar lo planificado en la propuesta se obtienen las siguientes recomendaciones primordiales.

- Se recomienda el uso del modelo como una herramienta de apoyo al juicio clínico, y no como un sustituto de la decisión médica. Especialmente pertinente en etapas tempranas del tratamiento, durante el asesoramiento con la paciente y en la planificación de estrategias terapéuticas, donde la estimación del éxito puede ayudar a optimizar recursos y reducir intervenciones innecesarias.
- Realizar validaciones externas del modelo en poblaciones independientes y en otros centros clínicos, a fin de determinar su aplicabilidad más allá del conjunto de datos original y reducir posibles sesgos asociados al carácter retrospectivo del conjunto de datos utilizado.
- Se recomienda la personalización del umbral de decisión en función del escenario clínico. En contextos donde el objetivo principal sea no perder ciclos con potencial reproductivo, puede priorizarse una mayor sensibilidad, mientras que en otros escenarios resulta más adecuado un equilibrio entre precisión y sensibilidad.

- Extender el enfoque metodológico desarrollado en este estudio a otros fenotipos reproductivos complejos, como mujeres con síndrome de ovario poliquístico (PCOS) o endometriosis. El uso de bases de datos similares y arquitecturas de modelado comparables permitiría desarrollar nuevas herramientas de apoyo a la decisión clínica específicas para estas condiciones, manteniendo la interpretabilidad como eje central del análisis.
- El modelo podría ofrecerse bajo esquemas de suscripción anual, licencias institucionales o pago por uso. Un costo anual estimado entre USD 500 y 1.500 por clínica resulta marginal frente al costo de uno o dos ciclos de FIV, pero con un impacto significativo en eficiencia y calidad asistencial. Este enfoque facilita su adopción por parte de centros de reproducción asistida y refuerza su viabilidad como herramienta de innovación clínica.
- Se recomienda explorar la integración del modelo en entornos clínicos simulados, con el objetivo de evaluar su utilidad práctica, su aceptación por parte de profesionales de la salud y su potencial impacto en la optimización de estrategias terapéuticas.

Referencias

- World Health Organization. (2023). Infertility prevalence estimates 1990–2021.
<https://www.who.int>
- European Society of Human Reproduction and Embryology. (2022). ART fact sheet.
<https://www.eshre.eu>
- La Marca, A., & Sunkara, S. K. (2014). Individualization of controlled ovarian stimulation in IVF using ovarian reserve markers: from theory to practice. *Human Reproduction Update*, 20(1), 124–140. <https://doi.org/10.1093/humupd/dmt037>
- Griesinger, G., Verweij, P. J., & Gates, D. (2011). Individualized ovarian stimulation in IVF/ICSI. *Human Reproduction Update*, 17(3), 321–335.
<https://doi.org/10.1093/humupd/dmq052>
- Sunkara, S. K., Rittenberg, V., Raine-Fenning, N., Bhattacharya, S., Zamora, J., & Coomarasamy, A. (2011). Association between the number of eggs and live birth in IVF treatment: an analysis of 400, 135 treatment cycles. *Human Reproduction* (Oxford, England), 26(7), 1768–1774. <https://doi.org/10.1093/humrep/der106>
- Ji, J., Liu, Y., Tong, X., Luo, L., Ma, J., & Zhang, Y. (2013). The optimum number of oocytes retrieved in IVF treatment: An analysis of 2455 cycles. *Human Reproduction*, 28(10), 2728–2734. <https://doi.org/10.1093/humrep/det315>
- Steward, R. G., Lan, L., Shah, A. A., Yeh, J. S., Price, T. M., Goldfarb, J. M., & Muasher, S. J. (2014). Oocyte number as a predictor for ovarian hyperstimulation syndrome and live birth: An analysis of 256,381 IVF cycles. *Fertility and Sterility*, 101(4), 967–973.
<https://doi.org/10.1016/j.fertnstert.2013.12.026>

American Society for Reproductive Medicine Practice Committee. (2016). Prevention and treatment of moderate and severe ovarian hyperstimulation syndrome: A guideline. *Fertility and Sterility*, 106(7), 1634–1647.

<https://doi.org/10.1016/j.fertnstert.2016.08.048>

Papanikolaou, E. G., Tournaye, H., Verpoest, W., Camus, M., Vernaeye, V., Van Steirteghem, A., & Devroey, P. (2005). Early and late ovarian hyperstimulation syndrome: early pregnancy outcome and profile. *Human Reproduction* (Oxford, England), 20(3), 636–641. <https://doi.org/10.1093/humrep/deh638>

Practice Committee of the American Society for Reproductive Medicine. Electronic address: asrm@asrm.org. (2024). Prevention of moderate and severe ovarian hyperstimulation syndrome: a guideline. *Fertility and Sterility*, 121(2), 230–245.

<https://doi.org/10.1016/j.fertnstert.2023.11.013>

Cordeiro, F. B., Cataldi, T. R., do Vale Teixeira da Costa, L., de Lima, C. B., Stevanato, J., Zylbersztejn, D. S., Ferreira, C. R., Eberlin, M. N., Cedenho, A. P., & Turco, E. G. L. (2015). Follicular fluid lipid fingerprinting from women with PCOS and hyper response during IVF treatment. *Journal of Assisted Reproduction and Genetics*, 32(1), 45–54. <https://doi.org/10.1007/s10815-014-0375-0>

Cordeiro, F. B., Cataldi, T. R., de Souza, B. Z., Rochetti, R. C., Fraietta, R., Labate, C. A., & Lo Turco, E. G. (2018). Hyper response to ovarian stimulation affects the follicular fluid metabolomic profile of women undergoing IVF similarly to polycystic ovary syndrome. *Metabolomics*, 14(4), 51. <https://doi.org/10.1007/s11306-018-1350-z>

Cordeiro, F., Constantine-Macías, A., Loayza, F. R., Tapia-Rosero, A., Valarezo-Añazco, E., Turco, E. L., Ferreira, F. P., & Jarrín, E. P. (2025). Embryo quality and in vitro

fertilization outcome prediction using machine learning and explainability. 2025 Eleventh International Conference on eDemocracy & eGovernment (ICEDEG), 239–247.

Moreira Cedeño, D. M., Zambrano Solórzano, D. J., & Cruz Marte, I. A. (2025). Métodos de reproducción asistida en Ecuador: análisis jurídico comparado. *Revista Científica de Educación Superior y Gobernanza Interuniversitaria* Aula 24, 53.
<https://doi.org/10.56124/aula24.v1i2-25.005>

Jiang, V. S., & Bormann, C. L. (2023). Artificial intelligence in the in vitro fertilization laboratory: A review of advancements over the last decade. *Fertility and Sterility*, 120(1), 17–23. <https://doi.org/10.1016/j.fertnstert.2023.05.004>

Sadegh-Zadeh, S., Koupaei, R., & Azadi, H. (2024). Machine learning approaches for predicting live birth outcomes in IVF. *Frontiers in Artificial Intelligence*, 6, 1392611.
<https://doi.org/10.3389/frai.2024.1392611>

Wang, C.-W., Kuo, C.-Y., Chen, C.-H., Hsieh, Y.-H., & Su, E. C.-Y. (2022). Predicting clinical pregnancy using clinical features and machine learning algorithms in in vitro fertilization. *PloS One*, 17(6), e0267554.
<https://doi.org/10.1371/journal.pone.0267554>

Gleicher, N., Kushnir, V. A., Barad, D. H., & Weghofer, A. (2017). Precision medicine in assisted reproduction. *Reproductive Biology and Endocrinology*, 15(1), 1–8.
<https://doi.org/10.1186/s12958-017-0268-0>

Instituto Quiteño de Infertilidad. (s. f.). Costos de tratamientos de fertilización in vitro. Recuperado el 23 de enero de 2026, de <https://iqui.ec/>

Devroey, P., Polyzos, N. P., & Blockeel, C. (2011). An OHSS-free clinic by segmentation of IVF treatment. *Human Reproduction*, 26(10), 2593–2597.

<https://doi.org/10.1093/humrep/der251>

Steyerberg, E. W. (2010). Clinical prediction models: A practical approach to development, validation and updating by STEYERBERG, E. w. *Biometrics*, 66(2), 661–662.

<https://doi.org/10.1111/j.1541-0420.2010.01431.x>

Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44–56. [https://doi.org/10.1038/s41591-018-](https://doi.org/10.1038/s41591-018-0300-7)

0300-7

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions.

Neural Information Processing Systems, 30, 4768–4777.

<https://doi.org/10.48550/ARXIV.1705.07874>

Ponce-Bobadilla, A. V., Schmitt, V., Maier, C. S., Mensing, S., & Stodtmann, S. (2024).

Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development. *Clinical and Translational Science*, 17(11), e70056.

<https://doi.org/10.1111/cts.70056>

Niculescu-Mizil, A., & Caruana, R. (2005). Predicting good probabilities with supervised learning. In *Proceedings of the 22nd International Conference on Machine Learning (ICML '05)* (pp. 625–632). Association for Computing Machinery.

<https://doi.org/10.1145/1102351.1102430>

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). Wiley. <https://doi.org/10.1002/9781118548387>

- Nasiri, N., & Eftekhari-Yazdi, P. (2015). An overview of the available methods for morphological scoring of pre-implantation embryos in in vitro fertilization. *Cell Journal*, 16(4), 392–405. <https://doi.org/10.22074/cellj.2015.486>
- Khosravi, P., Kazemi, E., Zhan, Q. et al. Deep learning enables robust assessment and selection of human blastocysts after in vitro fertilization. *npj Digit. Med.* 2, 21 (2019). <https://doi.org/10.1038/s41746-019-0096-y>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Luong, T.-M.-T., Ho, N.-T., Hwu, Y.-M., Lin, S.-Y., Ho, J. Y.-P., Wang, R.-S., Lee, Y.-X., Tan, S.-J., Lee, Y.-R., Huang, Y.-L., Hsu, Y.-C., Le, N.-Q.-K., & Tzeng, C.-R. (2024). Beyond black-box models: explainable AI for embryo ploidy prediction and patient-centric consultation. *Journal of Assisted Reproduction and Genetics*, 41(9), 2349–2358. <https://doi.org/10.1007/s10815-024-03178-7>
- Bagherzadeh-Khiabani, F., Ramezankhani, A., Azizi, F., Hadaegh, F., Steyerberg, E. W., & Khalili, D. (2016). A tutorial on variable selection for clinical prediction models: Feature selection methods in data mining could improve the results. *Journal of Clinical Epidemiology*, 71, 76–85. <https://doi.org/10.1016/j.jclinepi.2015.10.002>
- World Medical Association. World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Participants. *JAMA*. 2025;333(1):71–74. doi:10.1001/jama.2024.21972

Apéndice

Anexo 1. Link de la clínica NeoVita que proporciono la base de datos.

<https://www.neovita.med.br/human-reproduction-clinic-in-brazil>

Figura 8

Gráfico de dependencia SHAP que muestra el efecto no lineal del porcentaje de blastocistos biopsiados sobre la predicción de implantación clínica, evidenciando cambios en la contribución del predictor a lo largo de su rango.

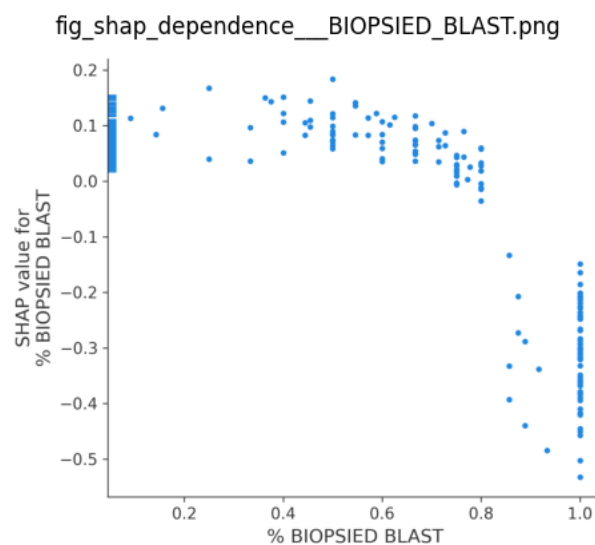


Figura 9

Relación entre el número medio de ovocitos recolectados y su contribución al modelo, mostrando una tendencia decreciente del impacto predictivo a medida que aumenta la cantidad de ovocitos.

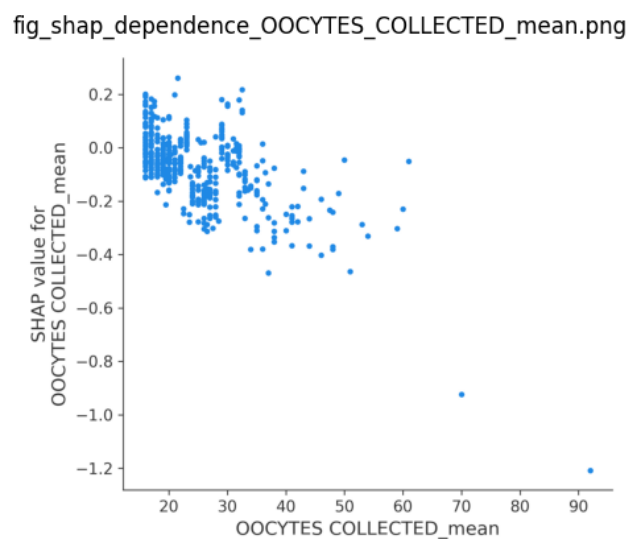
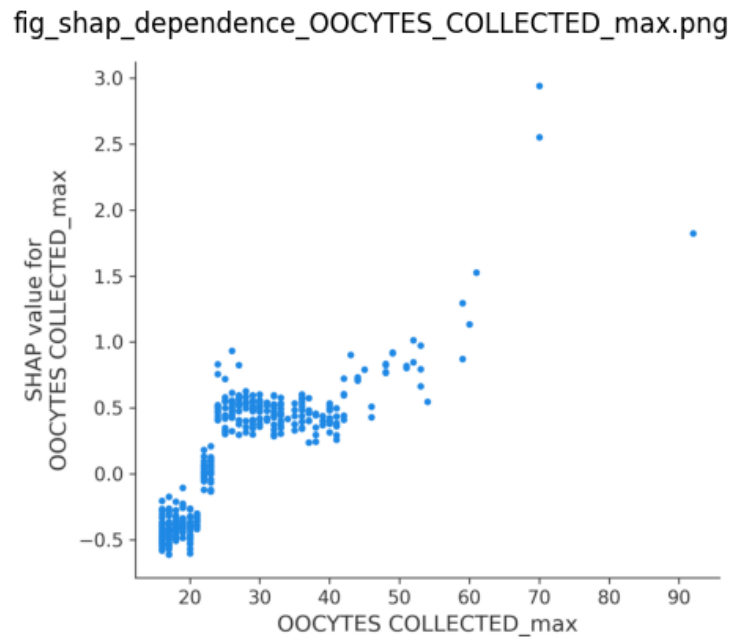


Figura 10

Gráfico que ilustra el efecto del valor máximo de ovocitos recolectados sobre la predicción, destacando una contribución positiva solo en determinados rangos y un aumento de la variabilidad en valores extremos.

**Figura 11**

Dependencia SHAP del número de embriones de alta calidad en día 5 (AA, AB, BA, BB), evidenciando su contribución positiva y consistente a la probabilidad de implantación clínica.

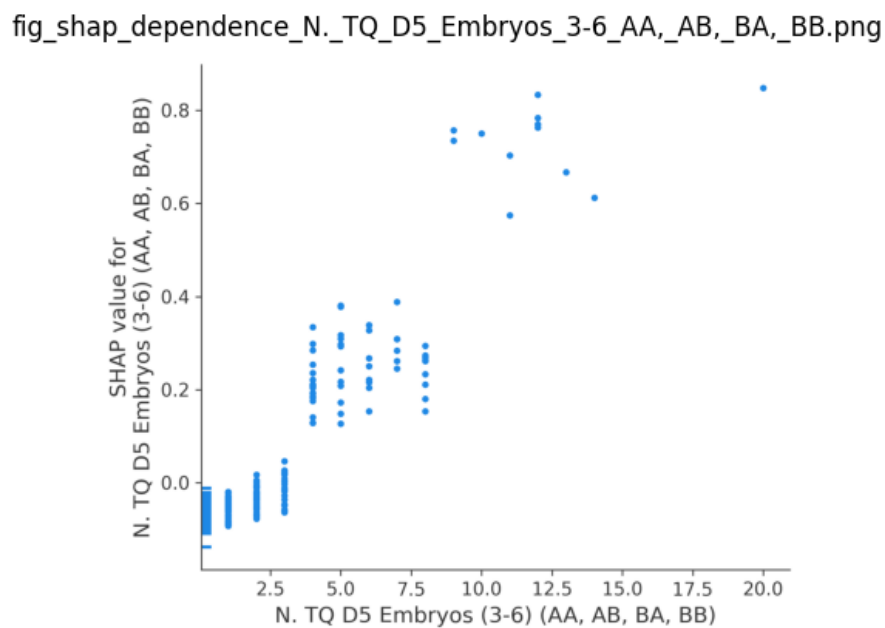
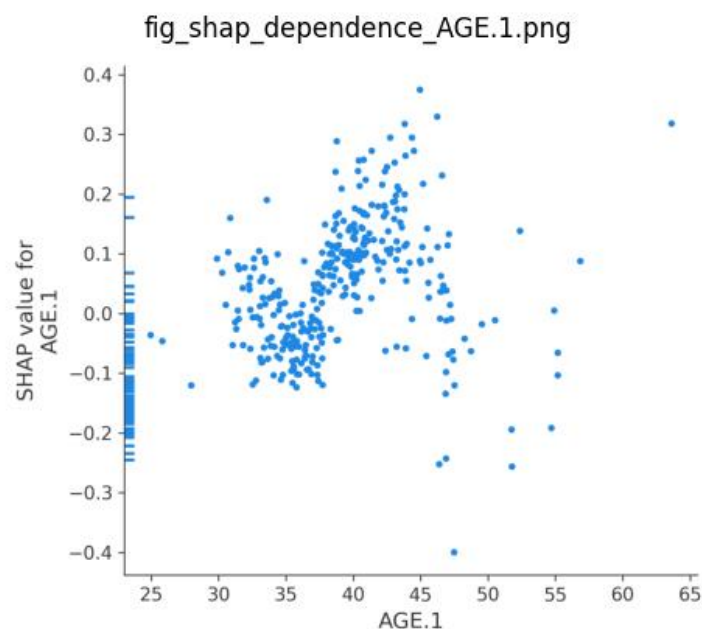


Figura 12

Gráfico que representa la relación entre la edad materna y su impacto en la predicción del modelo, mostrando un patrón no lineal coherente con la biología reproductiva.

**Figura 13**

Curva dosis-respuesta que muestra la tasa observada de embarazo clínico en función de deciles del número de ovocitos recolectados, evidenciando un patrón de rendimientos decrecientes.

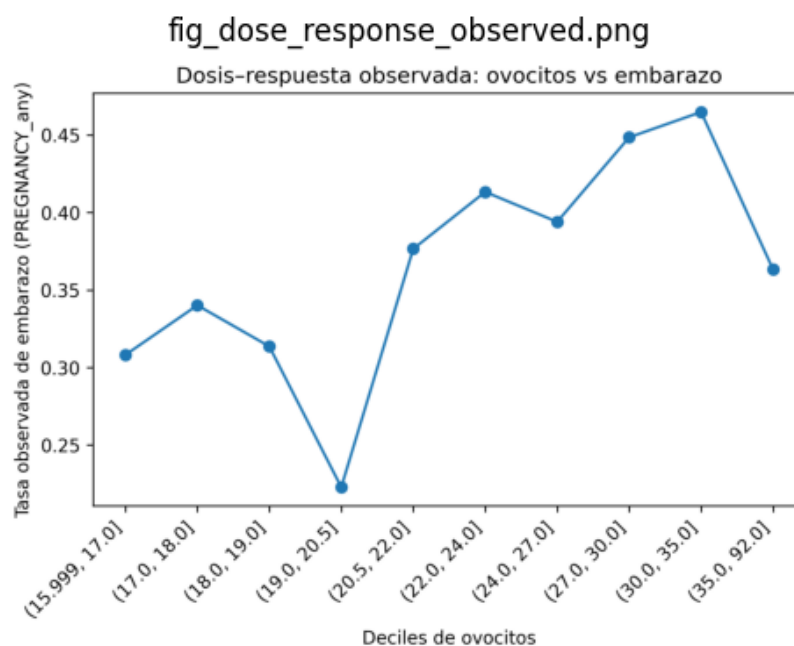


Figura 14

Curva de calibración que compara probabilidades predichas y observadas para el modelo de regresión logística, permitiendo evaluar la confiabilidad de las estimaciones.

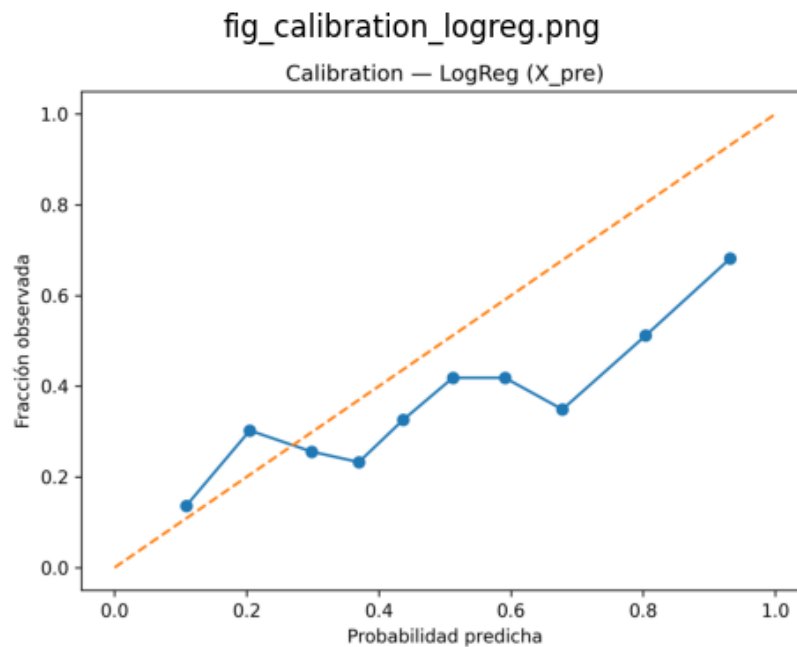
**Figura 15**

Gráfico de calibración posterior a Platt scaling que muestra una mejora en la correspondencia entre riesgo predicho y riesgo observado en el modelo CatBoost.

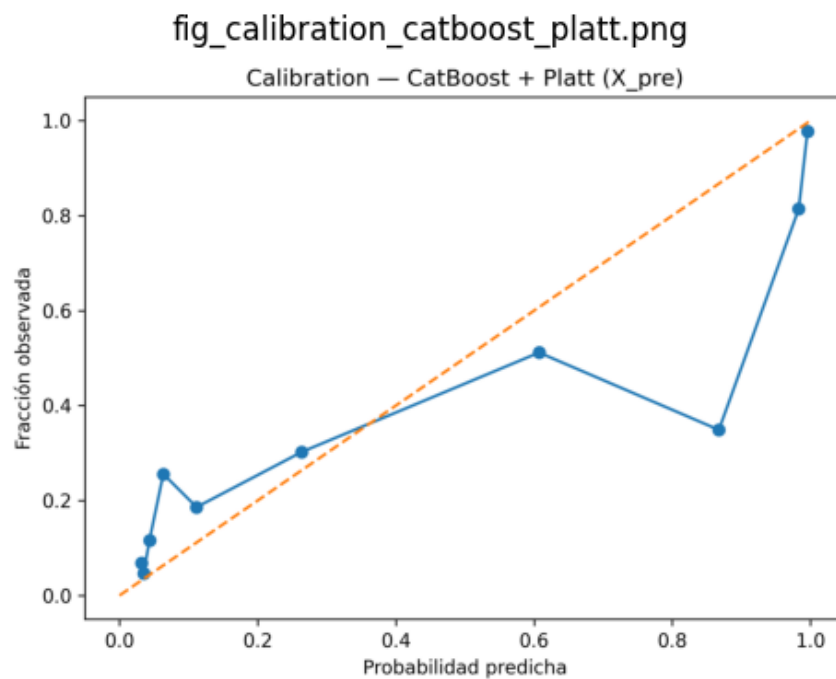
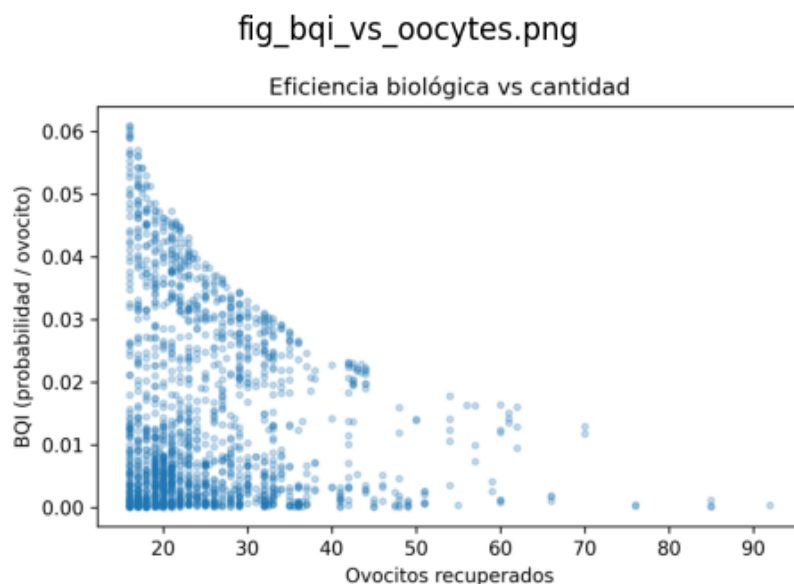


Figura 16

Gráfico de dispersión que relaciona la eficiencia biológica (probabilidad de implantación por ovocito) con el número total de ovocitos recuperados, evidenciando una disminución de eficiencia a altas cantidades.

**Figura 17**

Curva de calibración que compara probabilidades predichas y observadas para el modelo de regresión logística, permitiendo evaluar la confiabilidad de las estimaciones.

