



Programa de Doctorado 3139

Medicina

TESIS DOCTORAL

**VALIDACIÓN DE UN ALGORITMO DE
INTELIGENCIA ARTIFICIAL PARA LA
SELECCIÓN EMBRIONARIA EN
TRATAMIENTOS DE FECUNDACIÓN
IN VITRO (FIV)**

Directores:

Dr. Marcos Meseguer Escrivá

Dra. Daniela Galliano

Dr. Antonio Pellicer

MARCO TOSCHI

Valencia, abril 2024

Programa de Doctorado 3139 Medicina:

Obstetricia, ginecología y medicina regenerativa.

**VALIDACIÓN DE UN ALGORITMO DE
INTELIGENCIA ARTIFICIAL PARA LA
SELECCIÓN EMBRIONARIA EN
TRATAMIENTOS DE FECUNDACIÓN IN
VITRO (FIV).**

TESIS DOCTORAL

Presentada por:

Marco Toschi

Dirigida por:

Dr. Marcos Meseguer Escrivá

Dr. Antonio Pellicer Martínez

Dra. Daniela Galliano

Valencia, abril 2024

“La única constante en la vida es el cambio”

Heráclito

Dedicado a mis padres



Programa de Doctorado 3139 Medicina:

Obstetricia, ginecología y medicina regenerativa.

**VALIDACIÓN DE UN ALGORITMO DE INTELIGENCIA
ARTIFICIAL PARA LA SELECCIÓN EMBRIONARIA EN
TRATAMIENTOS DE FECUNDACIÓN IN VITRO (FIV).**

*VALIDATION OF AN ARTIFICIAL INTELLIGENCE
ALGORITHM FOR EMBRYO SELECTION IN IN VITRO
FERTILIZATION (IVF) TREATMENTS.*

TESIS DOCTORAL

Marco Toschi

Dirigida por:

Dr. Marcos Meseguer Escrivá, Dr. Antonio Pellicer
Martínez y Dra. Daniela Galliano

Valencia, abril 2024

INFORME DIRECTORES/AS Y TUTOR/A PARA DEPÓSITO DE TESIS

Director (es) / Codirector (es) y Tutor/a:

- 1.- Apellidos y nombre: MESEGUER ESCRIVÁ, MARCOS N.I.F. 73776663T, Departamento/Instituto: Laboratorio de Fecundación in vitro Centro: IVI Valencia
- 2.- Apellidos y nombre: GALLIANO, DANIELA N.I.F. CA32664CC Departamento/Instituto: Directora Clínica PMA, Centro: IVI Roma
- 3.- Apellidos y nombre: PELLICER MARTÍNEZ, ANTONIO N.I.F. 19979620A, Departamento/Instituto: Obstetricia y ginecología Centro: IVI Valencia

Directores/as y Tutor/a de la tesis doctoral: "Validación de un algoritmo de Inteligencia Artificial para la selección embrionaria en tratamientos de fecundación in vitro (FIV)."

de D/Dña. Marco Toschi,

estudiante del Programa de Doctorado **3139 Medicina** (RD99/2011) en Medicina de la Universitat de València, emiten informe favorable para la realización del depósito y la defensa de la tesis doctoral.

Fecha: 15/04/2024

Fdo.: Marcos Meseguer Escrivá
MESEGUER ESCRIVÁ MARCOS
- 73776663T
Director/a

Firmado digitalmente por MESEGUER ESCRIVÁ MARCOS - 73776663T
Fecha: 2024.05.07
23:58:43 +02'00'

Fdo.: Daniela Galliano
Firmado digitalmente da
Daniela Rina Camilla Galliano
C = IT
Dada, 5 de mayo de 2024 a las 17:20:28
Director/a

Fdo.: Antonio Pellicer Martinez
Firmado digitalmente da
ANTONIO PELLICER MARTINEZ
CN = ANTONIO PELLICER MARTINEZ
C = IT
Director/a y Tutor/a

ESCUELA DOCTORAL
UNIVERSITAT DE VALÈNCIA

Agradecimientos

Mediante estas palabras, deseo expresar mi más profundo agradecimiento a todas las personas que contribuyeron al éxito de esta investigación.

Quisiera comenzar expresando mi gratitud a mis directores de tesis, el Dr. Marcos Meseguer, la Dra. Daniela Galliano y el Dr. Antonio Pellicer, cuya colaboración fue fundamental para la realización de los estudios presentados en esta tesis doctoral.

A todo el equipo de IVIRMA Valencia e IVIRMA Roma. Sobre todo, a la Dra. Lorena Bori Arnal por su apoyo científico y humano. Gracias a todas las personas que han colaborado directa o indirectamente en este trabajo.

A mi familia, por su amor incondicional, apoyo constante y ánimo inquebrantable a lo largo de este viaje. Su presencia ha sido mi mayor fortaleza y motivación. Gracias por estar siempre a mi lado.

Índice

Índice

Símbolos, abreviaturas y siglas	17
Lista de figuras	23
Lista de tablas	28
Resumen	32
Summary	38
Parte 1. Prólogo	44
1.1. Introducción: Inteligencia artificial	48
Conceptos básicos	48
La inteligencia artificial en medicina	51
Automatización de la evaluación embrionaria	52
Uso de inteligencia artificial en la selección de embriones ..	59
1.2. Justificación	72
1.3. Objetivos	76
1.4. Procedimientos intrínsecos de los tratamientos de FIV analizados	78
Estimulación ovárica y receptividad uterina	78
Recuperación de ovocitos e inyección intracitoplasmática de espermatozoides (ICSI)	79
Incubación y evaluación de embriones	80
Diagnóstico genético preimplantacional de aneuploidías (PGT-A)	81
Vitrificación y desvitrificación embrionaria	81
Transferencia de embriones y soporte lúteo	82

Parte 2. Descripción de estudios	84
2.1. Objetivo específico 1. Diseñar una herramienta interpretable para automatizar la evaluación de embriones humanos mediante machine learning.	86
Propósito	86
Materiales y métodos específicos	86
Resultados	92
Discusión	95
2.2. Objetivo específico 2. Validar externamente la puntuación automática asignada mediante deep learning para clasificar embriones en función de su potencial de implantación y nacimiento vivo.	100
Propósito	100
Materiales y métodos específicos	100
Resultados	104
Discusión	115
2.3. Objetivo específico 3. Analizar el nivel de acuerdo entre la selección de embriones mediante inteligencia artificial y la decisión de los embriólogos.	120
Propósito	120
Material y métodos específicos	120
Resultados	121
Discusión	124
Parte 3. Epílogo	126
3.1. Consecución de objetivos y relevancia científica	128
3.2. Perspectivas futuras	132

3.3. Conclusiones.....	134
Material suplementario.....	136
Contribuciones científicas.....	138
Figuras suplementarias	140
Tablas suplementarias.....	144
Bibliografía.....	146

Símbolos, abreviaturas y siglas

μl: Microlitros

AG: Algoritmo genético

AI: Inteligencia artificial, del inglés *artificial intelligence*

ANN: Red neuronal artificial, del inglés *artificial neural network*

ASEBIR: Asociación para el estudio de la biología de la reproducción

AUC: Área bajo la curva

BE: Blastocisto expandido

BHi: Blastocisto eclosionando

CO²: Dióxido de carbono

DS: Desviación estándar

Eeva: Early embryo viability assessment

EMA: Embryology management assistant

GA: Algoritmos genéticos, del inglés *Genetic algorithms*

GEE: Ecuación de estimación generalizada

GG: Sistema de clasificación de Gardner

GLCM: Matriz de Co-ocurrencia de Niveles de Gris

GnRH: hormona liberadora de gonadotropina

h: Horas

hCG: Hormona gonadotropina coriónica

Hpi: Horas postICSI

IA: Inteligencia artificial

IC: Intervalo de confianza.

ICSI: Inyección intracitoplasmática

IMC: Índice de masa corporal

IVF: Fecundación in vitro, del inglés *in vitro fertilization*.

K: coeficiente Kappa

KID: Datos de implantación conocidos

LWG: Life whisperer genetic

LWV: Life whisperer viability

MCI: Masa celular interna

Mg: Miligramos

MII: Ovocitos metafase II

ml: Mililitros

Mm: Milímetros

O²: Oxígeno

OR: Odds ratio

PGT-A: Test genético preimplantatorio para aneuploidías

PLN: Procesamiento del lenguaje natural

PN: Pronúcleos

RNV: Recién nacido vivo

RR: Riesgo relativo

SET: Transferencia de un único embrión, del inglés *single embryo transfer*

SPSS: Statistical Package for Social Sciences

t2: Tiempo de división a dos células

t3: Tiempo de división a tres células

t4: Tiempo de división a cuatro células

t5: Tiempo de división a cinco células

t6: Tiempo de división a seis células

t7: tiempo de división a siete células

t8: Tiempo de división a ocho células

tB: Tiempo de formación de blastocisto

TE: Transferencia embrionaria

tEB: Tiempo de blastocisto expandido

tHib: Tiempo de inicio de eclosión

TL: Lapso de tiempo, del inglés *time-lapse*

tM: Tiempo de formación de mórula

tPNa: Tiempo de aparición de los pronúcleos

tPNf: Tiempo de desaparición de los pronúcleos⁷

TRA: Técnicas de reproducción asistida

tSB: Tiempo de inicio de blastulación

tSC: tiempo de inicio de compactación

UI: Unidades internacionales

v1: Primera versión

v2: Segunda versión

v3: Tercera versión

Lista de figuras

Figura 1. Comparación de la estructura de una neurona biológica (a) con un modelo basado en redes neuronales artificiales (b).

Figura 2. Parámetros morfocinéticos y algoritmo de selección embrionaria para predecir implantación.

Figura 3. Ajuste del nivel de confianza para las anotaciones guiadas.

Figura 4. Demostración de la categorización proporcionada por el algoritmo Eeva para una cohorte embrionaria en el software Geri Connect & Assess.

Figura 5. Presentación de la clasificación proporcionada por el algoritmo KIDScore D5 v3 para un grupo de 16 embriones en el software EmbryoViewer®.

Figura 6. Demostración del software EMA.

Figura 7. Demostración del software Life Whisperer Viability.

Figura 8. Informe para el paciente generado por la plataforma Life Whisperer Genetics,

Figura 9. Demostración del software Chloe.

Figura 10. Presentación del software Embryo Assist.

Figura 11. Demostración de la clasificación proporcionada por el algoritmo iDAScore v2 para una cohorte embrionaria.

Figura 12. A) Imagen de un blastocisto que muestra la estandarización de una imagen de embrión humano. B) Blastocisto humano aislado con la función de Hough para determinar la circunferencia que mejor caracteriza al embrión.

Figura 13. Secuencia de pasos necesarios para segmentar una imagen digital del blastocisto humano.

Figura 14. Secuencia de pasos necesarios para segmentar una imagen digital del blastocisto humano

Figura 15. Distribución de la puntuación automática según la morfología convencional de acuerdo con los criterios ASEBIR.

Figura 16. Distribución de la puntuación automática según la ploidía.

Figura 17. Tasa de embarazo, implantación y recién nacido vivo (RNV) según la puntuación automática (iDAScore v2) por cuartiles.

Figura 18. Distribución de la puntuación automática de embriones según dos resultados clínicos.

Lista de tablas

Objetivo específico 1

Tabla 1. Precisión general de la herramienta de inteligencia artificial para la expansión del blastocisto, la calidad de la masa celular interna y el trofoblasto en los datos de prueba y entrenamiento.

Tabla 2. Rendimiento del modelo con valores de AUC para la expansión del blastocisto, la masa celular interna (MCI) y la calidad del trofoectodermo (TE) en los datos de prueba y entrenamiento.

Tabla 3. Coeficientes Kappa de Cohen obtenidos de la evaluación realizada por cinco embriólogos de tres características de los blastocistos: grado de expansión, masa celular interna (MCI) y trofoectodermo.

Tabla 4. Coeficientes de Kappa de Cohen de la evaluación realizada por inteligencia artificial comparada con la moda de 5 embriólogos de tres características de los blastocistos: grado de expansión, masa celular interna y trofoectodermo.

Objetivo específico 2

Tabla 5. Tabla demográfica de tratamientos dividida por subtipos.

Tabla 6. Tasa de implantación esperada según el origen de los ovocitos, el tipo de transferencia embrionaria (TE) y la puntuación automática asignada por iDAScore v2.

Tabla 7. Ecuaciones de estimación generalizada (GEE) que evalúan la correlación entre iDAScore y los resultados de implantación controlando otros posibles factores de confusión.

Tabla 8. Ecuaciones de estimación generalizada (GEE) que evalúan la correlación entre iDAScore y los resultados de nacimiento vivo, controlando otros posibles factores de confusión.

Tabla 9. Métricas de rendimiento para las ecuaciones de estimación generalizada (GEEs) de los dos sistemas de clasificación, iDAScore y evaluación morfológica, para la predicción de implantación (a) y nacimientos vivos (b).

Objetivo específico 3

Tabla 10. Número de embriones transferidos (en concordancia y discordancia entre los algoritmos de IA y los embriólogos senior) e incidencia de implantación.

Resumen

La inteligencia artificial (IA) es un campo de la informática que se enfoca en desarrollar sistemas capaces de llevar a cabo tareas que requieren inteligencia humana. En medicina, la IA está transformando el panorama al proporcionar herramientas y técnicas avanzadas para mejorar el diagnóstico, tratamiento y manejo de enfermedades. Las redes neuronales artificiales, inspiradas en el cerebro humano, son fundamentales en la IA. En el ámbito médico, estas redes se emplean para diversas aplicaciones, como el análisis de imágenes médicas, incluyendo radiografías, tomografías computarizadas y resonancias magnéticas, para detectar patrones y características que pueden pasar desapercibidas al ojo humano. En medicina reproductiva, la inteligencia artificial y las redes neuronales están empezando a desempeñar un papel crucial. Por ejemplo, en la fecundación in vitro (FIV), las redes neuronales artificiales pueden ayudar a prever la probabilidad de éxito de un ciclo de FIV basándose en factores como la edad de la paciente, la calidad de los embriones y el historial médico. Dentro de la FIV, la IA está transformando la forma en que se evalúan y seleccionan los embriones, con el objetivo de mejorar las tasas de éxito de los tratamientos de fertilidad.

Uno de los principales retos en el laboratorio de FIV es la identificación de los embriones con mayor probabilidad de implantación y desarrollo exitoso. Históricamente, esta evaluación se ha basado en criterios morfológicos que pueden ser subjetivos. Sin embargo, con la incorporación de técnicas de inteligencia artificial, como las redes neuronales artificiales, se busca lograr una

evaluación más precisa y objetiva. Estas redes pueden analizar imágenes de embriones en diferentes etapas de desarrollo y ofrecer una evaluación automatizada y precisa de su viabilidad. Se han creado diversos modelos basados en inteligencia artificial para calificar o clasificar embriones según su viabilidad. Sin embargo, la mayoría de estos modelos de inteligencia artificial muestran una falta de transparencia, lo que dificulta su interpretación, y no han sido validados externamente con un tamaño muestral suficiente para su aplicación generalizada.

En esta tesis doctoral, planteamos la hipótesis de que el uso de redes neuronales artificiales podría ser una opción para desarrollar un modelo de inteligencia artificial que anote de manera objetiva y transparente las características morfológicas de los blastocistos. Además, considerando la amplia aplicación de las metodologías basadas en deep learning en la evaluación y selección de embriones, buscamos evaluar la confiabilidad de estos algoritmos y comparar su concordancia con la de los embriólogos experimentados. Por lo tanto, nuestro objetivo principal fue clasificar las categorías morfológicas de los blastocistos según el éxito de un tratamiento de fecundación in vitro (FIV) mediante modelos de inteligencia artificial (IA) basados en redes neuronales artificiales, y evaluar la precisión y consistencia de la IA en comparación con la clasificación realizada por los embriólogos. Para lograrlo, establecimos tres objetivos específicos:

Objetivo específico 1. Diseñar una herramienta interpretable para automatizar la evaluación de embriones humanos mediante machine learning.

Objetivo específico 2. Validar externamente la puntuación automática asignada mediante deep learning para clasificar embriones en función de su potencial de implantación y nacimiento vivo.

Objetivo específico 3. Analizar el nivel de acuerdo entre la selección de embriones mediante inteligencia artificial y la decisión de los embriólogos

Summary

Artificial intelligence (AI) is a field of computer science that focuses on developing systems capable of performing tasks that require human intelligence. In medicine, AI is transforming the landscape by providing advanced tools and techniques to enhance the diagnosis, treatment, and management of diseases. Artificial neural networks, inspired by the human brain, are pivotal in AI. In the medical field, these networks are employed for various applications, such as the analysis of medical images, including X-rays, computed tomography scans, and magnetic resonance imaging, to detect patterns and features that may go unnoticed by the human eye. In reproductive medicine, artificial intelligence and neural networks are beginning to play a crucial role. For instance, in in vitro fertilization (IVF), artificial neural networks can help predict the likelihood of success of an IVF cycle based on factors such as the patient's age, embryo quality, and medical history. Within IVF, AI is transforming the way embryos are evaluated and selected, aiming to improve the success rates of fertility treatments.

One of the main challenges in the IVF laboratory is the identification of embryos with the highest likelihood of implantation and successful development. Historically, this evaluation has been based on morphological criteria that can be subjective. However, with the incorporation of artificial intelligence techniques, such as artificial neural networks, the aim is to achieve a more precise and objective assessment. These networks can analyze images of embryos at different stages of development and provide automated and accurate assessment of their viability. Various artificial intelligence-based

models have been created to grade or classify embryos according to their viability. However, most of these artificial intelligence models lack transparency, making interpretation difficult, and have not been externally validated with a large enough sample size for widespread application.

In this doctoral thesis, we hypothesized that the use of artificial neural networks could be an option to develop an artificial intelligence model that objectively and transparently annotates the morphological characteristics of blastocysts. Additionally, considering the widespread application of deep learning methodologies in the evaluation and selection of embryos, we sought to assess the reliability of these algorithms and compare their agreement with that of experienced embryologists. Therefore, our main objective was to classify the morphological categories of blastocysts according to the success of in vitro fertilization (IVF) treatment using artificial intelligence (AI) models based on artificial neural networks, and evaluate the accuracy and consistency of AI compared to the classification performed by embryologists. To achieve this, we established three specific objectives:

Specific objective 1. Design an interpretable tool to automate the evaluation of human embryos using machine learning.

Specific objective 2. Externally validate the automatically assigned score using deep learning to classify embryos based on their potential for implantation and live birth.

Specific objective 3. Analyze the level of agreement between embryo selection using artificial intelligence and the decision of embryologists.

Parte 1. Prólogo

La estructura de esta tesis doctoral se divide en tres secciones principales: un prólogo que presenta la justificación y los objetivos de la investigación, junto con una introducción general sobre la inteligencia artificial y la descripción de los procedimientos intrínsecos de los tratamientos de fecundación in vitro (FIV) incluidos; una sección que detalla los estudios llevados a cabo para alcanzar los objetivos, organizados en tres objetivos específicos; y un epílogo que resume el logro de los objetivos, la importancia de los resultados, nuevas direcciones de investigación y conclusiones finales. Además, se incluye un apartado al final del documento que contiene las contribuciones científicas, figuras y tablas adicionales, así como la lista de referencias bibliográficas utilizadas en la tesis.

El objetivo específico 1 se basa en un extenso estudio en el que se diseña una herramienta interpretable para automatizar la evaluación de embriones humanos mediante la integración de redes neuronales artificiales (ANN) y algoritmos genéticos (GA).

En el objetivo específico 2 se evalúa masivamente la puntuación automática asignada por un algoritmo de deep learning como herramienta de apoyo en los laboratorios de FIV.

El objetivo específico 3 se centra en analizar la concordancia y discordancia entre embriólogos senior y la puntuación automática asignada por varios modelos de inteligencia artificial (IA) para cada cohorte de embriones.

1.1. Introducción: Inteligencia artificial

Conceptos básicos

La inteligencia artificial (IA) es un campo de la informática que se centra en la creación de sistemas capaces de realizar tareas que requieren inteligencia humana (Simopoulou et al., 2018). Estos sistemas pueden aprender, razonar, percibir, comprender y actuar de manera similar a los seres humanos, lo que los hace extremadamente versátiles y potencialmente útiles en una amplia variedad de aplicaciones.

Uno de los conceptos básicos de la inteligencia artificial es el aprendizaje automático (machine learning), que es una rama que se enfoca en desarrollar algoritmos y técnicas que permiten a las computadoras aprender de los datos sin ser explícitamente programadas para ello. Dentro del aprendizaje automático, las redes neuronales artificiales (ANN) juegan un papel crucial. Estas son estructuras inspiradas en la biología (Figura 1) que consisten en capas de nodos interconectados que procesan información (Curchoe & Bormann, 2019). El aprendizaje en estas redes se basa en ajustar los pesos de las conexiones entre nodos para que el modelo pueda realizar tareas específicas, como reconocimiento de patrones, clasificación de imágenes o procesamiento de lenguaje natural.

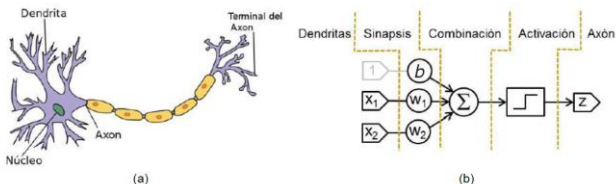


Figura 1. Comparación de la estructura de una neurona biológica (a) con un modelo basado en redes neuronales artificiales (b). Las dendritas reciben señales que actúan como entradas a la neurona, y en el núcleo estas señales se combinan mediante una suma ponderada, donde los pesos representan la fuerza de las conexiones sinápticas. La neurona se activa o dispara si el resultado de esta combinación supera un umbral específico, transmitiendo luego la información a la siguiente neurona. El axón, a su vez, funciona como la salida de la neurona, permitiendo la generación de señales de salida a partir de la combinación de las señales de entrada y una regla de activación específica (conocida en las neuronas artificiales como función de activación).

El aprendizaje profundo (deep learning) es un subcampo del aprendizaje automático que se centra en entrenar modelos de redes neuronales artificiales con múltiples capas de procesamiento para aprender representaciones jerárquicas de datos (Lecun et al., 2015). A diferencia de los modelos de aprendizaje automático tradicionales, que pueden requerir la extracción manual de características, el aprendizaje profundo es capaz de aprender automáticamente características relevantes directamente de los datos (Kim, 2016)

Otro concepto fundamental es el procesamiento del lenguaje natural (PLN), que se refiere a la capacidad de las computadoras para

comprender, interpretar y generar lenguaje humano de manera natural. Esto incluye tareas como la traducción automática, la generación de texto y el análisis de sentimientos. Sus aplicaciones en el ámbito sanitario son muy amplias y crecen considerablemente día a día (por ejemplo, para reducir de la carga que supone la documentación clínica) (Voytovich et al., 2022).

La IA también incluye áreas como la visión por computadora, que se centra en permitir que las computadoras vean y comprendan el mundo visual, y la robótica, que busca desarrollar robots inteligentes capaces de interactuar con su entorno de manera autónoma (Danuser, 2011). La embriología puede obtener ventajas del empleo de análisis de imágenes a través de visión computacional, una técnica que ya se ha utilizado para abordar problemas médicos y biológicos en niveles moleculares, subcelulares y supracelulares (Peng et al., 2016).

En resumen, la inteligencia artificial es un campo multidisciplinario que combina la informática, las matemáticas, la estadística y la ingeniería para crear sistemas inteligentes capaces de simular la inteligencia humana en diversas aplicaciones. Su potencial para transformar industrias y mejorar la vida cotidiana es enorme, y su desarrollo continúa avanzando a pasos agigantados, en gran parte gracias al uso de redes neuronales artificiales como una de sus metodologías principales.

La inteligencia artificial en medicina

La inteligencia artificial está revolucionando el campo de la medicina al proporcionar herramientas y técnicas avanzadas para mejorar el diagnóstico, el tratamiento y la gestión de enfermedades (Gennatas & Chen, 2021).

En medicina, las redes neuronales artificiales se utilizan para una variedad de aplicaciones, incluido el análisis de imágenes médicas, como radiografías, tomografías computarizadas y resonancias magnéticas, para detectar patrones y características que pueden ser difíciles de identificar para el ojo humano (Garbade, 2018). Las ANN pueden emplearse en la predicción de enfermedades, el desarrollo de modelos de riesgo y la personalización de tratamientos basados en datos clínicos y genéticos de los pacientes.

En el campo de la medicina reproductiva, la inteligencia artificial y las redes neuronales están empezando a desempeñar un papel crucial. Por ejemplo, en la fecundación in vitro, las ANN pueden ayudar a predecir la probabilidad de éxito de un ciclo de FIV en función de diversos factores, como la edad de la paciente, la calidad de los embriones y el historial médico. Además, en el diagnóstico de problemas de fertilidad, las redes neuronales pueden analizar datos hormonales, ecográficos y genéticos para identificar posibles causas de infertilidad y ayudar a guiar el tratamiento. En el análisis de imagen se utilizan ANNs complejas, como las redes neuronales convolucionales (Matusevičius et al., 2017). Por ejemplo, para anticipar la calidad de los blastocistos utilizando este método, el

proceso de aprendizaje se realizaría utilizando miles de embriones, sin necesidad de extraer variables de las imágenes (Khosravi et al., 2019).

Otro campo importante es el monitoreo del embarazo, donde las ANN pueden analizar datos de ultrasonido, monitoreo fetal y otros parámetros para detectar posibles complicaciones y predecir el riesgo de parto prematuro u otras complicaciones obstétricas (Gupta et al., 2022).

En resumen, las redes neuronales artificiales son una herramienta poderosa en la inteligencia artificial aplicada a la medicina, con un potencial significativo para mejorar la precisión del diagnóstico, optimizar el tratamiento y avanzar en la medicina reproductiva para ayudar a las parejas a concebir y tener embarazos saludables.

Automatización de la evaluación embrionaria

La medicina reproductiva es una disciplina que está experimentando avances significativos gracias al uso de la inteligencia artificial. En particular, dentro del ámbito de la fecundación in vitro, la IA está revolucionando la forma en que se evalúan y seleccionan los embriones para mejorar las tasas de éxito de los tratamientos de fertilidad (Dimitriadis et al., 2022).

Uno de los principales desafíos en el laboratorio de FIV es identificar los embriones con mayor probabilidad de implantación y desarrollo exitoso. Tradicionalmente, esta evaluación se ha basado en criterios morfológicos que pueden ser subjetivos, pero con la introducción de

técnicas de IA, como las redes neuronales artificiales, se intenta lograr una evaluación más precisa y objetiva. En el contexto de la medicina reproductiva, estas redes pueden analizar imágenes de embriones en diferentes etapas de desarrollo y proporcionar una evaluación automatizada y precisa de su viabilidad.

Parámetros morfocinéticos y anotaciones manuales

La implementación de incubadoras equipadas con sistemas de time-lapse (TL) en los laboratorios de FIV ha permitido la observación continua del desarrollo embrionario en tiempo real. Esta tecnología se ha convertido en una herramienta valiosa para examinar el progreso dinámico del embrión sin perturbar las condiciones de cultivo. Además, proporciona información cualitativa y cuantitativa de manera objetiva y precisa (Aparicio-Ruiz et al., 2016). La introducción de la tecnología time-lapse en los laboratorios de FIV ha facilitado la determinación de los parámetros morfocinéticos del desarrollo embrionario. Esta combinación de parámetros morfológicos y morfocinéticos ha dado lugar al desarrollo de varios algoritmos (Figura 2) para explorar variables relevantes en la selección de embriones (Cruz et al., 2012; Dal Canto et al., 2012; Meseguer et al., 2011; Motato et al., 2016).

Se han propuesto anotaciones manuales de parámetros cuantitativos y cualitativos para prever la formación de blastocistos, el potencial de implantación e incluso el nacimiento vivo. La escisión temprana (Milewski et al., 2015) y el tiempo de formación de la mórula (Motato et al., 2016) se han utilizado como indicadores para

Además, se considera que patrones de división anormales, multinucleación, fragmentación o colapso son parámetros con pronóstico desfavorable para clasificar la calidad del embrión (Zaninovic et al., 2017). Estos eventos han sido asociados con la mala formación de blastocistos (Wirka et al., 2014), poca tasa de embriones euploides (Zhan et al., 2016) y bajas tasas de resultados clínicos (Aguilar et al., 2016; Azzarello et al., 2017; Desai et al., 2014, 2016; Desch et al., 2017; Ebner et al., 2017; Goodman et al., 2016; Rubio et al., 2012; Zhan et al., 2016).

El principal desafío para la amplia adopción de los algoritmos de evaluación y selección embrionaria radica en la falta de reproducibilidad (Barrie et al., 2017; Fréour et al., 2015). Se han identificado posibles factores de confusión que podrían influir en el rendimiento de los algoritmos de sistemas TL, como las diferencias en el entorno y las técnicas utilizadas en cada laboratorio, las características específicas de los pacientes y la subjetividad inherente a las anotaciones manuales (ESHRE Working group on time-lapse technology et al., 2020; Martínez-Granados et al., 2017; Sundvall et al., 2013).

Anotaciones automáticas

La incorporación de sistemas de visión computacional en el campo de la medicina reproductiva ha resultado en una reducción significativa de la subjetividad en la selección de embriones. Estos sistemas, como señala Danuser (2011), son capaces de extraer

información objetiva y estandarizada de imágenes, lo que permite anotaciones automáticas y una evaluación más precisa.

En la embriología humana, Wong et al. (2010) llevaron a cabo la primera aproximación predictiva utilizando imágenes proporcionadas por sistemas de time-lapse. Desarrollaron un algoritmo para cuantificar automáticamente la duración de la citocinesis y los tiempos entre mitosis hasta el estadio de 4 células. Este sistema, conocido como Eeva™ (Early Embryo Viability Assessment), se convirtió en la primera plataforma clínicamente validada que integra time-lapse, dos parámetros predictivos y un software automático. Este sistema puede ser integrado en un incubador estándar y captura imágenes digitales de alta resolución en un solo plano a intervalos de 5 minutos. Cuenta con un microscopio invertido digital de campo oscuro, una cámara digital, y un software especializado para la adquisición de imágenes. Además, incluye una placa diseñada para el cultivo de embriones en grupo (Conaghan et al., 2013).

Asimismo, otros sistemas como el incubador Geri® (Merck, Darmstadt, Alemania) han implementado sistemas de anotaciones automáticas en su software de monitorización digital Geri® Connect and Assess 2.0. Este sistema es capaz de detectar automáticamente una serie de eventos clave en el desarrollo embrionario, como el tiempo de aparición/desaparición de los pronúcleos, tiempos de división celular y la formación de mórula y blastocisto. Las anotaciones automáticas se pueden ajustar manualmente según la

evaluación del embriólogo. Además, el sistema tiene la capacidad de identificar eventos como la división reversa o la presencia de fragmentación. También, el software emite una advertencia si alguna anotación está fuera del rango temporal establecido: desaparición de los pronúcleos, tPNf (17-30 horas); división a 2 células, t2 (20-40 horas); a 3 células, t3 (30-48 horas); a 4 células, t4 (32-54 horas); a 5 células, t5 (38-68 horas); a 6 células, t6 (46-78 horas); formación de mórula, tM (64-100 horas); formación de blastocisto temprano, tSB (86-126 horas); formación de blastocisto expandido, tBE (86-192 horas), y el inicio de eclosión a través de la zona pelúcida, tBHi (86-192 horas) (Alpha Scientists, 2011; Ciray et al., 2014).

Los incubadores EmbryoScope® y EmbryoScope Plus® (Vitrolife, Copenhague, Dinamarca) también cuentan con sistemas de anotación automática en su software EmbryoViewer® (Guided Annotations tool). Este sistema guía a los embriólogos en el proceso de realizar anotaciones, proporcionando una evaluación más simplificada y coherente del desarrollo embrionario. Todas las variables definidas en la estrategia de anotación seleccionada se presentan automáticamente junto con una imagen del embrión en ese momento específico. Posteriormente, el usuario podría modificar la anotación o confirmar las anotaciones sugeridas por la herramienta.

La función del sistema detecta automáticamente la división celular y los eventos morfológicos, e introduce estimaciones de tiempos de división y parámetros morfológicos como el número de pronúcleos y

el tamaño del embrión. El análisis automático de imágenes calcula los tiempos de división celular y los parámetros morfológicos con diferentes niveles de confianza (Figura 3).

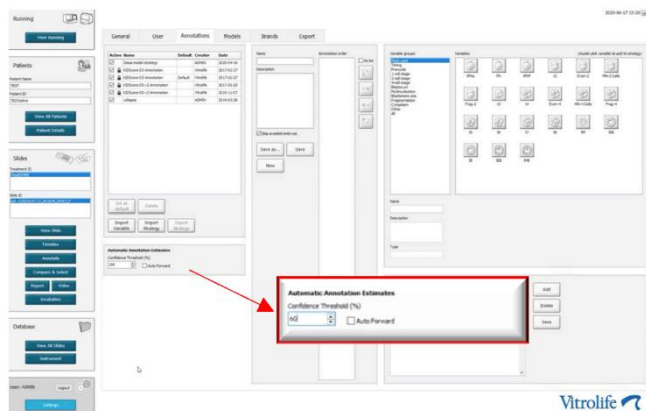


Figura 3. Ajuste del nivel de confianza para las anotaciones guiadas. La imagen completa muestra un umbral estricto, del 100%. La imagen ampliada muestra un umbral de confianza del 60%.

Al establecer un umbral de confianza alto, la mayoría de las estimaciones de tiempos y parámetros morfológicos serán precisas, aunque requerirán una verificación manual más extensa. Por otro lado, al establecer un umbral de confianza bajo, las estimaciones serán menos precisas, pero requerirán menos verificaciones manuales. La selección del umbral de confianza debe basarse en la tolerancia específica de cada clínica.

En resumen, estos sistemas de visión computacional con anotaciones automáticas permiten una evaluación objetiva y estandarizada del desarrollo embrionario en el laboratorio de FIV, mejorando la precisión y la eficiencia del proceso de selección de embriones.

Uso de inteligencia artificial en la selección de embriones

Al entrenar las redes neuronales con datos de historiales clínicos, resultados de tratamientos anteriores y características morfológicas de los embriones, se podría desarrollar un sistema de IA capaz de predecir con alta precisión la probabilidad de éxito de cada embrión. Esto permitiría a los especialistas en fertilidad tomar decisiones informadas sobre qué embriones transferir durante el ciclo de FIV, maximizando las posibilidades de embarazo.

En última instancia, el uso de IA en la evaluación de embriones en el laboratorio de FIV no solo mejoraría las tasas de éxito de los tratamientos de fertilidad, sino que también reduciría el tiempo y los costos asociados con el proceso. Además, al proporcionar una evaluación objetiva y basada en datos, ayudaría a mitigar el sesgo humano en la selección de embriones.

Eeva™

En la actualidad, el sistema Eeva™ se encuentra integrado en el incubador Geri® (Figura 4). La utilización de este software predictivo con anotaciones automáticas, combinado con la evaluación

morfológica en el tercer día de desarrollo embrionario, ha demostrado mejorar la capacidad del embriólogo para identificar embriones con potencial para formar blastocistos de alta calidad. Esta integración facilita una selección más precisa entre embriones con morfología adecuada y ayuda a reducir la variabilidad entre operadores (Conaghan et al., 2013). Además, el software Eeva también ha mostrado correlación con mejoras en la tasa de implantación y en la tasa de embarazo clínico (Adamson et al., 2016; Vermilyea et al., 2014), así como en el embarazo evolutivo (Aparicio-Ruiz et al., 2016; Valera et al., 2023).



Figura 4. Demostración de la categorización proporcionada por el algoritmo Eeva para una cohorte embrionaria en el software Geri Connect & Assess. La línea temporal muestra las anotaciones automáticas para cada evento del desarrollo embrionario. Las marcas rosas corresponden a las variables consideradas por el modelo de evaluación.

KIDScore™

Los sistemas time-lapse de Vitrolife también incluyen una herramienta de apoyo a la toma de decisiones para los embriólogos, conocida como KIDScore D5™ (software EmbryoViewer; Vitrolife), que mejora la objetividad y coherencia en la evaluación de embriones al clasificarlos según su probabilidad de implantación según anotaciones automáticas. La primera versión del algoritmo (KIDScore D5™ v1) se basaba en el número de pronúcleos (PN), la morfología del blastocisto y ocho parámetros morfocinéticos: el tiempo de desvanecimiento de los PN (tPNf), el tiempo de división en dos células (t2), tres células (t3), cuatro células (t4), cinco células (t5) y ocho células (t8), el tiempo hasta la blastulación temprana (tSB) y el tiempo hasta la formación completa del blastocisto (tB). En la segunda versión del algoritmo (KIDScore D5™ v2), se redujeron las variables de entrada a cinco parámetros morfocinéticos (t2, t3, t4, t5 y tB). Ambas versiones han sido validadas recientemente de manera retrospectiva como herramientas de apoyo a la toma de decisiones para los embriólogos (Gazzo et al., 2020; Reignier et al., 2019).

La versión más reciente, denominada KIDScore™ D5 versión 3, clasifica los embriones según su ritmo de desarrollo y la calidad del blastocisto. Fue desarrollada utilizando una base de datos multicéntrica que incluye información de más de 5.000 embriones con datos conocidos sobre implantación (embriones KID). Según los desarrolladores, esta última versión del modelo considera el número

de pronúcleos, los tiempos de división (t_2 , t_3 , t_4 , t_5 , t_B), la masa celular interna y la calidad del trofoectodermo.

Se asigna una puntuación lineal final a cada embrión, que varía de 1 a 9,9 y que representa una probabilidad de implantación que va desde baja hasta alta (Figura 5). Se recomienda utilizar la puntuación automática de embriones para aquellos procedentes de FIV convencional o ICSI, y que hayan sido incubados en condiciones de bajo oxígeno (4-6%). Esta última versión también fue validada para implantación y recién nacido vivo (Bori et al., 2022).

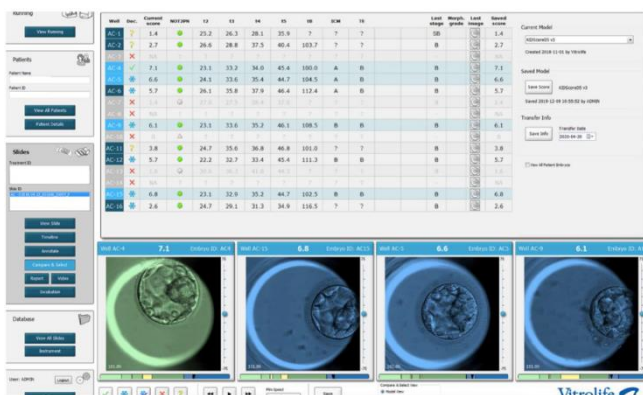


Figura 5. Presentación de la clasificación proporcionada por el algoritmo KIDScore D5 v3 para un grupo de 16 embriones en el software EmbryoViewer®. En la sección superior se muestran las anotaciones automáticas realizadas en cada embrión, mientras que la sección inferior

exhibe los cuatro embriones con la puntuación máxima en la cohorte embrionaria.

EMA™

EMA de AIVF™ (Tel Aviv, Israel) presenta un enfoque de evaluación de embriones totalmente automatizado de principio a fin, que tiene en cuenta la calidad visual, la morfocinética y la probabilidad de ploidía para cada embrión (Figura 6).

Según los desarrolladores cuenta con distintos módulos de IA basados en deep learning que agregan una capa adicional de conocimiento:

Modelo de resiliencia embrionaria: Un algoritmo que fue entrenado para identificar transitorios sustanciales en el embrión, para garantizar robustez frente a intervenciones médicas como la eclosión asistida o el colapso artificial, así como escenarios biológicos que pueden ocurrir durante el desarrollo embrionario in vitro.

Modelo de morfocinética: Un algoritmo basado en aprendizaje profundo que fue entrenado para detectar automáticamente eventos dinámicos durante el desarrollo embrionario in vitro, que van desde el momento de desvanecimiento de los pronúcleos hasta el momento de la blastulación. Incluye: T2, T3, T4, T5, T8, TM, TB, así como otras características de la etapa de blastocisto.

Modelo de segmentación: Un algoritmo de segmentación basado en aprendizaje profundo que ayuda a detectar y medir la forma, tamaño,

estructura y posición del embrión dentro del pocillo de cultivo para evaluaciones de imágenes significativas y precisas por parte del modelo de aprendizaje automático.

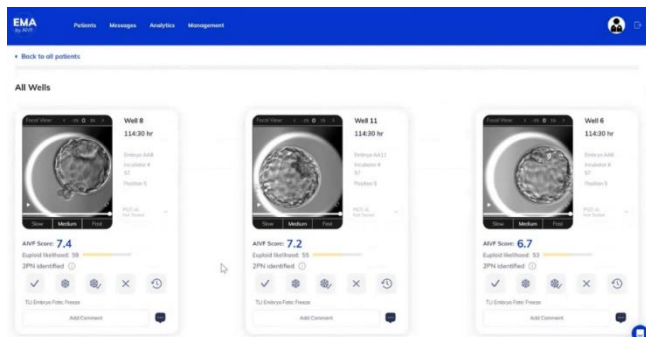


Figura 6. Demostración del software EMA. Tres embriones incubados en sistemas time-lapse y puntuados por AivF Score según viabilidad (0-10) y probabilidad de euploidía (0-100).

El sistema EMA se validó de forma retrospectiva y los resultados se han mostrado en congresos internacionales para viabilidad (Bori et al., 2021) y ploidía (Meseguer et al., 2022).

Life Whisperer Viability (LWV) and Life Whisperer Genetics (LWG)

Ambas son aplicaciones web instantáneas, objetivas y no invasivas diseñadas para ayudar en el proceso de toma de decisiones para selección de embriones. Las dos plataformas son fáciles de usar y completamente automatizadas, sin necesidad de anotaciones

manuales. Además, funcionan con sistemas de cámara estándar o TL. Simplemente se arrastran y sueltan imágenes estáticas de embriones en la plataforma y se reciben resultados instantáneos.

Los algoritmos de IA evalúan características morfológicas que se correlacionan con la probabilidad de embarazo clínico (LWV) (Figura 7) y con la probabilidad de que un embrión sea genéticamente normal o euploide (LWG) (Figura 8).

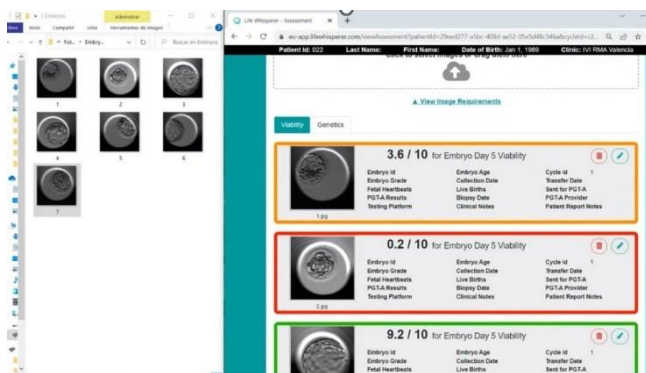


Figura 7. Demostración del software Life Whisperer Viability. La parte izquierda muestra las imágenes de los embriones sujetos a evaluación. La parte derecha muestra la puntuación obtenida por cada embrión según el algoritmo LWV (0-10).

Estudios clínicos de la propia casa comercial han demostrado que LWV es hasta un 25% más preciso que los métodos de calificación manual para identificar embriones viables que conducen a un

embarazo clínico (VerMilyea et al., 2021), lo que puede reducir el tiempo hasta el embarazo en un 12% (Diakiw, Hall, VerMilyea, Lim, et al., 2022). Otro estudio propio ha demostrado que LWG identificó al embrión clasificado como el de mayor calidad como euploide en un 82% de las cohortes (Diakiw, Hall, VerMilyea, Amin, et al., 2022a).

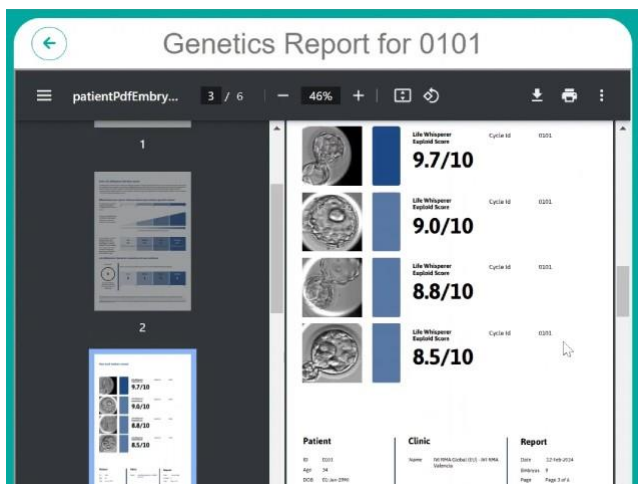


Figura 8. Informe para el paciente generado por la plataforma Life Whisperer Genetics, incluyendo las imágenes de los embriones y la puntuación asignada según probabilidad de euploidía.

CHLOE™

Chloe es una herramienta de Fairtily™ para la ayuda en la toma de decisiones del laboratorio de FIV (Figura 9). De forma similar a los algoritmos nombrados anteriormente, Chloe proporciona

anotaciones automáticas del desarrollo embrionario y puntuaciones en función del potencial de implantación en cada día post inseminación. Según los desarrolladores mejora el flujo de trabajo y estandariza la evaluación embrionaria ya que elimina la entrada manual de datos, reduce el error humano, aumenta el número de ciclos por embriólogo y reduce el agotamiento del personal.

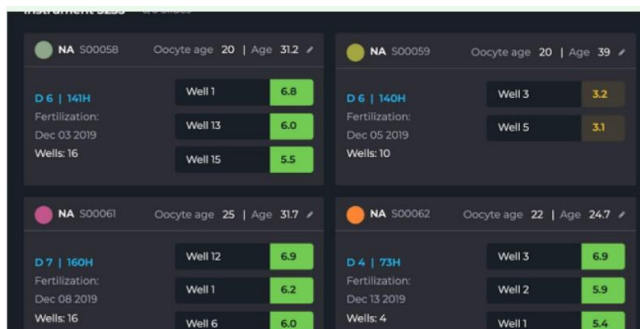


Figura 9. Demostración del software Chloe con una visión general de cuatro pacientes, su edad, edad de los ovocitos y puntuaciones para los embriones mejor evaluados de la cohorte.

No existe estudio clínico en revistas científicas que avale su uso, aunque se ha demostrado en una comunicación propia en el congreso de la Sociedad Europea de Reproducción Humana y Embriología (2022) una precisión en la predicción de blastocisto del 96% y en predicción de implantación del 73% (Yelke et al., 2022).

Embryo Assist™

La función de clasificación de Embryo Assist aprovecha el aprendizaje automático para ayudarlo a determinar qué embriones tienen la mayor probabilidad de éxito (Figura 10).

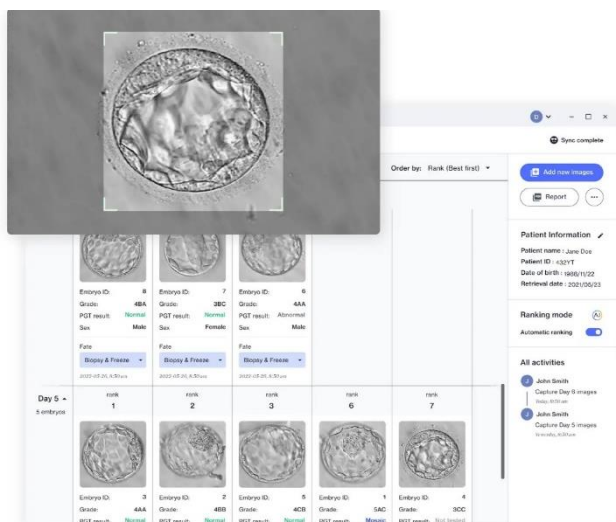


Figura 10. Presentación del software Embryo Assist con la clasificación embrionaria para una cohorte en día 5 de desarrollo.

El algoritmo utiliza un modelo complejo para clasificar los embriones disponibles y fue entrenado con datos históricos de 12.600 resultados de transferencias ocurridas entre 2014 y 2021, recopilados de cinco clínicas líderes en los Estados Unidos. Puede alternar entre clasificación manual y automatizada para activar o

desactivar la función de clasificación, o para comparar sus clasificaciones manuales con las del algoritmo.

Los mismos desarrolladores publicaron un estudio clínico validando el beneficio de la herramienta en la predicción de embarazo clínico (Loewke et al., 2022).

iDAScore™

El modelo iDAScore, desarrollado por Vitrolife en Aarhus, Dinamarca, es un modelo comercial de aprendizaje profundo diseñado para evaluar la calidad de los blastocistos (Figura 11).

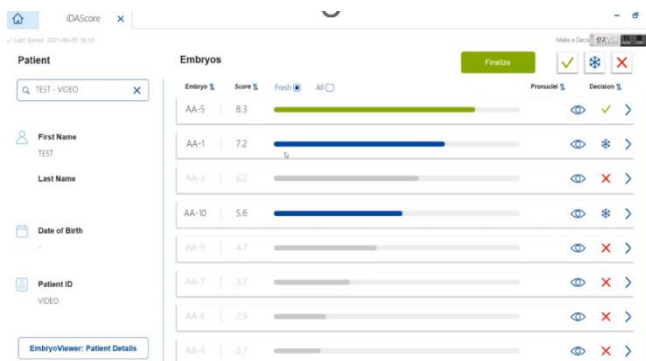


Figura 11. Demostración de la clasificación proporcionada por el algoritmo iDAScore v2 para una cohorte embrionaria.

Como humanos, no es posible conocer exactamente las variables específicas de estos modelos, ya que las técnicas de aprendizaje profundo no requieren la extracción previa de características. Sin embargo, se ha informado una correlación significativa entre

iDAScore versión 1.0 (v1) y características morfocinéticas y morfológicas de embriones preimplantatorios (Ezoe et al., 2022). Notablemente, factores como el número de células, la presencia de fragmentación de blastómeros y la morfología del blastocisto mostraron una asociación significativa con este sistema de puntuación automatizado. Además, se observó que los embriones con puntuaciones bajas presentaban una compactación, blastulación y expansión del blastocisto retrasadas. Además, varios estudios han demostrado una correlación significativa entre iDAScore v1 y diversos resultados reproductivos, incluidos la implantación, el embarazo en curso, el aborto espontáneo, el nacimiento vivo y las tasas de euploidía (Glatstein et al., 2023; Kato et al., 2023; S. Ueno et al., 2021; Zhan et al., 2020). Recientemente, este modelo ha sido actualizado a su segunda versión (iDAScore v2), que utiliza la misma arquitectura de aprendizaje profundo que la v1 pero incorpora un conjunto de datos de entrenamiento ampliado y la capacidad de proporcionar puntuaciones en los días 2/3 de desarrollo embrionario. Se basa en el uso de redes convolucionales 3D, lo que permite la identificación simultánea de patrones espaciales (morfológicos) y temporales (morfolinéticos) en datos de time-lapse (Theilgaard Lassen et al., 2023). Según los desarrolladores, este modelo se calibró para obtener una relación lineal entre las predicciones del modelo y las tasas de implantación. Se entrenó con un conjunto de datos extenso y diverso que incluye 181.428 embriones de 22 clínicas de FIV en todo el mundo. La puntuación automática se validó internamente en datos de prueba

independientes para embriones transferidos después de 2, 3 y 5/6 días de incubación (Theilgaard Lassen et al., 2023).

La presente tesis incluye la primera validación externa con una muestra considerable que examina la puntuación automática asignada por el algoritmo iDAScore v2 para clasificar blastocistos según su potencial de implantación y nacimiento vivo.

Se han descrito brevemente algunos de los modelos de selección embrionaria basados en IA disponibles; sin embargo, es importante destacar que existen otros comercialmente disponibles, con funcionalidades similares, que también utilizan imágenes o vídeos del desarrollo embrionario.

1.2. Justificación

La infertilidad es un desafío emocional y físico que enfrentan muchas parejas en todo el mundo (Cox et al., 2022). Para aquellos que luchan por concebir de manera natural, los tratamientos de FIV han surgido como una esperanza. Estos procedimientos ofrecen una oportunidad para superar diversos problemas de fertilidad, como obstrucciones en las trompas de Falopio, endometriosis, problemas de ovulación o espermatozoides de baja calidad. La FIV implica la unión de óvulos y espermatozoides en un laboratorio, seguido de la transferencia de embriones al útero de la mujer. Este proceso ha demostrado ser efectivo para muchas parejas, permitiéndoles realizar su deseo de tener un hijo. Además de ofrecer una opción para quienes luchan contra la infertilidad, la FIV también ha contribuido significativamente a la investigación médica y al entendimiento de la reproducción humana, abriendo nuevas puertas para el tratamiento y la prevención de la infertilidad en el futuro.

Se estima que más de 8 millones de bebés han nacido gracias a técnicas de reproducción asistida (TRA), con alrededor de 400.000 nacimientos al año, y esta cifra sigue en aumento.(Faddy et al., 2018) A pesar de los avances en tecnologías de reproducción asistida, la tasa de éxito de los tratamientos de FIV sigue siendo relativamente baja, alrededor del 30% por transferencia embrionaria (Wyns et al., 2017). La evaluación y selección de embriones son factores determinantes para el éxito de la FIV.

La comunidad científica en medicina reproductiva ha establecido criterios para evaluar la morfología embrionaria, junto con sistemas estandarizados para la clasificación y selección (Alpha Scientists, 2011). Sin embargo, la evaluación de la viabilidad del embrión todavía presenta desafíos debido a su naturaleza subjetiva e inconsistente (Martínez-Granados et al., 2017; Storr et al., 2017).

El uso de inteligencia artificial con redes neuronales artificiales ha ganado atención significativa en la investigación y las conferencias de fertilidad en todo el mundo (Curchoe & Bormann, 2019). Evaluar con precisión la viabilidad del embrión basada en la morfología es crucial para optimizar los resultados del tratamiento de FIV. La IA tiene un tremendo potencial para procesar conjuntos de datos complejos, detectar patrones sutiles pero cruciales y proporcionar una evaluación más objetiva de los embriones en comparación con los humanos. Sin embargo, se requiere un progreso significativo antes de implementar éticamente la IA para este propósito (Afnan et al., 2021). Además, los desafíos clave de la IA incluyen preocupaciones sobre la privacidad de los datos, la falta de creatividad humana y el potencial de desplazamiento laboral (Bori & Meseguer, 2021; Rolfes et al., 2023). Actualmente, la mayoría de los modelos de IA exhiben opacidad, lo que los hace difíciles de interpretar. Esto da lugar a varias preocupaciones epistémicas y éticas, incluidos problemas de confianza, posibles problemas de generalización entre diferentes poblaciones, implicaciones económicas adversas para las clínicas de FIV, y posibles desalineaciones con los valores del paciente o implicaciones más

amplias para la sociedad. Para abordar estas preocupaciones, el uso de modelos interpretables, diseñados para ser fácilmente comprensibles y explicables por los humanos, podría ser una solución.

En el transcurso de la presente tesis doctoral se diseñó una herramienta interpretable para automatizar la evaluación de embriones humanos (Objetivo específico 1).

Hoy en día, existe variación entre los modelos de IA en su enfoque hacia la automatización. Algunos métodos dependen de una preselección manual por parte de los embriólogos, lo que resulta en una categorización semiautomatizada. Estos métodos suelen entrenarse en embriones transferidos y, por lo tanto, tienen una exposición limitada a embriones de baja calidad (Bormann, Kanakasabapathy, et al., 2020; Miyagi, 2019; VerMilyea et al., 2020). Por otro lado, algunos métodos buscan la automatización completa al entrenar con todos los embriones, independientemente de si fueron transferidos o no (Loewke et al., 2022; Rimestad et al., 2020). Estos últimos son los más novedosos, aunque son popularmente conocidos como “cajas negras”, lo que implica que al menos algunos aspectos de sus procesos permanecen no interpretables (Afnan et al., 2021). En consecuencia, ha surgido el debate sobre si se requiere un fundamento biológico para respaldar los algoritmos generados por computadora en la selección de embriones (ESHRE Working group on time-lapse technology et al., 2020). En esta línea, el algoritmo iDAScore se ha relacionado con varias variables del

desarrollo embrionario (Berntsen et al., 2022a; Diakiw, Hall, VerMilyea, Lim, et al., 2022; Ezoe et al., 2022; Joo et al., 2023; Theilgaard Lassen et al., 2023) y está siendo ampliamente utilizado, a pesar de no existir una validación externa considerable.

El estudio incluido en esta tesis doctoral representa la primera validación externa con un tamaño de muestra sustancial que examina la puntuación automática asignada por el algoritmo iDAScore v2 para clasificar los blastocistos en función de su potencial de implantación y de nacidos vivos.

1.3. Objetivos

El objetivo principal de esta tesis es clasificar las categorías morfológicas de los blastocistos de acuerdo con el éxito de un tratamiento de fecundación in vitro (FIV), a través de modelos de inteligencia artificial (IA) basados en redes neuronales artificiales, y evaluar la precisión y consistencia de la IA en comparación con la clasificación de los embriólogos.

Para ello, se abordaron los siguientes objetivos específicos:

Objetivo específico 1. Diseñar una herramienta interpretable para automatizar la evaluación de embriones humanos mediante machine learning.

Objetivo específico 2. Validar externamente la puntuación automática asignada mediante deep learning para clasificar embriones en función de su potencial de implantación y nacimiento vivo.

Objetivo específico 3. Analizar el nivel de acuerdo entre la selección de embriones mediante inteligencia artificial y la decisión de los embriólogos.

1.4. Procedimientos intrínsecos de los tratamientos de FIV analizados

Estimulación ovárica y receptividad uterina

Las donantes fueron sometidas a estimulación ovárica controlada utilizando un protocolo de agonistas de la GnRH con Decapeptyl 1 (Ipsen Pharma, España) hasta que al menos ocho folículos alcanzaran un diámetro promedio de 18 mm, seguido de la recuperación transvaginal de ovocitos 36 horas después. Las receptoras recibieron preparación endometrial según lo descrito por Cerrillo et al. (2017), seguido de 400 mg de progesterona micronizada vaginal (Progeffik, Lab. Effik, Madrid, España) cada 12 horas después de la transferencia de embriones para el soporte de la fase lútea.

La estimulación ovárica utilizó protocolos de antagonistas de la GnRH, ocasionalmente omitiendo protocolos largos de agonistas de la GnRH o la supresión de la oleada de LH (Cozzolino et al., 2020). Las dosis de gonadotropina (FSH recombinante que varía de 150 a 300 UI/día) se ajustaron según las características de la paciente, a menudo combinadas con 75 UI de HMG altamente purificada (Menopur, Ferring Pharmaceuticals). El estrógeno se administró desde la fase lútea media en los ciclos de antagonistas de la GnRH. Los ciclos largos de agonistas de la GnRH implicaron la administración de 0,1 mg/día de GnRHa (Decapeptyl, Ipsen Pharma) a partir del día 21 del ciclo anterior.

Las ecografías transvaginales se realizaron en el día 5 de estimulación y posteriormente cada dos días. El antagonista de la GnRH (Orgalutran, MSD; o Cetrotide, Merck Serono) se introdujo cuando al menos un folículo alcanzó un diámetro promedio de 14 mm. La maduración final del ovocito se desencadenó utilizando 0,2 mg de GnRHa y/o 250 mcg de rhCG (Ovitrelle, Merck Serono) cuando al menos tres folículos alcanzaron un diámetro promedio de 17-18 mm en los ciclos de antagonistas de la GnRH. En los ciclos largos de agonistas de la GnRH, solo se administró rhCG.

Recuperación de ovocitos e inyección intracitoplasmática de espermatozoides (ICSI)

La aspiración folicular a través de punción ovárica transvaginal recuperó los ovocitos, que fueron lavados en medio de Gametos (Cook Medical, Australia) y cultivados en medio de fertilización Origio (Cooper Surgical, Dinamarca) al 6% de CO₂, 5% de O₂ y 37°C. La denudación de ovocitos, 4 horas después de la recuperación, involucró métodos mecánicos y químicos (pipeteado con 40 UI/ml de hialuronidasa). La inyección intracitoplasmática de espermatozoides (ICSI) se realizó utilizando medio de fertilización Origio a 400x de aumento con un microscopio Olympus IX7.

Después de la ICSI, los ovocitos se colocaron en EmbryoSlides pre-equilibrados (Vitrolife) hasta alcanzar el estadio de blastocisto. El volumen del medio de cultivo (Gems, Genea Biomedx, Sydney, Australia) fue de 28 µl para EmbryoScope convencional o 180 µl para

EmbryoScope Plus, con embriones cubiertos con 1.6 ml de aceite mineral durante el cultivo.

Incubación y evaluación de embriones

Los embriones fueron cultivados individualmente en el EmbryoScope convencional o en grupos de hasta ocho en el sistema EmbryoScope Plus hasta el quinto/sexta día de desarrollo. Se realizó una imagen automática cada 10-20 minutos, capturando imágenes de hasta 11 planos focales. Se utilizó software de análisis informático externo, específicamente la estación de trabajo EmbryoViewer™ de Vitrolife, para evaluar el desarrollo embrionario.

La evaluación de la fecundación tuvo lugar entre 16 y 19 horas después de la ICSI y fue confirmada por la presencia de dos pronúcleos y dos corpúsculos polares. Los tiempos de división a 2 células (t2), 3 células (t3), 4 células (t4), 5 células (t5), así como el tiempo de formación de blastocisto (tB), y la calidad de la masa celular interna (MCI) y el trofoectodermo fueron automáticamente anotados utilizando la herramienta de anotaciones guiadas dentro del EmbryoViewer, bajo la supervisión de los embriólogos senior. Simultáneamente, los blastocistos fueron evaluados por embriólogos senior a las 114-118 horas post-inseminación (Día 5) y 136-140 horas post-inseminación (Día 6). La evaluación incluyó el monitoreo de la expansión de la cavidad blastocística, así como la evaluación de la calidad de la masa celular interna y el trofoectodermo según los criterios establecidos por la Asociación para el Estudio de la

Biología de la Reproducción (ASEBIR) (Figuras suplementarias 1,2 y Tablas suplementarias 1 y 2).

Diagnóstico genético preimplantacional de aneuploidías (PGT-A)

En los ciclos que involucran PGT-A, los embriones fueron sometidos a eclosión asistida en el Día 3 después del recuento celular, utilizando el láser Hamilton-Thorne LykosVR. Al alcanzar la etapa de blastocisto, se biopsiaron 5-6 células del trofoectodermo, y su ploidía fue evaluada utilizando la tecnología de Secuenciación de Próxima Generación (NGS) de Thermo Fisher Scientific, EE. UU.

Vitrificación y desvitrificación embrionaria

La técnica de vitrificación y desvitrificación de los embriones se llevó a cabo siguiendo el protocolo de Kitazato (Japón), como se describe en la literatura (Cobo et al., 2010).. En resumen, los ovocitos y embriones fueron equilibrados a temperatura ambiente en una solución que contenía 7,5% de etilenglicol y 7,5% de dimetilsulfóxido (DMSO) en medio de cultivo de tejidos M 199 (TCM199) con un 20% de sustituto de suero sintético. Posteriormente, se transfirieron a una solución de vitrificación con una concentración diferente, que consistía en 15% de etilenglicol, 15% de DMSO y sacarosa al 0,5 M. Tras un minuto en esta solución, se colocaron en el dispositivo Cryotop con un volumen mínimo y se sumergieron directamente en nitrógeno líquido.

Para la desvitrificación, el Cryotop se retiró del nitrógeno líquido y se colocó en una solución compuesta por sacarosa al 1,0 M en medio TCM199 con un 20% de sustituto de suero sintético, a una temperatura de 37 °C. Después de un minuto, los embriones se transfirieron a otra solución que contenía sacarosa al 0,5 M en medio TCM199 con un 20% de sustituto de suero sintético, a temperatura ambiente durante tres minutos.

Finalmente, después de dos lavados de cinco y un minuto cada uno en medio TCM199 con un 20% de sustituto de suero sintético, los embriones se incubaron en medio Gems hasta la transferencia.

Transferencia de embriones y soporte lúteo

Los embriones mejor calificados de cada cohorte o los dos mejores embriones fueron seleccionados para la transferencia siguiendo las pautas de ASEBIR. Las transferencias de embriones congelados y descongelados posteriores se llevaron a cabo según la solicitud del paciente, según estuvieran disponibles. El soporte lúteo involucró dosis diarias de progesterona intravaginal micronizada (400 mg después de las transferencias frescas y 800 mg después de las transferencias congeladas y descongeladas) (Progeffik, Effik, España). Los embarazos bioquímicos fueron determinados mediante niveles de β hCG en sangre medidos 11 días después de la transferencia de embriones, mientras que la implantación se confirmó en la octava semana de embarazo mediante observación ecográfica del saco gestacional. Los nacimientos vivos fueron confirmados a través de comunicación directa con los pacientes.

Parte 2. Descripción de estudios

2.1. Objetivo específico 1. Diseñar una herramienta interpretable para automatizar la evaluación de embriones humanos mediante machine learning.

Propósito

El objetivo del estudio actual ha sido diseñar una herramienta interpretable para automatizar la evaluación de embriones humanos mediante la integración de redes neuronales artificiales (ANN) y algoritmos genéticos (GA).

Para justificar su uso, primero hemos comparado el acuerdo entre embriólogos en la evaluación de embriones. Luego, hemos comparado la correspondencia entre las evaluaciones de los embriólogos y las decisiones basadas en el aprendizaje automático.

Materiales y métodos específicos

Diseño del estudio

En este estudio de cohorte retrospectivo, se incluyeron 223 imágenes de blastocistos humanos (de 74 pacientes) obtenidas utilizando el Embryoscope® entre enero y junio de 2020. Todas las imágenes de microscopía de lapso temporal o time-lapse (TL) de embriones humanos capturadas a las 110 horas posteriores a la inyección (hpi) fueron evaluadas y clasificadas según su expansión, masa celular interna y trofoectodermo por cinco embriólogos

experimentados de tres países diferentes (con más de 5 años de experiencia). Luego, las imágenes se ordenaron aleatoriamente y se designaron para su uso como conjunto de datos de entrada para el sistema de inteligencia artificial, con un 70% asignado para entrenamiento, un 15% para validación y un 15% para pruebas o test.

Evaluación de la morfología del blastocisto y procesamiento de análisis de imágenes

Los blastocistos fueron evaluados rutinariamente y se les asignaron puntuaciones basadas en el grado de expansión, masa celular interna y apariencia del trofoectodermo, siguiendo el Sistema de Clasificación de Gardner (GG) (Gardner et al., 2000) ampliamente aceptado y adaptado (Figura suplementaria 3). Las imágenes del grupo fueron evaluadas de manera independiente por cinco embriólogos experimentados utilizando los criterios de GG. Luego, todas las imágenes de blastocistos fueron capturadas utilizando el EmbryoScope® y se utilizó el valor más frecuentemente asignado para la expansión, la masa celular interna y el trofoectodermo entre los cinco embriólogos (es decir, la moda) para entrenar la herramienta de IA.

En este estudio, el modelo de IA utilizó un algoritmo que importa automáticamente las imágenes de los blastocistos al software MatLab® y las estandariza. El proceso de estandarización incluyó la conversión de las imágenes a escala de grises, ajuste del ratio resolución y aspecto (T. Ueno et al., 2020) y conversión a escala de

grises de 8 bits (Figura 12A). Para garantizar la uniformidad para el análisis, las imágenes con diferentes resoluciones y proporciones se ajustaron en consecuencia. Además, se saturó el 1% de la información de píxeles entre píxeles claros y oscuros en cada imagen de blastocisto, lo que facilitó el paso de segmentación posterior (Hassaballah et al., 2019). Después de la estandarización, el blastocisto fue aislado de la imagen circundante modificando el gradiente de la imagen para definir sus límites mediante la Transformada de Hough (Vijayarajeswari et al., 2019) (Figura 12B).

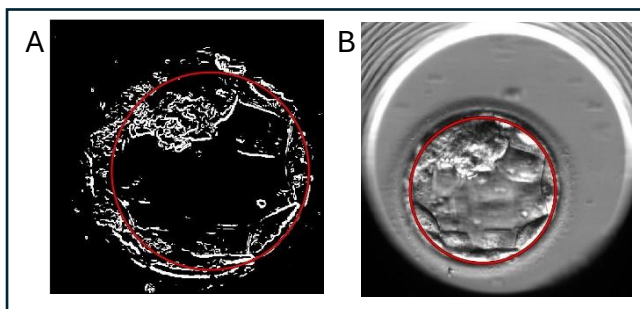


Figura 12. A) Imagen de un blastocisto que muestra la estandarización de una imagen de embrión humano. Para garantizar el contraste y la consistencia de la imagen, se satura el 1% de toda la información entre píxeles claros y oscuros. B) Blastocisto humano aislado con la función de Hough para determinar la circunferencia que mejor caracteriza al embrión.

El procesamiento de la imagen involucró múltiples algoritmos que operan de manera independiente. Se utilizó la Matriz de Co-

ocurrencia de Niveles de Gris (GLCM) para el análisis de texturas, la transformación de Watershed para la segmentación de imágenes, y el filtro de Gabor para diferenciar diversas texturas dentro de la imagen. Después de aplicar estas técnicas, se identificaron por separado las regiones del trofoectodermo y la masa celular interna (Figura 13). Tras el proceso de segmentación de la imagen, la información extraída proporcionó un vector numérico que describe las características clave de una imagen de blastocisto en cantidades adecuadas para aplicaciones digitales.

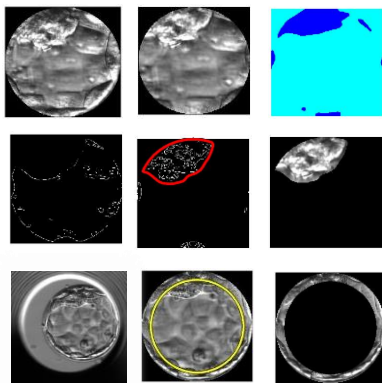


Figura 13. Secuencia de pasos necesarios para segmentar una imagen digital del blastocisto humano.

Análisis estadístico

Inteligencia artificial y algoritmos genéticos (AG): Después de analizar y procesar las imágenes, las variables cuantificadas de los

embriones fueron sometidas a la técnica de AG. Esta técnica implicaba la construcción de una población de individuos que representaban diferentes arquitecturas de ANNs. Las ANNs se generaron aleatoriamente para formar la población inicial, con 100 o 200 individuos en cada población. Cada individuo en la población representaba una configuración específica de ANN. Los parámetros que definían las ANN incluían el número máximo y mínimo de neuronas por capa, el número de capas, la tasa de aprendizaje, las funciones de transferencia (es decir, modelos que son una representación en el dominio de la frecuencia de sistemas lineales e invariantes en el tiempo) y las funciones de aprendizaje (es decir, funciones de entrenamiento que facilitan el proceso de entrenamiento o aprendizaje de las redes neuronales artificiales). El desarrollo de estas redes se llevó a cabo en la plataforma Matlab®.

Una vez que las ANN fueron generadas como individuos, toda la población fue sometida a entrenamiento, validación y pruebas utilizando la base de datos de imágenes de blastocistos. Los individuos más adecuados fueron seleccionados en función del menor error en el conjunto de prueba al aplicar la técnica de ANN. Una estructura esquemática del sistema de IA utilizado para la clasificación de embriones según el sistema de clasificación de Gardner se muestra en la Figura 14. Finalmente, el rendimiento de las predicciones fue comparado utilizando el área bajo la curva ROC (AUC).

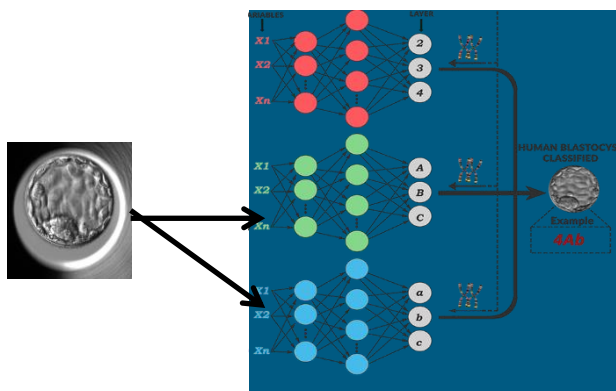


Figura 14. Secuencia de pasos necesarios para segmentar una imagen digital del blastocisto humano.

Coefficiente Kappa de Cohen: Para evaluar la concordancia entre observadores y dentro de un único observador, se empleó el coeficiente Kappa de Cohen (k). Los cálculos se realizaron utilizando el software estadístico SPSS (versión 24.0.0.1, IBM Corp). El coeficiente kappa es una medida estadística que tiene en cuenta el acuerdo por azar y cuantifica el nivel de concordancia entre dos conjuntos de datos.

La interpretación de los valores del coeficiente kappa es la siguiente: acuerdo absoluto ($k = 1$), acuerdo excelente ($k \geq 0,80$), acuerdo bueno ($k = 0,60-0,79$), acuerdo moderado ($k = 0,40-0,59$), acuerdo pobre ($k = 0,20-0,39$), acuerdo muy pobre ($k < 0,20$) y sin acuerdo ($k = 0$), según se sugiere. Para aplicaciones clínicas de laboratorio, se

definió una clasificación mínima aceptable como acuerdo excelente ($k \geq 0,80$).

Para evaluar la variabilidad entre observadores, se comparó la clasificación entre los embriólogos para cada parámetro de clasificación de los embriones (BE, MCI y TE). El acuerdo intraobservador se determinó evaluando la consistencia de cada embriólogo al calificar las imágenes. Cada embriólogo calificó todas las imágenes tres veces diferentes, y a partir de estas clasificaciones se calculó el valor de Kappa.

Resultados

Herramienta de IA para la evaluación de blastocistos.

La precisión general del sistema de IA en el análisis de las características clave del blastocisto fue aceptablemente alta. Los detalles se muestran en la Tabla 1.

Tabla 1. Precisión general de la herramienta de inteligencia artificial para la expansión del blastocisto, la calidad de la masa celular interna y el trofoblasto en los datos de prueba y entrenamiento.

Característica morfológica	Entrenamiento	Test
Expansión del blastocisto	93,9%	81,5%
Masa celular interna	93,0%	78,8%
Trofectoderma	78,8%	78,3%

El rendimiento de la herramienta de IA demostró ser más alto en la clasificación de la expansión del blastocisto (AUC de 0,923 a 1) en

comparación con la MCI (AUC de 0,605 a 0,854) o la calidad del trofoectodermo (AUC de 0,726 a 0,769) en el conjunto de datos de prueba (Tabla 2).

Tabla 2. Rendimiento del modelo con valores de AUC para la expansión del blastocisto, la masa celular interna (MCI) y la calidad del trofoectodermo (TE) en los datos de prueba y entrenamiento.

Grado de expansión	Entrenamiento	Test
Temprano	0,956	1
Completo	0,888	0,923
Expandido	0,946	1
MCI	Entrenamiento	Test
Grado A	0,979	0,854
Grado B	0,967	0,691
Grado C	0,961	0,605
Trofoectodermo	Entrenamiento	Test
Grado A	0,900	0,726
Grado B	0,863	0,733
Grado C	0,914	0,769

Grado de acuerdo entre los embriólogos

Cinco embriólogos de diferentes clínicas participaron en la clasificación de los blastocistos según su expansión, masa celular interna y trofoectodermo. Se observó una considerable variación en las calificaciones asignadas, lo que resultó en valores de acuerdo bajos medidos por el coeficiente Kappa (0,4 para la expansión y 0,3 para la MCI y el TE) (Tabla 3).

Tabla 3. Coeficientes Kappa de Cohen obtenidos de la evaluación realizada por cinco embriólogos de tres características de los blastocistos: grado de expansión, masa celular interna (MCI) y trofoectodermo. Los valores Kappa se expresan con intervalos de confianza (IC) del 95% entre paréntesis.

Grado de expansión	Temprano	Completo	Expandido	Media
<i>Kappa</i>	0,422 (0,464-0,381)	0,222 (0,181-0,264)	0,508 (0,466-0,549)	0,371 (0,342-0,400)
<i>P value</i>	<0,001	<0,001	<0,001	<0,001
MCI	Grado A	Grado B	Grado C	Media
<i>Kappa</i>	0,404 (0,363-0,446)	0,178 (0,137-0,220)	0,285 (0,243-0,326)	0,267 (0,236-0,298)
<i>P value</i>	<0,001	<0,001	<0,001	<0,001
Trofoectodermo	Grado A	Grado B	Grado C	Media
<i>Kappa</i>	0,353 (0,312-0,395)	0,209 (0,167-0,250)	0,376 (0,334-0,417)	0,299 (0,268-0,331)
<i>P value</i>	<0,001	<0,001	<0,001	<0,001

Grado de acuerdo entre el sistema automatizado y la moda de los embriólogos

Al comparar los resultados obtenidos por la técnica de IA con el valor modal (el valor calculado en función de la moda de los resultados proporcionados por los cinco embriólogos), se observó una mejora significativa en el acuerdo de clasificación de los blastocistos, como indican los valores de Kappa asignados (Tabla 4).

Tabla 4. Coeficientes de Kappa de Cohen de la evaluación realizada por inteligencia artificial comparada con la moda de 5 embriólogos de tres características de los blastocistos: grado de expansión, masa celular interna y trofoectodermo. Los valores de Kappa se expresan con intervalos de confianza (CI) del 95% entre corchetes.

Grado de expansión	Temprano	Completo	Expandido	Media
<i>Kappa</i>	0,751 (0,620-0,883)	0,652 (0,521-0,783)	0,790 (0,659-0,921)	0,729 (0,632-0,826)
<i>P value</i>	<0,001	<0,001	<0,001	<0,001
MCI	Grado A	Grado B	Grado C	Media
<i>Kappa</i>	0,779 (0,648-0,91)	0,688 (0,558-0,818)	0,681 (0,551-0,811)	0,705 (0,606-0,804)
<i>P value</i>	<0,001	<0,001	<0,001	<0,001
Trofoectodermo	Grado A	Grado B	Grado C	Media
<i>Kappa</i>	0,382 (0,256-0,507)	0,402 (0,273-0,531)	0,501 (0,370-0,631)	0,438 (0,332-0,544)
<i>P value</i>	<0,001	<0,001	<0,001	<0,001

Discusión

Se ha desarrollado una herramienta automatizada para la clasificación morfológica de blastocistos, utilizando información temporal de imágenes de lapso de tiempo y técnicas de aprendizaje automático. Este estudio se alinea con otros esfuerzos de investigación dirigidos a mejorar la calidad de la reproducción asistida mediante el aprovechamiento del potencial de la inteligencia artificial para ayudar a embriólogos y médicos.

En estudios iniciales sobre el uso de la IA en los laboratorios de FIV, Bormann et al. en 2020 aplicaron redes neuronales convolucionales para evaluar la consistencia en la clasificación de la calidad embrionaria y facilitar la selección para biopsia y criopreservación. También utilizaron algoritmos genéticos e imágenes de embriones individuales para lograr una precisión del 90% en la elección del embrión de mayor calidad disponible. Sus resultados fueron muy prometedores para distinguir entre blastocistos viables y no viables, y embriones con diferentes potenciales de implantación. Sin embargo, no analizaron la concordancia en la evaluación morfológica de blastocistos, donde nuestros resultados muestran un rendimiento superior por análisis de imagen (Table 4 vs. Tabla 3).

Varios autores emplearon técnicas de aprendizaje profundo para evaluar la calidad embrionaria y lograron resultados superiores en comparación con los embriólogos. LifeWhisperer realizó una comparación binaria de la clasificación de embriones viables/no viables y demostró una mejora del 24,7% sobre la precisión de los embriólogos (VerMilyea et al., 2021). Del mismo modo, STORK llevó a cabo una clasificación entre embriones buenos y pobres y comparó la predicción con el voto mayoritario de los embriólogos, logrando un coeficiente Kappa de 0,63 (Khosravi et al., 2019). Nuestro sistema similar obtuvo coeficientes más altos para dos características del blastocisto (expansión del blastocisto y masa celular interna), pero no para el trofoectodermo. Chavez-Badiola et al., exploraron la aplicación de la IA para determinar la viabilidad del embrión y predecir la ploidía, obteniendo resultados prometedores. Varios

estudios posteriores promueven la estandarización de la evaluación embrionaria (Cimadomo et al., 2023; Erlich et al., 2022; Kato et al., 2023; Loewke et al., 2022; Polsky et al., 2020; Rimestad et al., 2020; VerMilyea et al., 2020), y se han desarrollado nuevos sistemas para predecir el embarazo clínico mediante la combinación de imágenes de embriones con datos clínicos (Enatsu et al., 2022). Su sistema (FITTE) logró una mayor precisión que la evaluación del puntaje de Gardner (59,8% vs. 65,2%). Dado el gran número de modelos de IA que han surgido recientemente, es responsabilidad de los embriólogos evaluar cada uno de ellos lo mejor posible. Por ejemplo, Diakiw et al. en 2022 esbozaron métodos que deberían considerarse más ampliamente para evaluar algoritmos automatizados. Sin embargo, los métodos de inteligencia artificial utilizados para la selección de embriones a menudo se perciben como evaluaciones de caja negra. Para infundir confianza en las predicciones que generan, es crucial examinar y validar a fondo estos modelos, asegurando su consistencia con las características clínicas establecidas de la calidad embrionaria.

Nuestro estudio ha priorizado la evaluación objetiva y consistente de la morfología del blastocisto, en lugar de aspirar a crear un nuevo sistema automatizado de puntuación para la clasificación de embriones. Hasta donde sabemos, el sistema propuesto en este estudio es el único que proporciona resultados interpretables para cada parte del blastocisto, imitando la evaluación de un embriólogo pero obteniendo resultados más objetivos. Reconocemos el desafío de estandarizar la evaluación de la morfología embrionaria,

considerando su conexión crucial con los resultados clínicos posteriores. La búsqueda de objetividad en la evaluación embrionaria juega un papel crucial en el éxito y avance de la medicina reproductiva. De acuerdo con nuestros resultados, clínicas de diferentes países e incluso embriólogos dentro de la misma clínica informan de resultados variables (Cimadomo et al., 2022; Martínez-Granados et al., 2017; Storr et al., 2017; Wyns et al., 2020). Las discrepancias en la clasificación de embriones se originan en factores como las diferencias en la formación, variaciones en el control de calidad y numerosas otras posibilidades inherentes al trabajo humano. Según nuestros hallazgos, la IA puede abordar estas deficiencias y discrepancias que surgen durante la evaluación embrionaria. A través del análisis de imágenes en intervalos de tiempo, nuestros resultados han demostrado un rendimiento superior para la IA en comparación con los embriólogos en la clasificación de blastocistos humanos. Además, tal tecnología puede tener algunas ventajas: es rentable, no invasiva y proporciona una mayor consistencia en comparación con las evaluaciones realizadas por operadores individuales. La IA puede analizar miles de imágenes a un ritmo más rápido que los humanos, mientras continúa aprendiendo e integrando información adicional sobre los embriones.

La principal limitación de nuestro estudio es su naturaleza retrospectiva. Aunque la tecnología de IA se ha propuesto como una revolución en la evaluación de la morfología embrionaria, ofreciendo una mayor precisión y consistencia en la clasificación entre

individuos y clínicas, son necesarios más estudios independientes para validar y asegurar la reproducibilidad de estos hallazgos antes de implementarlos en entornos clínicos. Además, es probable que el bajo acuerdo entre operadores se deba al uso de imágenes 2D únicas con un plano focal central fijo, aunque el software EmbryoScope® garantizó una resolución e iluminación consistentes en todas las imágenes. También somos conscientes de que nuestro tamaño de muestra no es muy grande. Sin embargo, nuestra técnica de IA (ANN con GA) analizó miles de arquitecturas de ANN para encontrar la óptima. Típicamente, los modelos desarrollados con ANNs pueden ser susceptibles al sobreajuste. Sin embargo, en este caso, hicimos esfuerzos para mitigar este problema definiendo cuidadosamente las variables de entrada y salida. Además, la validación de nuestra herramienta debería implicar una comparación entre los resultados clínicos basados en la calidad embrionaria asignada por la IA y por el embriólogo. Esto no se llevó a cabo en este estudio ya que no todos los embriones fueron transferidos.

2.2. Objetivo específico 2. Validar externamente la puntuación automática asignada mediante deep learning para clasificar embriones en función de su potencial de implantación y nacimiento vivo.

Propósito

En este segundo objetivo, nuestro propósito ha sido evaluar la puntuación automática asignada por el algoritmo iDAScore como herramienta de apoyo en los laboratorios de FIV.

En primer lugar, hemos analizado la asociación entre la puntuación automática de los embriones y los resultados clínicos; y en segundo lugar, hemos evaluado la relevancia de esta puntuación en la tasa de implantación y recién nacido vivo.

Materiales y métodos específicos

Diseño del estudio

Este análisis retrospectivo abarcó un total de 70.456 embriones incubados en IVI Valencia (España) durante un período de cinco años, desde 2016 hasta 2021. El estudio se centró en pacientes y

receptores que se sometieron a ciclos de ICSI, con sus embriones incubados hasta el quinto o sexto día de desarrollo utilizando los sistemas de lapso temporal EmbryoScope y EmbryoScope Plus (Vitrolife, Aarhus, Dinamarca). Los embriones excluidos del análisis fueron aquellos que no pudieron ser evaluados por el algoritmo iDAScore v2.0 (por ejemplo, debido a la formación de burbujas o pérdida de grabación del desarrollo embrionario).

La tabla 5 ilustra la composición de la población de estudio. Entre la población total, se evaluaron 68.471 embriones mediante la puntuación automática, y 7.082 embriones fueron incluidos en el programa de pruebas genéticas preimplantacionales para aneuploidía (PGT-A). En última instancia, se obtuvieron datos de implantación conocidos (KID) de 10.054 embriones resultantes de 9.116 transferencias de blastocistos.

La Fase 1 consistió en un análisis de la asociación entre la puntuación automática de los embriones y los resultados clínicos. Esta sección reveló un modelo que indica las tasas de implantación esperadas correspondientes a diferentes puntuaciones automáticas.

La Fase 2 implicó evaluar la relevancia de la puntuación automática de los embriones para la tasa de implantación y el resultado del nacimiento vivo. Esta evaluación se llevó a cabo de manera individual y en comparación con la morfología convencional.

Tabla 5. Tabla demográfica de tratamientos dividida por subtipos. PGT-A: prueba genética preimplantacional para aneuploidía; ICSI: inyección intracitoplasmática de espermatozoides; IMC: índice de masa corporal; MII: metafase II; pn: pronúcleo; TE: transferencia embrionaria.

	No PGT-A		PGT-A (n=1.337)	TOTAL ciclos (n=7.352)
	Pacientes (n=2.700)	Ovodon (n=3.315)		
Edad ovocito*	36,9±4,4	25,3±4,7	35,9±5,7	31,5±7,4
IMC pacientes*	23,2±3,9	23,3±3,9	23,2±3,6	23,2±3,8
Ovocitos MII*	7,7±5,3	11,3±2,9	11,2±6,0	9,9±4,8
Ovocitos fecundados (2pn) *	5,5±4,3	8,7±2,7	8,5±4,8	7,5±4,1
Embriones puntuados	19.222	35.357	13.892	68.471
TE simple**	1.832	4.656	1.690	8.178
TE doble**	482	1.264	130	1.876
TE en D5**	1.988	5.226	1.459	8.673
TE en D6**	326	694	361	1.381
TE fresco**	957	2.597	-	3.554
TE en diferido**	1.357	3.323	1.820	6.500

*Media y desviación estándar por ciclo. **Número de embriones KID.

Análisis estadístico

Se analizaron las diferencias en las puntuaciones de los embriones dentro de cada grupo en función de diversos factores, como el grado morfológico, el estado de ploidía (euploide versus aneuploide), el resultado de la b-hCG (positivo versus negativo), el resultado de implantación (implantado versus no implantado), el resultado de

nacimiento vivo (positivo versus negativo) y el aborto espontáneo. Se llevó a cabo un análisis estadístico utilizando pruebas de ANOVA y chi-cuadrado para comparar las puntuaciones.

El rendimiento de la clasificación iDAScore para predecir la implantación y el nacimiento vivo se evaluó mediante la ecuación de estimación generalizada (GEE), midiendo la correlación entre el sistema de clasificación y el resultado al ajustar para posibles factores de confusión. Este enfoque trata a múltiples embriones de un mismo paciente como una variable intra-sujeto, controlando los factores relacionados con el paciente. Tras la evaluación de numerosos posibles factores de confusión, solo se incorporó la combinación que produjo el modelo óptimo en el análisis final. Finalmente, se creó un modelo GEE con la puntuación automática como variable predictiva y las siguientes variables como posibles factores de confusión: origen del óvulo (donado versus autólogo), tipo de transferencia de embriones (fresca versus congelada), edad del óvulo, IMC del paciente, día de la transferencia embrionaria (quinto versus sexto día de desarrollo), estrategia de cultivo (en grupo versus individual), morfología del blastocisto (grado A versus C y grado B versus C), e indicación para el tratamiento de la infertilidad (por ejemplo, etiología femenina, etiología masculina, etiología mixta, factor social o desconocido). Se construyeron modelos GEE adicionales de manera independiente para ciclos convencionales con óvulos autólogos en diferentes grupos de edad, el programa de donación de óvulos y ciclos de PGT-A. También se formularon GEEs comparables empleando la clasificación ASEBIR

como predictor en lugar de iDAScore, facilitando una evaluación comparativa de ambos sistemas de clasificación.

Las razones de probabilidad (OR, *del inglés odds ratio*) que indican el impacto de las variables incluidas en los GEE en las medidas de resultado se presentaron con intervalos de confianza del 95% (IC), y se determinó la significancia estadística para valores de $P < 0,05$. Para abordar los posibles efectos de variables no medidas, se proporcionaron los valores de E. Estos se calcularon utilizando la calculadora de valores de E en línea (Altman, 1991; Mathur et al., 2018), ofreciendo información sobre la fuerza mínima de asociación en la escala de riesgo que un factor de confusión no medido requeriría para explicar una asociación particular entre ciclos y resultados. Se generaron curvas ROC a partir de los valores de probabilidad derivados de los modelos GEE, y las áreas bajo las curvas ROC (AUC) se calcularon como métricas de evaluación del rendimiento del modelo.

Resultados

La puntuación más alta observada en la población de estudio fue de 9,9, mientras que la puntuación más baja fue de 1. Se observó un rango significativo entre las puntuaciones máxima (8,9 puntos) y mínima (0 puntos) para todos los pacientes, con una diferencia media de $5,2 \pm 3,0$.

Del total de la población, 10.054 embriones fueron datos de implantación conocidos (KID), incluyendo tanto transferencias

frescas como congeladas. Los resultados clínicos subsiguientes fueron los siguientes: una tasa de embarazo bioquímico del 59,5%, una tasa de implantación del 50,5% y una tasa de nacimientos vivos del 42,3%.

Fase 1: Asociación entre la puntuación automática de los embriones y los resultados clínicos.

La puntuación media de los embriones varió significativamente ($p < 0,001$) entre diferentes categorías morfológicas según los criterios ASEBIR (Figura 15).

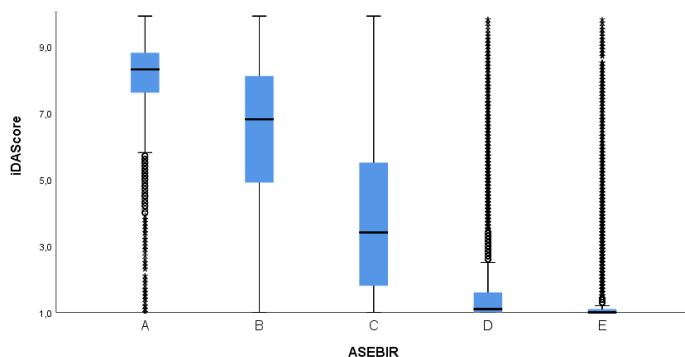


Figura 15. Distribución de la puntuación automática según la morfología convencional de acuerdo con los criterios ASEBIR.

Los embriones clasificados como A tuvieron una puntuación media de $8,0 \pm 1,4$ ($n=2.775$), los embriones de grado B tuvieron una puntuación media de $6,4 \pm 2,2$ ($n=15.523$), los embriones de grado C tuvieron una puntuación media de $3,8 \pm 2,3$ ($n=11.047$), los embriones

de grado D tuvieron una puntuación media de $1,7 \pm 1,3$ ($n=12.778$) y los embriones excluidos tuvieron una puntuación media de $1,2 \pm 0,7$ ($n=26.348$). Además, hubo una diferencia significativa en la puntuación media entre los embriones euploides ($6,3 \pm 2,5$, $n=3.208$) y los embriones aneuploides ($5,1 \pm 2,5$, $n=3.874$) (Figura 16).

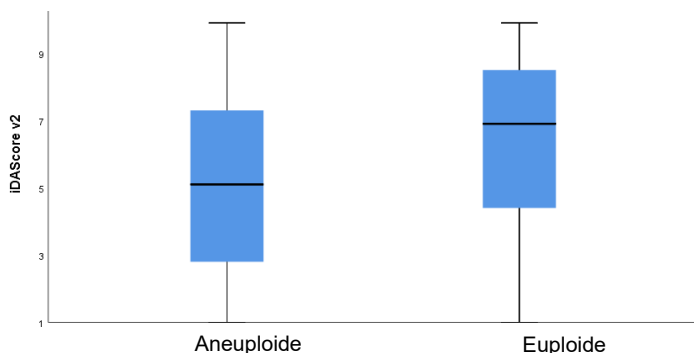


Figura 16. Distribución de la puntuación automática según la ploidía.

Los embriones que resultaron en una beta-hCG positiva tuvieron una puntuación media de $6,7 \pm 2,2$, mientras que aquellos que resultaron en una beta-hCG negativa tuvieron una puntuación media de $5,7 \pm 2,5$ ($P < 0,001$). De manera similar, los embriones que llevaron a una implantación positiva tuvieron una puntuación media de $6,8 \pm 2,2$, mientras que aquellos que resultaron en una implantación negativa tuvieron una puntuación media de $5,7 \pm 2,5$ ($P < 0,001$). Además, los embriones que resultaron en un nacimiento vivo positivo tuvieron una puntuación media de $6,9 \pm 2,2$, mientras que aquellos que resultaron

en un nacimiento vivo negativo tuvieron una puntuación media de $5,9 \pm 2,5$ ($P < 0,001$). Además, los embriones que resultaron en aborto clínico tuvieron una puntuación automática media más baja que aquellos que alcanzaron el parto ($6,6 \pm 2,3$ vs. $6,9 \pm 2,2$; $p < 0,001$).

Cuando los embriones se dividieron en cuartiles de tamaños de muestra similares utilizando el software SPSS, se observó una clara tendencia significativa ($P < 0,001$) hacia mejores resultados clínicos a medida que aumentaba la puntuación de los embriones (Figura 17).

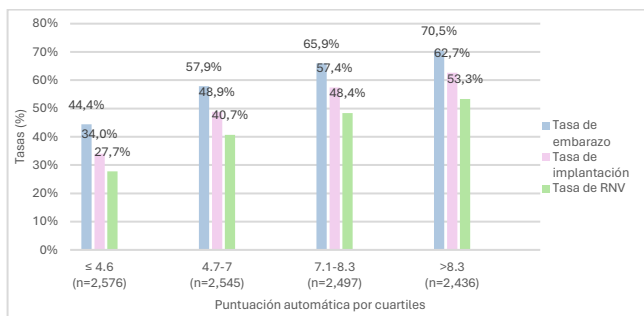


Figura 17. Tasa de embarazo, implantación y recién nacido vivo (RNV) según la puntuación automática (iDAScore v2) por cuartiles.

El software estadístico definió siete rangos de puntuación como categorías de calidad de embriones para las transferencias embrionarias en el quinto día de desarrollo. Luego se calcularon y compararon retrospectivamente las tasas de implantación alcanzadas para cada categoría en un total de 6.543 tratamientos

sin PGT-A, lo que resultó en el modelo de implantación esperada que se muestra en la Tabla 6.

Tabla 6. Tasa de implantación esperada según el origen de los ovocitos, el tipo de transferencia embrionaria (TE) y la puntuación automática asignada por iDAScore v2.

iDAScore v2	Tasa de implantación % [IC 95%]			
	TE fresco		TE congelado	
	Ovocitos propios	Ovocitos donados	Ovocitos propios	Ovocitos donados
≤ 3,5	22 [15,5-28,4]	43,4 [35,8-51]	28,5 [21,4-35,6]	33,8 [28,5-39]
3,5-5,1	27,3 [19,5-35,2]	50,2 [43,7-56,8]	37,5 [29,7-45,3]	37,1 [32,1-42,1]
5,1-6,8	48,3 [40,7-55,8]	59 [54-63,9]	50,1 [44,3-57,4]	44,4 [40,4-48,3]
6,8-7,6	54,4 [45,9-62,9]	61,3 [56,1-66,4]	59,4 [51,3-67,6]	53,7 [48,6-58,8]
7,6-8,4	56,4 [48,1-64,7]	66,1 [62,5-69,7]	59,8 [52,8-66,8]	52,7 [48,5-56,9]
8,4-9	66,1 [57,3-74,9]	67,6 [63,5-71,7]	64,9 [57,1-72,6]	56,8 [51,9-61,8]
≥ 9	72,3 [59,1-85,6]	68,2 [61,7-74,7]	75,4 [64,9-85,8]	61 [53,4-68,5]

Fase 2: Relevancia de la puntuación automática de los embriones para la tasa de implantación y el resultado del nacimiento vivo.

La distribución de los embriones puntuados automáticamente según el éxito del tratamiento (implantación y nacimiento vivo) se puede visualizar en la Figura 18.

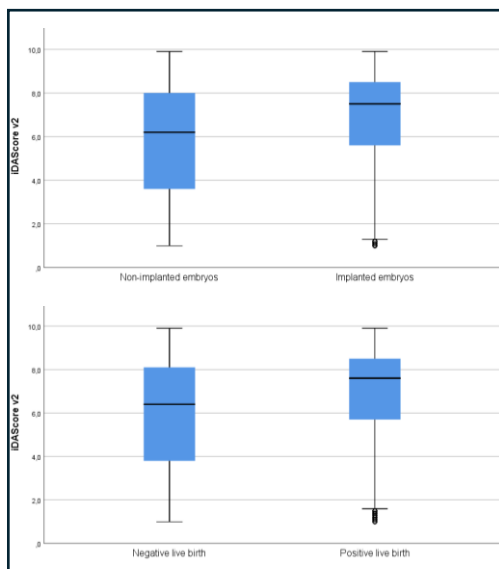


Figura 18. Distribución de la puntuación automática de embriones según dos resultados clínicos. A) implantación y B) recién nacido vivo.

La capacidad predictiva de iDAScore fue confirmada por un modelo GEE, que produjo un OR de 1,18 para la implantación y 1,17 para el nacimiento vivo en toda la población. La puntuación del embrión contribuyó significativamente al resultado de la implantación en tratamientos convencionales y programas de donación de óvulos, pero no en ciclos PGT-A (Tabla 7). Sin embargo, sí contribuyó a la predicción del nacimiento vivo en todas las subpoblaciones por separado, como se presenta en la Tabla 8.

Se construyeron modelos GEE predictivos adicionales tanto para la implantación como para los resultados del nacimiento vivo, utilizando la clasificación morfológica convencional como variable predictiva en lugar de iDAScore. Se incluyó el mismo conjunto de factores de confusión empleados en los modelos GEE anteriores (Tabla 9). Las AUC fueron mayores cuando se consideró iDAScore v2 (en lugar de morfología) para pacientes de diferentes grupos de edad y para el programa de donación de óvulos, pero no para los tratamientos PGT-A.

Tabla 7. Ecuaciones de estimación generalizada (GEE) que evalúan la correlación entre iDAScore y los resultados de implantación controlando otros posibles factores de confusión. La tabla muestra únicamente aquellas variables con significancia estadística ($p < 0,05$) en la asociación entre variables y resultados.

	Implantación			
	OR ^a	IC 95%	P valor	Valor E ^b
Ovocitos propios (n=2.314)				
Puntuación automática	1,31	[1,25-1,36]	<0,001	1,55
Edad ovocito	0,94	[0,91-0,96]	<0,001	1,21
Programa donación de ovocitos (n=5.920)				
Puntuación automática	1,17	[1,14-1,20]	<0,001	1,38
TE Fresco (43,9%) vs. Diferido (56,1%)	0,63	[0,56-0,71]	<0,001	1,83
Estrategia de cultivo Individual (62%) vs. Grupo (38%)	0,82	[0,72-0,93]	0,005	1,44

^aOrs calculados a partir de GEE.

^bMínima fuerza de asociación, en la escala de razón de riesgos, que un factor de confusión no medido necesitaría tener tanto con el tratamiento como con el resultado para explicar completamente la asociación entre la variable y el resultado.

Tabla 8. Ecuaciones de estimación generalizada (GEE) que evalúan la correlación entre iDAScore y los resultados de nacimiento vivo, controlando otros posibles factores de confusión. La tabla muestra únicamente aquellas variables con significancia estadística ($p < 0,05$) en la asociación entre variables y resultados.

	RNV			
	ORa	IC 95%	P valor	Valor E b
Ovocitos propios (n=2.314)				
Puntuación automática	1,27	[1,21-1,33]	<0,001	1,51
Edad ovocito	0,93	[0,90-0,96]	<0,001	1,23
Programa donación de ovocitos (n=5.920)				
Puntuación automática	1,16	[1,13-1,19]	<0,001	1,37
TE Fresco (43.9%) vs. Diferido (56.1%)	0,66	[0,58-0,74]	<0,001	1,76
Estrategia de cultivo Individual (62%) vs. Grupo (38%)	0,83	[0,73-0,94]	0,003	1,43
PGT-A (n=1.820)				
Puntuación automática	1,05	[1-1,10]	0,044	1,18
IMC	0,97	[0,94-1]	0,024	1,14

^aORs calculados a partir de GEE. ^bMínima fuerza de asociación, en la escala de razón de riesgos, que un factor de confusión no medido necesitaría tener

tanto con el tratamiento como con el resultado para explicar completamente la asociación entre la variable y el resultado.

Tabla 9. Métricas de rendimiento para las ecuaciones de estimación generalizada (GEEs) de los dos sistemas de clasificación, iDAScore y evaluación morfológica, para la predicción de implantación (a) y nacimientos vivos (b). AUC: Área bajo la curva. OR: Odds ratio.

a) IMPLANTACIÓN		ASEBIR		iDAScore v2	
		OR	AUC	OR	AUC
OVOCITOS PACIENTE	<35 (n=1.117)	3,7 [2,4-5,8] (A vs. C)	0,654 (0,623-0,686)	1,3 [1,2-1,4]	0,692 (0,661-0,722)
		2,1 [1,4-3,0] (B vs. C)			
	35-37 (n=877)	5,5 [3,2-9,5] (A vs. C)	0,657 (0,621-0,693)	1,3 [1,2-1,4]	0,673 (0,637-0,708)
		2,3 [1,5-3,4] (B vs. C)			
	38-40 (n=236)	6,5 [2,3-18,1] (A vs. C)	0,697 (0,628-0,767)	1,3 [1,1-1,5]	0,714 (0,646-0,783)
		2,4 [1,1-5,2] (B vs. C)			
	>40 (n=84)	NS	-	1,5 [1,1-2,1]	0,772 (0,654-0,890)
	OVODON (n=5.920)	2,6 [2,2-3,2] (A vs. C)	0,641 (0,627-0,655)	1,2 [1,1-1,2]	0,646 (0,632-0,660)
		1,7 [1,4-2,0] (B vs. C)			
	PGTA (n=1.820)	2,0 [1,2-3,4] (A vs. C)	0,574 (0,547-0,600)	NS	-
		1,7 [1,4-2,1] (B vs. C)			

b) RNV		ASEBIR		iDAScore v2	
		OR	AUC	OR	AUC
OVOCITOS PACIENTE	<35 (n=1.117)	3,2 [2,1-5,0] (A vs. C)	0,657 (0,625-0,689)	1,3 [1,2-1,3]	0,677 (0,646-0,708)
		1,7 [1,2-2,6] (B vs. C)			
	35-37 (n=877)	5,3 [3,1-9,1] (A vs. C)	0,652 (0,615-0,689)	1,2 [1,2-1,3]	0,654 (0,617-0,691)
		2,2 [1,4-3,4] (B vs. C)			
	38-40 (n=236)	6,6 [2,1-21,1] (A vs. C)	0,707 (0,630-0,784)	1,4 [1,2-1,6]	0,722 (0,650-0,794)
		2,4 [1,0-5,7] (B vs. C)			
	>40 (n=84)	35,7 [2,3-546,7] (A vs. C)	0,817 (0,698-0,936)	2,1 [1,5-3,1]	0,894 (0,824-0,963)
		8,2 [1,4-48,7] (B vs. C)			
	OVODON (n=5.920)	2,4 [2,0-3,0] (A vs. C)	0,629 (0,615-0,643)	1,2 [1,1-1,2]	0,636 (0,622-0,650)
		1,7 [1,4-2,0] (B vs. C)			
	PGTA (n=1.820)	1,9 [1,1-3,3] (A vs. C)	0,586 (0,560-0,612)	1,1 [1,0-1,1]	0,558 (0,531-0,584)
		1,8 [1,5-2,3] (B vs. C)			

Discusión

Según nuestra validación externa y única con una muestra de gran tamaño, se espera que la transferencia de embriones con las puntuaciones más altas de la cohorte mejore las tasas de implantación en comparación con la selección tradicional basada en la morfología convencional.

Indudablemente, la inteligencia artificial muestra la capacidad de procesar conjuntos de datos complejos, discernir patrones sutiles pero cruciales, y mantener un mayor nivel de objetividad en comparación con la evaluación humana en el contexto de la evaluación de embriones. Sin embargo, muchos modelos de IA a menudo son considerados como 'cajas negras', lo que implica que al menos algunos aspectos de sus procesos permanecen ininterpretables (Afnan et al., 2021). En consecuencia, ha surgido el debate sobre si una justificación biológica es un requisito para respaldar los algoritmos generados por computadora en la selección de embriones (ESHRE Working group on time-lapse technology et al., 2020). En esta línea, el algoritmo iDAScore ha sido vinculado a diversas variables de desarrollo embrionario (Berntsen et al., 2022b; Diakiw, Hall, VerMilyea, Lim, et al., 2022; Ezoe et al., 2022; Joo et al., 2023; Theilgaard Lassen et al., 2023). En alineación con esto, nuestros hallazgos indican que la puntuación asignada automáticamente por un modelo basado en aprendizaje profundo correlaciona directamente con la calidad morfológica convencional determinada por los embriólogos.

La comunidad científica ya está al tanto de que la puntuación automatizada asignada por varios algoritmos está vinculada a resultados clínicos (Ahlström et al., 2023; Salih et al., 2023; Sarandi et al., 2023; S. Ueno et al., 2021, 2022; Zhu et al., 2023). Hasta donde sabemos, solo Ueno 2023 ha evaluado la utilidad del iDAScore v2 para transferencias en estadio de blastocisto, reportando un área bajo la curva (AUC) de 0,736 para el embarazo en curso. Sin embargo, la generalizabilidad de sus hallazgos es cuestionable, ya que la población de pacientes consistió en aquellos sometidos a estimulación mínima y ciclos de FIV naturales. Además, la clínica participó en el desarrollo del algoritmo. Aunque nuestros resultados se alinean con los suyos, las AUC fueron más bajas en casi todos los escenarios de tratamiento, posiblemente debido a que nuestros datos se consideraron una validación externa completa. Además, realizar comparaciones resulta 'desafiante' sin implementar ajustes, como la estandarización, para mitigar las discrepancias relacionadas con la edad (Johansen et al., 2023).

Según nuestros hallazgos, un aumento de una unidad en la puntuación del embrión, independientemente de otras variables clínicas analizadas, conduce a mejores resultados clínicos. El efecto más significativo en los tratamientos con óvulos de la paciente ha sido reportado previamente para otros modelos basados en aprendizaje automático con parámetros morfocinéticos (Bori et al., 2022). Además, nuestros hallazgos actuales muestran que el rendimiento de los modelos GEE con puntuación automática mejoró en pacientes de mayor edad. Esta mejora puede atribuirse a que los

embriones euploides tienen puntuaciones más altas que los aneuploides. Del mismo modo, otros modelos basados en IA también han indicado puntuaciones más altas para embriones cromosómicamente normales (Barnes et al., 2023; Cimadomo et al., 2023; Diakiw et al., 2022b). Por lo tanto, en tratamientos donde hay una mayor probabilidad de embriones cromosómicamente anormales, el algoritmo mostró un mejor rendimiento.

La discrepancia en el rendimiento del modelo nos motivó a crear una plantilla de implantación esperada. La implantación esperada para embriones con la misma puntuación variaba según el tratamiento. Los embriones con puntuaciones idénticas mostraron un mayor potencial de implantación en programas de donación de óvulos que en tratamientos con óvulos propios después de la transferencia de embriones frescos (TE), especialmente en categorías de puntuación más bajas ($p < 0,05$ hasta una puntuación de 6,80). En la TE congelada, no hubo diferencias significativas entre óvulos propios y donados, excepto para embriones con la puntuación máxima. Solo se observaron diferencias entre óvulos de la paciente y donados en la TE fresca, lo que resalta el impacto potencial de la estimulación ovárica en la receptividad endometrial. Los embriones con la misma puntuación mostraron un mayor potencial de implantación para la TE fresca que para la TE congelada en el programa de donación de óvulos ($p < 0,05$ hasta puntuaciones ≥ 9). Sin embargo, la tendencia fue la opuesta para los óvulos propios, aunque en general no fue estadísticamente significativa. En los tratamientos con óvulos propios, este efecto podría estar enmascarado por la estimulación

ovárica, lo que podría reducir la implantación en las transferencias frescas.

En nuestra base de datos, la utilización del iDAScore para la selección de embriones produjo resultados superiores en comparación con la morfología convencional en todas las subpoblaciones excepto en tratamientos PGT-A. Una justificación potencial podría ser la eclosión asistida a la que fueron sometidos todos los embriones en el día tres de desarrollo, lo que afecta el desarrollo embrionario en etapas tardías (Cimadomo et al., 2018) y, consecuentemente, el análisis de imágenes. En este escenario, los embriólogos eran conscientes del impacto del procedimiento, mientras que el algoritmo, no entrenado en embriones sometidos a este procedimiento, no lo consideró en la evaluación.

Una limitación significativa de nuestro estudio radica en su naturaleza retrospectiva. Para respaldar la importancia clínica del iDAScore™ v2, sería esencial realizar un estudio prospectivo aleatorizado. A pesar de ser una validación externa exhaustiva que involucra una población de ICSI no seleccionada, la aplicabilidad del modelo a una escala más amplia debe abordarse con cautela debido a su diseño de centro único. Además, la aplicación de un modelo de predicción a mayor escala debería considerar condiciones de cultivo específicas. Es importante tener en cuenta que la comparación entre sistemas de clasificación automatizados y convencionales podría introducir sesgos, dado el resultado continuo del valor de IDAScore v2 versus el resultado factorial de ASEBIR. Además, las tasas de

implantación y nacimientos vivos están influenciadas por diversos factores, que abarcan no solo las variables consideradas en la regresión logística realizada en este estudio, sino también numerosos factores clínicos. Se requiere una investigación adicional para mejorar nuestra comprensión de estas limitaciones y mejorar la validez general del estudio.

2.3. Objetivo específico 3. Analizar el nivel de acuerdo entre la selección de embriones mediante inteligencia artificial y la decisión de los embriólogos.

Propósito

Nuestro objetivo fue analizar la concordancia y discordancia entre embriólogos senior y la puntuación automática asignada por varios modelos de inteligencia artificial (IA) para cada cohorte de embriones.

En primer lugar, se procedió al análisis del nivel de concordancia con el KIDScore D5 v3, seguido por la evaluación del iDAScore v1 y, finalmente, del iDAScore v2.

Material y métodos específicos

Diseño del estudio

Se analizaron un total de 7.742 transferencias de embriones de 4.510 pacientes que se sometieron a tratamientos de FIV en un único centro. Los embriones se cultivaron en sistemas de lapso de tiempo EmbryoScope® (Vitrolife, Dinamarca) y fueron evaluados rutinariamente por embriólogos senior según los criterios morfológicos ASEBIR. La selección de embriones fue realizada por los embriólogos y los embriones fueron puntuados retrospectivamente con modelos de IA de 1 a 9,9 (KIDScore D5 v3, iDAScore v1 e iDAScore v2). A continuación, se realizó un análisis

de cohortes sobre la concordancia entre el embrión seleccionado por el embriólogo y el embrión seleccionado por cada modelo de IA. Por último, se calculó el riesgo relativo (RR o incidencia de implantación) para todos los modelos.

Análisis estadístico

Se midió la concordancia entre la decisión del embriólogo y la decisión del modelo (es decir, el embrión con la puntuación más alta según cada modelo) en la primera transferencia mediante un simple porcentaje de acuerdo general, calculado como el número de casos en los que ambos evaluadores estuvieron de acuerdo dividido por el número total de casos considerados. Además, se calculó el coeficiente Kappa de Cohen para considerar el efecto del azar en este acuerdo. Los valores de Kappa se interpretaron de la siguiente manera: muy bueno (0,81–1,00), bueno (0,61–0,80), moderado (0,41–0,60), regular (0,21–0,40) y malo ($\leq 0,20$). Finalmente, se calculó el riesgo relativo (o incidencia de implantación) como la razón entre la probabilidad de que ocurra un evento con una exposición (transferencia del embrión con la puntuación más alta según cada modelo) y la probabilidad de que ocurra el evento sin la exposición.

Resultados

KIDScore D5 v3

Los embriones seleccionados por embriólogos senior para transferencia de un solo embrión en el ciclo fresco coincidieron con el embrión de mayor puntuación en el día 5 de desarrollo el 74,96% de las veces para KIDScore D5 v3 (n=2.133).

El coeficiente Kappa de Cohen fue de 0,75. La diferencia media entre el KIDScore del embrión seleccionado y el embrión de mayor puntuación para los casos que no coincidían fue de $0,68 \pm 0,63$. La concordancia fue mayor para los tratamientos convencionales con óvulos de la paciente en comparación con el programa de donación de óvulos (82,12% vs. 72,39%). En general, considerando las transferencias de un solo embrión fresco/congelado en el día 5 ($n=7.440$), la tasa de implantación fue mayor cuando el embrión transferido coincidía con el embrión de mayor puntuación: 58,22% coincidencia vs. 49,33% no coincidencia ($p<0,001$). El RR mostró que cada paciente tenía 1,18 veces más probabilidades de quedar embarazada si el embrión transferido coincidía con el embrión con la puntuación más alta asignada por KIDScore D5 v3 (Tabla 10).

iDAScore v1

Los embriones seleccionados por embriólogos senior para transferencia de un solo embrión en el ciclo fresco coincidieron con el embrión de mayor puntuación en el día 5 de desarrollo el 67,14% para iDAScore v1 ($n=2.170$).

El coeficiente Kappa de Cohen fue de 0,65. La diferencia media entre el iDAScore v1 del embrión seleccionado y el embrión de mayor puntuación para los casos que no coincidían fue de $0,67 \pm 0,90$. La concordancia fue mayor para los tratamientos convencionales con óvulos de la paciente en comparación con el programa de donación de óvulos (73,85% vs. 64,64%). En general, considerando las transferencias de un solo embrión fresco/congelado en el día 5, la

tasa de implantación fue mayor cuando el embrión transferido coincidía con el embrión de mayor puntuación: 56,38% coincidencia vs. 50,42% no coincidencia ($p<0,001$). El RR mostró que cada paciente tenía 1,12 veces más probabilidades de quedar embarazada si el embrión transferido coincidía con el embrión con la puntuación más alta asignada por iDAScore v1 (Tabla 10).

iDAScore v2

Los embriones seleccionados por embriólogos senior para transferencia de un solo embrión en el ciclo fresco coincidieron con el embrión de mayor puntuación en el día 5 de desarrollo el 63,8% del tiempo ($n=2.675$) para iDAScore v2.

El coeficiente Kappa de Cohen fue de 0.63. La diferencia media entre el iDAScore del embrión seleccionado y el embrión de mayor puntuación para los casos que no coincidían fue de $1,23\pm1,15$. La concordancia fue mayor para los tratamientos convencionales con óvulos de la paciente en comparación con el programa de donación de óvulos (71,38% vs. 60,97%). En general, considerando las transferencias de un solo embrión fresco/congelado en el día 5 ($n=7.126$), la tasa de implantación fue mayor cuando el embrión transferido coincidía con el embrión de mayor puntuación: 57,36% coincidencia vs. 49,98% no coincidencia ($p<0,001$). El RR mostró que cada paciente tenía 1,14 veces más probabilidades de quedar embarazada si el embrión transferido coincidía con el embrión con la puntuación más alta asignada por iDAScore v2 (Tabla 10).

Tabla 10. Número de embriones transferidos (en concordancia y discordancia entre los algoritmos de IA y los embriólogos senior) e incidencia de implantación.

		Embriones implantados	Embriones no implantados	n TOTAL	Riesgo relativo (RR)
KIDScore D5 v3	Concordancia	1.959	1.406	3.365	1,18
	Discordancia	2.010	2.065	4.075	
iDAScore v1	Concordancia	1.868	1.445	3.313	1,12
	Discordancia	2.090	2.055	4.145	
iDAScore v2	Concordancia	1.774	1.319	3.093	1,14
	Discordancia	2.177	2.179	4.356	

Discusión

El acuerdo entre el embriólogo y la IA fue bastante alto en tratamientos frescos, con una tasa de acuerdo entre 64-75%. Aunque hay evidencia de que no todos los modelos de selección de embriones basados en IA tienen el mismo nivel de acuerdo con los embriólogos (Zaninovic et al., 2023), observamos un buen grado de acuerdo. Este grado de acuerdo fue más alto en tratamientos convencionales, probablemente porque las cohortes de embriones eran más pequeñas y había menos blastocistos de alta calidad para seleccionar en comparación con las cohortes con óvulos donados. A favor de la IA, cabe destacar que cuando había acuerdo entre el embriólogo y el modelo, los resultados del tratamiento eran mejores.

Sin embargo, no pudimos cuantificar la mejora potencial si el embrión seleccionado por la IA hubiera sido transferido, dada la naturaleza retrospectiva del estudio. Si bien el riesgo relativo se usa predominantemente en estudios prospectivos, nuestro objetivo era calcularlo como una medida de la incidencia de implantación. Es decir, la odds ratio de implantación cuando había acuerdo entre el embriólogo y la IA fue siempre mayor que 1.

Parte 3. Epílogo

3.1. Consecución de objetivos y relevancia científica

La presente tesis doctoral ha abordado múltiples interrogantes planteados por la comunidad científica acerca del uso emergente de la inteligencia artificial en la evaluación y selección de embriones humanos en tratamientos de fecundación in vitro.

El propósito primordial de esta investigación doctoral consistía en clasificar las categorías morfológicas de los blastocistos de acuerdo con el éxito de un tratamiento de FIV, a través de modelos de IA basados en redes neuronales artificiales, y evaluar la precisión y consistencia de la IA en comparación con la clasificación de los embriólogos.

Con este propósito, se definieron tres objetivos específicos para investigar la aplicación de IA mediante diversas aproximaciones.

A continuación, se detalla y describe el logro de los diferentes objetivos del estudio:

Objetivo específico 1. Diseñar una herramienta interpretable para automatizar la evaluación de embriones humanos mediante machine learning: por un lado, la subjetividad en el análisis morfológico del desarrollo embrionario es un problema bien reportado en medicina reproductiva. La búsqueda de imparcialidad en la evaluación de embriones desempeña un papel fundamental en el éxito y progreso de un tratamiento de FIV. Según nuestros hallazgos, hay variaciones

en los resultados informados por clínicas de distintos países e incluso por embriólogos dentro de una misma clínica. Estas discrepancias pueden atribuirse a factores como la variabilidad en los estándares de calidad y otras circunstancias inherentes a la labor humana. Nuestros resultados sugieren que la inteligencia artificial (IA) puede abordar estas deficiencias y discrepancias que surgen durante la evaluación de embriones.

Por otro lado, al analizar imágenes a lo largo del tiempo, hemos observado un mejor rendimiento de la IA en comparación con los embriólogos en la clasificación de blastocistos humanos. Además, esta tecnología presenta diversas ventajas, como su rentabilidad, su carácter no invasivo y su capacidad para ofrecer una mayor consistencia en las evaluaciones en comparación con las realizadas por operadores individuales.

Objetivo específico 2. Validar externamente la puntuación automática asignada mediante deep learning para clasificar embriones en función de su potencial de implantación y nacimiento vivo: se ha suscitado discusión sobre si es necesario contar con una justificación biológica para respaldar los algoritmos desarrollados como “cajas negras” en la selección de embriones.

En primer lugar, algunos autores habían relacionado otros algoritmos con diversas variables del desarrollo embrionario. Nuestros resultados sugieren que la puntuación asignada por iDAScore v2 se correlaciona directamente con la calidad morfológica convencional determinada por los embriólogos.

En segundo lugar, la comunidad científica ya tenía conocimiento de que la puntuación automatizada asignada por varios algoritmos está asociada con resultados clínicos. Sin embargo, la aplicabilidad de sus resultados era discutible, ya que no existía una validación completamente externa en gran variedad de pacientes con el algoritmo iDAScore v2. Según nuestros descubrimientos, un aumento de una unidad en la puntuación del embrión, independientemente de otras variables clínicas analizadas, se asocia con mejores resultados clínicos.

En tercer lugar, la transferencia de embriones con la misma puntuación no conseguía los mismos resultados clínicos en todas las pacientes. Encontramos que la implantación esperada para embriones con la misma puntuación variaba según el tipo de tratamiento. Descubrimos que los embriones con puntuaciones idénticas tenían un mayor potencial de implantación en programas de donación de óvulos en comparación con tratamientos que utilizaban óvulos propios después de la transferencia de embriones frescos. Además, en el programa de donación de ovocitos, embriones con la misma puntuación tenían un potencial de implantación más bajo en las transferencias de embriones desvitrificados que en fresco. Sin embargo, esto no ocurrió en embriones de ovocitos propios de pacientes. Nuestro estudio culminó con el desarrollo de una plantilla de implantación esperada para distintos tipos de tratamiento, mejorando el asesoramiento a las pacientes.

Objetivo específico 3. Analizar el nivel de acuerdo entre la selección de embriones mediante inteligencia artificial y la decisión de los embriólogos: los profesionales senior empleaban con precaución las herramientas de inteligencia artificial debido a la preocupación de que puedan comprometer los resultados de los tratamientos de fecundación in vitro.

Por un lado, aunque se ha evidenciado que no todos los modelos de selección de embriones basados en inteligencia artificial coinciden en el mismo grado con las evaluaciones de los embriólogos, hemos observado un nivel satisfactorio de concordancia.

Por otro lado, es importante resaltar que los tratamientos mostraron mejores resultados cuando tanto el embriólogo como el modelo de inteligencia artificial estuvieron de acuerdo.

3.2. Perspectivas futuras

Los resultados de la presente tesis doctoral han proporcionado una comprensión profunda de varios aspectos y restricciones que podrían contribuir a la falta de consenso entre los profesionales en medicina reproductiva sobre la utilidad de la inteligencia artificial en la evaluación embrionaria. Estos resultados proporcionan confianza a los usuarios de la automatización en el laboratorio de fecundación in vitro y ofrecen orientación sobre cómo comprender y utilizar de manera efectiva los modelos basados en redes neuronales artificiales.

En primer lugar, la automatización de procedimientos como la evaluación morfológica de embriones sería factible desde una perspectiva objetiva. En este contexto, se podrían emplear herramientas como la mencionada en el objetivo específico 1. Se trata de una metodología completamente transparente que busca replicar el trabajo del embriólogo al asignar categorías a las estructuras de los blastocistos, con el propósito de reducir la disparidad entre observadores y mejorar el flujo de trabajo.

En segundo lugar, la puntuación automática asignada por un modelo de deep learning o “caja negra” podría utilizarse de forma rutinaria en los laboratorios de fecundación in vitro para evaluar embriones. La asociación directa entre la puntuación proporcionada por iDAScore v2 y el grado morfológico según criterios ASEBIR apoyaría la metodología de evaluación llevada a cabo desde el inicio de la embriología.

En tercer lugar, dicha puntuación automática también podría utilizarse de forma rutinaria en los laboratorios de fecundación in vitro para seleccionar embriones a transferir. La asociación entre una puntuación más alta y mejores tasas de euploidía, embarazo, implantación y recién nacidos vivos justifica el uso de esta herramienta y simplifica la labor del embriólogo al ofrecer una clasificación más detallada (1-9,9). La inclusión de un mayor número de categorías reduciría la probabilidad de tener que elegir entre dos embriones del mismo grado (A-C).

Finalmente, la implementación de herramientas basadas en inteligencia artificial para la evaluación y selección de embriones podría potenciar los resultados de los tratamientos, dado que la concordancia entre la inteligencia artificial y el embriólogo favoreció las tasas de implantación. La certeza de que no existe riesgo clínico asociado con la selección realizada por inteligencia artificial podría allanar el camino hacia el desarrollo de pautas consensuadas para su aplicación en el laboratorio de FIV, aliviando a los embriólogos de la carga de tomar decisiones sobre qué embrión transferir según los criterios actualmente establecidos.

3.3. Conclusiones

1. La inteligencia artificial es capaz de clasificar con precisión el grado de expansión de los blastocistos, la calidad de la masa celular interna y del trofoectodermo.
2. La inteligencia artificial demuestra resultados objetivos y más reproducibles al evaluar embriones tridimensionales basándose en una única imagen bidimensional en comparación con los operadores humanos.
3. Las puntuaciones automáticas asignadas por iDAScore v2 presentan una relación directa con las categorías morfológicas convencionales.
4. Las puntuaciones automáticas asignadas por iDAScore v2 son más altas en embriones con mayor probabilidad de euploidía.
5. Los embriones con mayor puntuación según iDAScore v2 tienen mayor probabilidad de dar lugar a un embarazo, de implantar y de conseguir un recién nacido vivo sano. No obstante, la transferencia de embriones con la misma puntuación, en distintos tipos de tratamientos, no tienen la misma probabilidad de éxito.
6. El rendimiento del modelo iDAScore v2 para seleccionar embriones es mayor que la morfología convencional en tratamientos con ovocitos de pacientes y de donantes sin diagnóstico genético preimplantacional. Sin embargo, el método de evaluación

convencional supera a la puntuación automática en tratamientos incluidos en el programa de PGT-A.

7. Confiar en la puntuación automática asignada por KIDScore D5 v3, iDAScore v1 o iDAScore v2 en casos de acuerdo entre el embriólogo y la IA resulta en mejores resultados del tratamiento de FIV.

8. El grado de acuerdo entre el embriólogo senior y la puntuación asignada por tres modelos de inteligencia artificial es bueno, entre 64% y 75%.

9. Aparentemente, no existe riesgo clínico asociado a seleccionar embriones por inteligencia artificial en los tratamientos de FIV. Sin embargo, sería necesario confirmar los hallazgos desde un punto de vista prospectivo. Esto implicaría transferir el embrión con mayor puntuación, en casos de discordancia con la morfología convencional.

10. Por tanto, la sinergia colaborativa entre la experiencia humana y la inteligencia artificial contribuye a resultados más favorables en los tratamientos de reproducción asistida.

Material suplementario

Contribuciones científicas

Khosravi P, Kazemi E, Zhan Q, Malmsten JE, **Toschi M**, Zisimopoulos P, Sigaras A, Lavery S, Cooper LAD, Hickman C, Meseguer M, Rosenwaks Z, Elemento O, Zaninovic N, Hajirasouliha I. Deep learning enables robust assessment and selection of human blastocysts after in vitro fertilization. NPJ Digit Med. 2019 Apr 4;2:21. doi: 10.1038/s41746-019-0096-y. PMID: 31304368; PMCID: PMC6550169.

Toschi M, Bori L, Rocha JC, Hickman C, Nogueira MFG, Satoshi A, Maffeis MC, Malmsten J, Zhan Q, Zaninovic N, Meseguer M. A combination of artificial intelligence with genetic algorithms on static time-lapse images improves consistency in blastocyst assessment: an interpretable tool to automate human embryo evaluation. Int J Fertil Steril. 2024 January In press; Available online doi: 10.22074/IJFS.2024.2008339.1510

Bori L, **Toschi M**, Esteve R, Delgado A, Pellicer A, Meseguer M. Largest external validation of a fully automated embryo evaluation with 68,471 embryos: Improved clinical outcomes with the top-ranked embryo. Fertil & Steril. Under Review, 2024.

O-177. October 17, 2023. What happens when the embryologist's opinion does not match the decision made by an artificial intelligence-based model? **Marco Toschi**, Lorena Bori Arnal, Alberto Tejera, Tamara Viloria, Antonio Pellicer, Marcos Meseguer.

Figuras suplementarias

D+4	D+5			
	Grado de expansión	MCI	Trofoectodermo	ASEBIR
Mórula compacta	Desde: "Iniciando la expansión" Hasta: "Ecllosionando"	A	A	A
			B	B
			C	C
			D	D
		B	A	A
			B	B
			C	C
			D	D
		C	A	A
			B	B
			C	C
			D	D
		D	A,B,C o D	D
	Blastocisto temprano o cavitando (ZP gruesa)			C
Mórula no compacta	Mórula			D

Figura suplementaria 1. Criterio para la evaluación de embriones in vitro según la asociación para el estudio de la biología de la reproducción (ASEBIR) en el quinto día del desarrollo embrionario.

D+5	D+6			
	Grado de expansión	MCI	Trofoectodermo	ASEBIR
Blastocisto temprano o cavitando (ZP gruesa)	Desde: "Iniciando la expansión" Hasta: "Ecllosionando"	A	A	B
			B	
			C	
			D	
		B	A	B
			B	
			C	
			D	
		C	A	B
			B	
			C	
			D	
		D	A,B,C o D	D
	Blastocisto temprano o cavitando (ZP gruesa)			D
Mórula no compacta	Mórula		Excluidos de la Clasificación	

Figura suplementaria 2. Criterio para la evaluación de embriones in vitro según la asociación para el estudio de la biología de la reproducción (ASEBIR) en el sexto día del desarrollo embrionario.

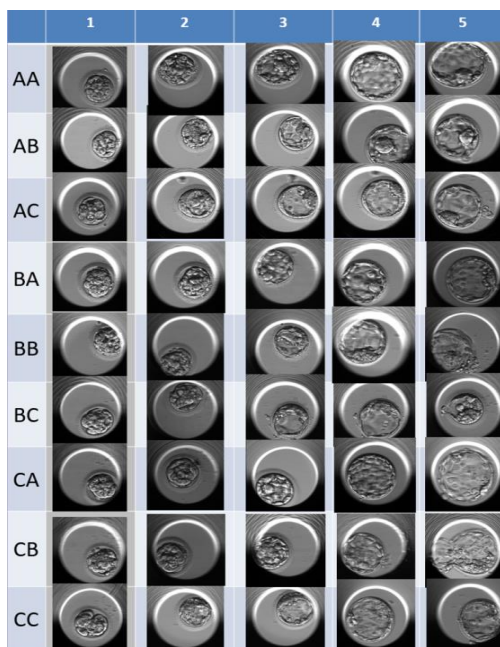


Figura suplementaria 3. Criterio para la evaluación de embriones in vitro según la clasificación de Gardner en quinto día del desarrollo embrionario.

Tablas suplementarias

Tabla suplementaria 1. Criterio para la evaluación de la masa celular interna según la asociación para el estudio de la biología de la reproducción (ASEBIR). MCI, masa celular interna.

Categoría	Tamaño MCI (μm^2)	Cohesión
A	3800-1900	Compacta
B	3800-1900	No compacta
C	1900	Indiferente
D	Signos de degeneración	
Excluidos	Degenerada	

Tabla suplementaria 2. Criterio para la evaluación del trofoectodermo según la asociación para el estudio de la biología de la reproducción (ASEBIR).

Categoría	Descripción del Trofoectodermo
A	Homogéneo, cohesionado y muchas células
B	Homogéneo; menos células
C	Pocas células
D	Signos de degeneración
Excluidos	Degenerado

Bibliografía

- Adamson, G. D., Abusief, M. E., Palao, L., Witmer, J., Palao, L. M., & Gvakharia, M. (2016). Improved implantation rates of day 3 embryo transfers with the use of an automated time-lapse-enabled test to aid in embryo selection. *Fertility and Sterility*, 105(2), 369-375.e6.
<https://doi.org/10.1016/j.fertnstert.2015.10.030>
- Afnan, M. A. M., Liu, Y., Conitzer, V., Rudin, C., Mishra, A., Savulescu, J., & Afnan, M. (2021). Interpretable, not black-box, artificial intelligence should be used for embryo selection. *Human Reproduction Open*, 2021(4), hoab040.
<https://doi.org/10.1093/hropen/hoab040>
- Aguilar, J., Rubio, I., Muñoz, E., Pellicer, A., & Meseguer, M. (2016). Study of nucleation status in the second cell cycle of human embryo and its impact on implantation rate. *Fertility and Sterility*, 106(2), 291–299.
<https://doi.org/10.1016/j.fertnstert.2016.03.036>
- Ahlström, A., Berntsen, J., Johansen, M., Bergh, C., Cimadomo, D., Hardarson, T., & Lundin, K. (2023). Correlations between a deep learning-based algorithm for embryo evaluation with cleavage-stage cell numbers and fragmentation. *Reproductive BioMedicine Online*, 47(6), 103408.
<https://doi.org/10.1016/j.rbmo.2023.103408>
- Alpha Scientists. (2011). The Istanbul consensus workshop on embryo assessment: proceedings of an expert meeting Alpha Scientists in Reproductive Medicine and ESHRE Special Interest Group of Embryology. *Human Reproduction*, 26(6), 1270–1283. <https://doi.org/10.1093/humrep/der037>
- Altman, D. (1991). *Practical statistics for medical research*. Chapman & Hall.

- Aparicio-Ruiz, B., Basile, N., Pérez Albalá, S., Bronet, F., Remohí, J., & Meseguer, M. (2016). Automatic time-lapse instrument is superior to single-point morphology observation for selecting viable embryos: retrospective study in oocyte donation. *Fertility and Sterility*, 106(6), 1379–1385. <https://doi.org/10.1016/j.fertnstert.2016.07.1117>
- Azzarello, A., Hoest, T., Hay-Schmidt, A., & Mikkelsen, A. L. (2017). Live birth potential of good morphology and vitrified blastocysts presenting abnormal cell divisions. *Reproductive Biology*, 17(2), 144–150. <https://doi.org/10.1016/j.repbio.2017.03.004>
- Barnes BS, B. J., Brendel MEng, M., Rajendran, S. B., Kim MEng, J., Zisimopoulos, P., Sigaras, A., Khosravi, P., Elemento, O., Hajirasouliha, I., Barnes, B. J., Brendel, M., Gao BS, V. R., Rajendran, S., Kim, J., Li, Q. B., Perelman, R. O., Cohen Center for Reproductive Medicine E Malmsten DPS, C. J., Zhan, Q., Rosenwaks, Z., ... Zaninovic, N. (2023). Department of Physiology and A non-invasive artificial intelligence approach for the prediction of human blastocyst ploidy: a retrospective model development and validation study. In *Articles Lancet Digit Health* (Vol. 5).
- Barrie, A., Homburg, R., McDowell, G., Brown, J., Kingsland, C., & Troup, S. (2017). Examining the efficacy of six published time-lapse imaging embryo selection algorithms to predict implantation to demonstrate the need for the development of specific, in-house morphokinetic selection algorithms. *Fertility and Sterility*, 107(3), 613–621. <https://doi.org/10.1016/j.fertnstert.2016.11.014>
- Berntsen, J., Rimestad, J., Lassen, J. T., Tran, D., & Kragh, M. F. (2022a). Robust and generalizable embryo selection based on artificial intelligence and time-lapse image sequences.

PLoS ONE, 17(2), 1–18.

<https://doi.org/10.1371/journal.pone.0262661>

Berntsen, J., Rimestad, J., Lassen, J. T., Tran, D., & Kragh, M. F. (2022b). Robust and generalizable embryo selection based on artificial intelligence and time-lapse image sequences. *PLoS ONE*, 17(2), 1–18.

<https://doi.org/10.1371/journal.pone.0262661>

Bori, L., Maor, R., Meseguer, F., Kottel, I., Seidman, D., Gilboa, D., & Meseguer, M. (2021). Artificial intelligence is moving closer to reproductive medicine: prediction of blastulation and embryo implantation. *American Society for Reproductive Medicine*.

Bori, L., Meseguer, F., Valera, M. A., Galan, A., Remohi, J., & Meseguer, M. (2022). 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. *Human Reproduction*, 37(6), 1148–1160.

<https://doi.org/10.1093/humrep/deac066>

Bori, L., & Meseguer, M. (2021). *Will the introduction of automated ART laboratory systems render the majority of embryologists redundant?* 43, 2021–2022. <https://doi.org/10.1016/j>

Bormann, C. L., Kanakasabapathy, M. K., Thirumalaraju, P., Gupta, R., Pooniwal, R., Kandula, H., Hariton, E., Souter, I., Dimitriadis, I., Ramirez, L. B., Curchoe, C. L., Swain, J., Boehnlein, L. M., & Shafiee, H. (2020). Performance of a deep learning based neural network in the selection of human blastocysts for implantation. *ELife*, 9, 1–14.

<https://doi.org/10.7554/ELIFE.55301>

Bormann, C. L., Thirumalaraju, P., Kanakasabapathy, M. K., Kandula, H., Souter, I., Dimitriadis, I., Gupta, R., Pooniwal,

- R., & Shafiee, H. (2020). Consistency and objectivity of automated embryo assessments using deep neural networks. *Fertility and Sterility. Scientific Congress Supplement: Oral and Poster Session Abstracts*, 113, 781–787.
- Cerrillo, M., Herrero, L., Guillén, A., Mayoral, M., & García-Velasco, J. A. (2017). Impact of Endometrial Preparation Protocols for Frozen Embryo Transfer on Live Birth Rates. *Rambam Maimonides Medical Journal*, 8(2), e0020.
<https://doi.org/10.5041/rmmj.10297>
- Chavez-Badiola, A., Flores-Sai, A., & Mendizabal-Ruiz, G. (2020). Predicting pregnancy test results after embryo transfer by image feature extraction and analysis using machine learning. *Scientific Reports*, 10(1), 4394.
- Cimadomo, D., Capalbo, A., Levi-Setti, P. E., Soscia, D., Orlando, G., Albani, E., Parini, V., Stoppa, M., Dovere, L., Tacconi, L., Ievoli, E., Maggiulli, R., Ubaldi, F. M., & Rienzi, L. (2018). Associations of blastocyst features, trophectoderm biopsy and other laboratory practice with post-warming behavior and implantation. *Human Reproduction*, 33(11), 1992–2001.
<https://doi.org/10.1093/humrep/dey291>
- Cimadomo, D., Chiappetta, V., Innocenti, F., Saturno, G., Taggi, M., Marconetto, A., Casciani, V., Albricci, L., Maggiulli, R., Coticchio, G., Ahlström, A., Berntsen, J., Larman, M., Borini, A., Vaiarelli, A., Ubaldi, F. M., & Rienzi, L. (2023). Towards Automation in IVF: Pre-Clinical Validation of a Deep Learning-Based Embryo Grading System during PGT-A Cycles. *Journal of Clinical Medicine*, 12(5).
<https://doi.org/10.3390/jcm12051806>

- Cimadomo, D., Sosa Fernandez, L., Soscia, D., Fabozzi, G., Benini, F., Cesana, A., Beatrice Dal Canto, M., Maggiulli, R., Muzzi, S., Scarica, C., Rienzi, L., & De Santis, L. (2022). Inter-centre reliability in embryo grading across several IVF clinics is limited: implications for embryo selection. *Reproductive BioMedicine Online*, 44, 39–48. <https://doi.org/10.1016/j>
- Ciray, H. N., Campbell, A., Agerholm, I. E., Aguilar, J., Chamayou, S., Esbert, M., & Sayed, S. (2014). Proposed guidelines on the nomenclature and annotation of dynamic human embryo monitoring by a time-lapse user group. *Human Reproduction*, 29(12), 2650–2660. <https://doi.org/10.1093/humrep/deu278>
- Cobo, A., Meseguer, M., Remohí, J., & Pellicer, A. (2010). Use of cryo-banked oocytes in an ovum donation programme: A prospective, randomized, controlled, clinical trial. *Human Reproduction*, 25(9), 2239–2246. <https://doi.org/10.1093/humrep/deq146>
- Conaghan, J., Chen, A. A., Willman, S. P., Ivani, K., Chenette, P. E., Boostanfar, R., Baker, V. L., Adamson, G. D., Abusief, M. E., Gvakharia, M., Loewke, K. E., & Shen, S. (2013). Improving embryo selection using a computer-automated time-lapse image analysis test plus day 3 morphology: Results from a prospective multicenter trial. *Fertility and Sterility*, 100(2), 412–419. <https://doi.org/10.1016/j.fertnstert.2013.04.021>
- Cox, C. M., Thoma, M. E., Tchangalova, N., Mburu, G., Bornstein, M. J., Johnson, C. L., & Kiarie, J. (2022). Infertility prevalence and the methods of estimation from 1990 to 2021: a systematic review and meta-analysis. *Human Reproduction Open*, 4, hoac051. <https://doi.org/10.1093/hropen/hoac051>

- Cozzolino, M., Franasiak, J., Andrisani, A., Ambrosini, G., & Vitagliano, A. (2020). "Delayed start" gonadotropin-releasing hormone antagonist protocol in Bologna poor-responders: A systematic review and meta-analysis of randomized controlled trials. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 244(Jan), 154–162.
- Cruz, M., Garrido, N., Herrero, J., Pérez-Cano, I., Muñoz, M., & Meseguer, M. (2012). Timing of cell division in human cleavage-stage embryos is linked with blastocyst formation and quality. *Reproductive BioMedicine Online*, 25(4), 371–381. <https://doi.org/10.1016/j.rbmo.2012.06.017>
- Curchoe, C. L., & Bormann, C. L. (2019). Artificial intelligence and machine learning for human reproduction and embryology presented at ASRM and ESHRE 2018. *Journal of Assisted Reproduction and Genetics*, 36(4), 591–600.
- Dal Canto, M., Coticchio, G., Mignini Renzini, M., De Ponti, E., Novara, P. V., Brambillasca, F., Comi, R., & Fadini, R. (2012). Cleavage kinetics analysis of human embryos predicts development to blastocyst and implantation. *Reproductive BioMedicine Online*, 25(5), 474–480. <https://doi.org/10.1016/j.rbmo.2012.07.016>
- Danuser, G. (2011). Essay Computer Vision in Cell Biology. *Cell*, 147(5), 973–978. <https://doi.org/10.1016/j.cell.2011.11.001>
- Del Gallego, R., Remohí, J., Meseguer, M., Affiliations, A., & Valencia, I. G. (2019). Time-Lapse Imaging : The State of the Art . *Biol Repod*, 101, 1146–1154.
- Desai, N., Ploskonka, S., Goodman, L., Attaran, M., Goldberg, J. M., Austin, C., & Falcone, T. (2016). Delayed blastulation, multinucleation, and expansion grade are independently associated with live-birth rates in frozen blastocyst transfer

- cycles. *Fertility and Sterility*, 106(6), 1370–1378.
<https://doi.org/10.1016/j.fertnstert.2016.07.1095>
- Desai, N., Ploskonka, S., Goodman, L. R., Austin, C., Goldberg, J., & Falcone, T. (2014). Analysis of embryo morphokinetics, multinucleation and cleavage anomalies using continuous time-lapse monitoring in blastocyst transfer cycles. *Reproductive Biology and Endocrinology*, 12, 54.
<https://doi.org/10.1186/1477-7827-12-54>
- Desch, L., Bruno, C., Luu, M., Barberet, J., Choux, C., Lamotte, M., Schmutz, E., Sagot, P., & Fauque, P. (2017). Embryo multinucleation at the two-cell stage is an independent predictor of intracytoplasmic sperm injection outcomes. *Fertility and Sterility*, 107(1), 97-103.e4.
<https://doi.org/10.1016/j.fertnstert.2016.09.022>
- Diakiw, S. M., Hall, J. M. M., VerMilyea, M. D., Amin, J., Aizpurua, J., Giardini, L., Briones, Y. G., Lim, A. Y. X., Dakka, M. A., Nguyen, T. V., Perugini, D., & Perugini, M. (2022a). Development of an artificial intelligence model for predicting the likelihood of human embryo euploidy based on blastocyst images from multiple imaging systems during IVF. *Human Reproduction*, 37(8), 1746–1759.
<https://doi.org/10.1093/humrep/deac131>
- Diakiw, S. M., Hall, J. M. M., VerMilyea, M. D., Amin, J., Aizpurua, J., Giardini, L., Briones, Y. G., Lim, A. Y. X., Dakka, M. A., Nguyen, T. V., Perugini, D., & Perugini, M. (2022b). Development of an artificial intelligence model for predicting the likelihood of human embryo euploidy based on blastocyst images from multiple imaging systems during IVF. *Human Reproduction*, 37(8), 1746–1759.
<https://doi.org/10.1093/humrep/deac131>

- Diakiw, S. M., Hall, J. M. M., VerMilyea, M., Lim, A. Y. X., Quangkananurug, W., Chanchamroen, S., Bankowski, B., Stones, R., Storr, A., Miller, A., Adaniya, G., van Tol, R. A., Hanson, R., Aizpurua, J., Giardini, L., Johnston, A., Van Nguyen, T., Dakka, M. A., Perugini, D., & Perugini, M. (2022). An artificial intelligence model correlated with morphological and genetic features of blastocyst quality improves ranking of viable embryos. *Reproductive BioMedicine Online*, 45(6), 1105–1117. <https://doi.org/10.1016/j.rbmo.2022.07.018>
- Dimitriadis, I., Zaninovic, N., Badiola, A. C., & Bormann, C. L. (2022). Artificial intelligence in the embryology laboratory: a review. *Reproductive Biomedicine Online*, 44(3), 435–448. <https://doi.org/10.1016/J.RBMO.2021.11.003>
- Ebner, T., Höggerl, A., Oppelt, P., Radler, E., Enzelsberger, S. H., Mayer, R. B., Petek, E., & Shebl, O. (2017). Time-lapse imaging provides further evidence that planar arrangement of blastomeres is highly abnormal. *Archives of Gynecology and Obstetrics*, 296(6), 1199–1205. <https://doi.org/10.1007/s00404-017-4531-5>
- Enatsu, N., Miyatsuka, I., An, L. M., Inubushi, M., Enatsu, K., Otsuki, J., Iwasaki, T., Kokeyuchi, S., & Shiotani, M. (2022). A novel system based on artificial intelligence for predicting blastocyst viability and visualizing the explanation. *Reproductive Medicine and Biology*, 21(1). <https://doi.org/10.1002/rmb2.12443>
- Erlach, I., Ben-Meir, A., Har-Vardi, I., Grifo, J., Wang, F., McCaffrey, C., McCulloh, D., Or, Y., & Wolf, L. (2022). Pseudo contrastive labeling for predicting IVF embryo developmental potential. *Scientific Reports*, 12(1). <https://doi.org/10.1038/s41598-022-06336-y>

- ESHRE Working group on time-lapse technology, 2020, Apter, S., Ebner, T., Freour, T., Guns, Y., Kovacic, B., Clef, N. Le, Marques, M., Meseguer, M., Montjean, D., Sfountouris, I., Sturmey, R., & Coticchio, G. (2020). Good practice recommendations for the use of time-lapse technology. *Human Reproduction Open*, 4, hoz025. <https://doi.org/10.1093/hropen/hoaa008>
- Ezoe, K., Shimazaki, K., Miki, T., Takahashi, T., Tanimura, Y., Amagai, A., Sawado, A., Akaike, H., Mogi, M., Kaneko, S., Okimura, T., & Kato, K. (2022). Association between a deep learning-based scoring system with morphokinetics and morphological alterations in human embryos. *Reproductive BioMedicine Online*, 45(6), 1124–1132. <https://doi.org/10.1016/j.rbmo.2022.08.098>
- Faddy, M. J., Gosden, M. D., & Gosden, R. G. (2018). A demographic projection of the contribution of assisted reproductive technologies to world population growth. *Reproductive BioMedicine Online*, 36(4), 455–458. <https://doi.org/10.1016/J.RBMO.2018.01.006>
- Fishel, S., Campbell, A., Montgomery, S., Smith, R., Nice, L., Duffy, S., Jenner, L., Berrisford, K., Kellam, L., Smith, R., D'Cruz, I., & Beccles, A. (2017). Live births after embryo selection using morphokinetics versus conventional morphology: a retrospective analysis. *Reproductive BioMedicine Online*, 35(4), 407–416. <https://doi.org/10.1016/j.rbmo.2017.06.009>
- Fishel, S., Campbell, A., Montgomery, S., Smith, R., Nice, L., Duffy, S., Jenner, L., Berrisford, K., Kellam, L., Smith, R., Foad, F., & Beccles, A. (2018). Time-lapse imaging algorithms rank human preimplantation embryos according to the probability of live birth. *Reproductive BioMedicine Online*, 37(3), 304–313. <https://doi.org/10.1016/j.rbmo.2018.05.016>

- Fréour, T., Fleuter, N., Lammers, J., Splingart, C., Reignier, A., & Barriere, P. (2015). External validation of a time-lapse prediction model. *Fertility and Sterility*, 103(4), 917–922. <https://doi.org/10.1016/j.fertnstert.2014.12.111>
- Garbade, D. M. J. (2018). Clearing the Confusion: AI vs Machine Learning vs Deep Learning Differences. *Towards Data Science*, 14.
- Gardner, D. K., Phil, D., Lane, M., Stevens, J., Schlenker, T., & Schoolcraft, W. B. (2000). *Blastocyst score affects implantation and pregnancy outcome: towards a single blastocyst transfer*.
- Gazzo, E., Peña, F., Valdéz, F., Chung, A., Bonomini, C., Ascenzo, M., & Escudero, E. (2020). The Kidscore TM D5 algorithm as an additional tool to morphological assessment and PGT-A in embryo selection: a time-lapse study. *Jornal Brasileiro de Reproducao Assistida*, 24(1), 55–60. <https://doi.org/10.5935/1518-0557.20190054>
- Gennatas, E. D., & Chen, J. H. (2021). Artificial intelligence in medicine: past, present, and future. *Artificial Intelligence in Medicine: Technical Basis and Clinical Applications*, 3–18. <https://doi.org/10.1016/B978-0-12-821259-2.00001-6>
- Glatstein, I., Chavez-Badiola, A., & Curchoe, C. L. (2023). New frontiers in embryo selection. In *Journal of Assisted Reproduction and Genetics* (Vol. 40, Issue 2, pp. 223–234). Springer. <https://doi.org/10.1007/s10815-022-02708-5>
- Goodman, L. R., Goldberg, J., Falcone, T., Austin, C., & Desai, N. (2016). Does the addition of time-lapse morphokinetics in the selection of embryos for transfer improve pregnancy rates? A randomized controlled trial. *Fertility and Sterility*, 105(2), 275–285. <https://doi.org/10.1016/j.fertnstert.2015.10.013>

- Gupta, K., Balyan, K., Lamba, B., Puri, M., Sengupta, D., & Kumar, M. (2022). Ultrasound placental image texture analysis using artificial intelligence to predict hypertension in pregnancy. *Journal of Maternal-Fetal and Neonatal Medicine*, 35(25), 5587–5594. <https://doi.org/10.1080/14767058.2021.1887847>
- Hassaballah, M., Alshazly, H. A., & Ali, A. A. (2019). Ear recognition using local binary patterns: A comparative experimental study. *Expert Systems with Applications*, 118, 182–200. <https://doi.org/10.1016/j.eswa.2018.10.007>
- Johansen, M. N., Parner, E. T., Mikkelsen, T., Kragh, F., Kato, K., Ueno, S., Palm, S., Kernbach, T., Manuel, Balaban, T., Başak, Keleş, İ., Gabrielsen, A. V., Lea, T., Iversen, H., & Berntsen, J. (2023). Comparing performance between clinics of an embryo evaluation algorithm based on time-lapse images and machine learning. *Journal of Assisted Reproduction and Genetics*, 40(3), 2129–2137. <https://doi.org/10.1007/s10815-023-02871-3>
- Joo, K., Nemes, A., Dudas, B., Berkes-Bara, E., Murber, A., Urbancsek, J., & Fancsovits, P. (2023). The importance of cytoplasmic strings during early human embryonic development. *Frontiers in Cell and Developmental Biology*, 11. <https://doi.org/10.3389/fcell.2023.1177279>
- Kato, K., Ueno, S., Berntsen, J., Kragh, M. F., Okimura, T., & Kuroda, T. (2023). Does embryo categorization by existing artificial intelligence, morphokinetic or morphological embryo selection models correlate with blastocyst euploidy rates? *Reproductive BioMedicine Online*, 46(2), 274–281. <https://doi.org/10.1016/j.rbmo.2022.09.010>
- Khosravi, P., Kazemi, E., Zhan, Q., Malmsten, J. E., Toschi, M., Zisimopoulos, P., Sigaras, A., Lavery, S., Cooper, L. A. D.,

- Hickman, C., Meseguer, M., & Rosenwaks, Z. (2019). Deep learning enables robust assessment and selection of human blastocysts after in vitro fertilization. *Npj Digital Medicine*, 2(21), 1–9. <https://doi.org/10.1038/s41746-019-0096-y>
- Kim, K. G. (2016). Deep Learning. *Healthcare Informatics Research*, 22(4), 351–354.
- Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521, 436–444. <https://doi.org/10.1038/nature14539>
- Loewke, K., Cho, J. H., Brumar, C. D., Maeder-York, P., Barash, O., Malmsten, J. E., Zaninovic, N., Sakkas, D., Miller, K. A., Levy, M., & VerMilyea, M. D. (2022). Characterization of an artificial intelligence model for ranking static images of blastocyst stage embryos. *Fertility and Sterility*, 117(3), 528–535. <https://doi.org/10.1016/j.fertnstert.2021.11.022>
- Martínez-Granados, L., Serrano, M., González-Utor, A., Ortíz, N., Badajoz, V., Olaya, E., Prados, N., Boada, M., & Castilla, J. A. (2017). Inter-laboratory agreement on embryo classification and clinical decision: Conventional morphological assessment vs. time lapse. *PLoS ONE*, 12(8), 1–13. <https://doi.org/10.1371/journal.pone.0183328>
- Mathur, M. B., Ding, P., Riddell, C. A., & VanderWeele, T. J. (2018). Web Site and R Package for Computing E-values. *Epidemiology*, 29(5), e45–e47. <https://doi.org/10.1097/EDE.0000000000000864>
- Matusevičius, A., Dirvanauskas, D., Maskeliūnas, R., & Raudonis, V. (2017). Embryo cell detection using regions with convolutional neural networks. *CEUR Workshop Proceedings*, 1856, 89–93.

- Meseguer, M., Herrero, J., Tejera, A., Hilligsøe, K. M., Ramsing, N. B., & Remoh, J. (2011). The use of morphokinetics as a predictor of embryo implantation. *Human Reproduction*, 26(10), 2658–2671. <https://doi.org/10.1093/humrep/der256>
- Meseguer, M., Maor, R., Bori, L., Shapiro, M., Pellicer, A., Seidman, D., Mercader, A., & Gilboa, D. (2022). Artificial intelligence (AI) based triage for preimplantation genetic testing (PGT); an AI model that detects novel features in the embryo associated with ploidy. *European Society of Human Reproduction and Embryology*.
- Milewski, R., Kuć, P., Kuczyńska, A., Stankiewicz, B., Łukaszuk, K., & Kuczyński, W. (2015). A predictive model for blastocyst formation based on morphokinetic parameters in time-lapse monitoring of embryo development. *Journal of Assisted Reproduction and Genetics*, 32(4), 571–579. <https://doi.org/10.1007/s10815-015-0440-3>
- Miyagi, Y. (2019). Feasibility of deep learning for predicting live birth from a blastocyst image in patients classified by age. *Reproductive Medicine and Biology*, 18, 190–203. <https://doi.org/10.1002/rmb2.12266>
- Motato, Y., de los Santos, M. J., Escriba, M. J., Ruiz, B. A., Remohí, J., & Meseguer, M. (2016). Morphokinetic analysis and embryonic prediction for blastocyst formation through an integrated time-lapse system. *Fertility and Sterility*, 105(2), 376–384. <https://doi.org/10.1016/j.fertnstert.2015.11.001>
- Peng, H., Zhou, J., Zhou, Z., Bria, A., Li, Y., Kleissas, D. M., Drenkow, N. G., Long, B., Liu, X., & Chen, H. (2016). Bioimage Informatics for Big Data. In *Focus on Bio-Image Informatics* (pp. 263–272). <https://doi.org/10.1007/978-3-319-28549-8>

- Petersen, B. M., Boel, M., Montag, M., & Gardner, D. K. (2016). Development of a generally applicable morphokinetic algorithm capable of predicting the implantation potential of embryos transferred on Day 3. *Human Reproduction*, 31(10), 2231–2244. <https://doi.org/10.1093/humrep/dew188>
- Polsky, A., Silver, D. H., Feder, M., Gold-Zamir, Y., Rosentraub, S., Shachor, E., Weinberger, A., Mazur, P., Zukin, V. D., & Bronstein, A. M. (2020). Deep learning model for improving embryo selection and deselection. *Human Reproduction. Abstracts of the 36th Virtual Annual Meeting of the ESHRE*, 35.
- Reignier, A., Girard, J., Lammers, J., Chtourou, S., & Lefebvre, T. (2019). Performance of Day 5 KIDScore™ morphokinetic prediction models of implantation and live birth after single blastocyst transfer. *Journal of Assisted Reproduction and Genetics*, 36(11), 2279–2285.
- Rienzi, L., Cimadomo, D., Delgado, A., Minasi, G., Fabozzi, G., Del Gallego, R., Stoppa, M., Bellver, J., Giancani, A., Esbert, M., Capalbo, A., Remohí, J., Greco, E., Ubaldi, F., & Meseguer, M. (2019). Time of morulation and trophectoderm quality are predictors of a live birth after euploid blastocyst transfer : a multicenter study. *Fertility and Sterility*, 112(6), 1080–1093.e1. <https://doi.org/10.1016/j.fertnstert.2019.07.1322>
- Rimestad, J., Kragh, M. F., Lassen, J. T., Tran, A., & Berntsen, J. (2020). Robust embryo scoring model based on artificial intelligence (AI) applied to a large time-lapse dataset. *Human Reproduction. Abstracts of the 36th Virtual Annual Meeting of the ESHRE*, 35.
- Rolfes, V., Bittner, U., Gerhards, H., Krüssel, J. S., Fehm, T., Ranisch, R., & Fangerau, H. (2023). Artificial Intelligence in

- Reproductive Medicine - An Ethical Perspective. *Geburtshilfe Und Frauenheilkunde*, 83(1), 106–115.
<https://doi.org/10.1055/a-1866-2792>
- Rubio, I., Kuhlmann, R., Agerholm, I., Kirk, J., & Herrero, J. (2012). Limited implantation success of direct-cleaved human zygotes : a time-lapse study. *Fertility and Sterility*, 98(6), 11–15. <https://doi.org/10.1016/j.fertnstert.2012.07.1135>
- Salih, M., Austin, C., Warty, R. R., Tiktin, C., Rolnik, D. L., Momeni, M., Rezaatofighi, H., Reddy, S., Smith, V., Vollenhoven, B., & Horta, F. (2023). Embryo selection through artificial intelligence versus embryologists: a systematic review. *Human Reproduction Open*, 2023(3), hoao031. <https://doi.org/10.1093/hropen/hoao031>
- Sarandi, S., Boumerdassi, Y., O'Neill, L., Puy, V., & Sifer, C. (2023). Interest of iDAScore (intelligent Data Analysis Score) for embryo selection in routine IVF laboratory practice: Results of a preliminary study. *Gynecologie Obstetrique Fertilité et Senologie*, 51(7–8), 372–377. <https://doi.org/10.1016/j.gofs.2023.05.001>
- Simopoulou, M., Sfakianoudis, K., Maziotis, E., Antoniou, N., Rapani, A., Anifandis, G., Bakas, P., Bolaris, S., Pantou, A., Pantos, K., & Koutsilieris, M. (2018). Are computational applications the “crystal ball” in the IVF laboratory? The evolution from mathematics to artificial intelligence. *Journal of Assisted Reproduction and Genetics*, 35(9), 1545–1557.
- Storr, A., Venetis, C. A., Cooke, S., Kilani, S., & Ledger, W. (2017). Inter-observer and intra-observer agreement between embryologists during selection of a single Day 5 embryo for transfer: A multicenter study. *Human Reproduction*, 32(2), 307–314. <https://doi.org/10.1093/humrep/dew330>

- Sundvall, L., Ingerslev, H. J., Breth Knudsen, U., & Kirkegaard, K. (2013). Inter- and intra-observer variability of time-lapse annotations. *Human Reproduction*, 28(12), 3215–3221. <https://doi.org/10.1093/humrep/det366>
- Theilgaard Lassen, J., Fly Kragh, M., Rimestad, J., Nygård Johansen, M., & Berntsen, J. (2023). Development and validation of deep learning based embryo selection across multiple days of transfer. *Scientific Reports*, 13(1), 4235–4235. <https://doi.org/10.1038/s41598-023-31136-3>
- Ueno, S., Berntsen, J., Ito, M., Okimura, T., & Kato, K. (2022). Correlation between an annotation-free embryo scoring system based on deep learning and live birth/neonatal outcomes after single vitrified-warmed blastocyst transfer: a single-centre, large-cohort retrospective study. *Journal of Assisted Reproduction and Genetics*, 39, 2089–2099. <https://doi.org/10.1007/s10815-022-02562-5>
- Ueno, S., Berntsen, J., Ito, M., Uchiyama, K., Okimura, T., Yabuuchi, A., & Kato, K. (2021). Pregnancy prediction performance of an annotation-free embryo scoring system on the basis of deep learning after single vitrified-warmed blastocyst transfer: a single-center large cohort retrospective study. *Fertility and Sterility*, 116(4), 1172–1180. <https://doi.org/10.1016/j.fertnstert.2021.06.001>
- Ueno, T., Ueno, S., Kakazu, Y., Akaike, N., & Nabekura, J. (2020). *Image Processing Toolbox TM 6 User's Guide. Image Process.: Vol. March.*
- Valera, M. A., Aparicio-Ruiz, B., Pérez-Albalá, S., Romany, L., Remohí, J., & Meseguer, M. (2023). Clinical validation of an automatic classification algorithm applied on cleavage stage embryos: analysis for blastulation, euploidy, implantation,

- and live-birth potential. *Human Reproduction*, 38(6), 1060–1075. <https://doi.org/10.1093/humrep/dead058>
- Vermilyea, M. D., Tan, L., Anthony, J. T., Conaghan, J., Ivani, K., Gvakharia, M., Boostanfar, R., Baker, V. L., Suraj, V., Chen, A. A., Mainigi, M., Coutifaris, C., & Shen, S. (2014). Computer-automated time-lapse analysis results correlate with embryo implantation and clinical pregnancy: A blinded, multi-centre study. *Reproductive BioMedicine Online*, 29(6), 729–736. <https://doi.org/10.1016/j.rbmo.2014.09.005>
- VerMilyea, M., Hall, J. M. M., Diakiw, S., Johnston, A., Nguyen, T., Dakka, M. A., Lim, A., Quangkananurug, W., Perugini, D., Murphy, A. P., & Perugini, M. (2020). Camera-agnostic self-annotating Artificial Intelligence (AI) system for blastocyst evaluation. *Human Reproduction. Abstracts of the 36th Virtual Annual Meeting of the ESHRE*, 35.
- VerMilyea, M., Hall, J. M. M., Diakiw, S. M., Johnston, A., Nguyen, T., Perugini, D., Miller, A., Picou, A., Murphy, A. P., & Perugini, M. (2021). Development of an artificial intelligence-based assessment model for prediction of embryo viability using static images captured by optical light microscopy during IVF. *Human Reproduction*, 35(4), 770–784. <https://doi.org/10.1093/HUMREP/DEAA013>
- Vijayarajeswari, R., Parthasarathy, P., Vivekanandan, S., & Basha, A. A. (2019). Classification of mammogram for early detection of breast cancer using SVM classifier and Hough transform. *Measurement: Journal of the International Measurement Confederation*, 146, 800–805. <https://doi.org/10.1016/j.measurement.2019.05.083>
- Voytovich, L., Greenberg, C., Voytovich, L., & Greenberg, C. (2022). Natural Language Processing: Practical Applications in

Medicine and Investigation of Contextual Autocomplete.
Machine Learning in Clinical Neuroscience, 207–214.
https://doi.org/10.1007/978-3-030-85292-4_24

- Wirka, K. A., Chen, A. A., Conaghan, J., & Ivani, K. (2014). Atypical embryo phenotypes identified by time-lapse microscopy : high prevalence and association with embryo development. *Fertility and Sterility*, 101(6), 1637-1648.e5.
<https://doi.org/10.1016/j.fertnstert.2014.02.050>
- Wong, C. C., Loewke, K. E., Bossert, N. L., Behr, B., De Jonge, C. J., Baer, T. M., & Pera, R. A. R. (2010). Non-invasive imaging of human embryos before embryonic genome activation predicts development to the blastocyst stage. *Nature Biotechnology*, 28(10), 1115–1121.
<https://doi.org/10.1038/nbt.1686>
- Wyns, C., Bergh, C., Geyter, C. De, Kupka, M. S., Motrenko, T., Rugescu, I., Smeenk, J., Vidakovic, S., & Goossens, V. (2020). ART in Europe , 2016 : results generated from European registries by ESHRE. *Human Reproduction Open*, 2020(3), 1–17. <https://doi.org/10.1093/hropen/hoaa032>
- Wyns, C., De Geyter, C., Calhaz-Jorge, C., Kupka, M., Motrenko, T., Smeenk, J., Bergh, C., Tandler-Schneider, A., Rugescu, I., Vidakovic, S., & Goossens, V. (2017). ART in Europe 2017: results generated from European registries by ESHRE † The European IVF-Monitoring Consortium (EIM) ‡ for the European Society of Human Reproduction and Embryology (ESHRE). *Human Reproduction Open*, 00(0), 1–17.
<https://doi.org/10.1093/hropen/hoab026>
- Yelke, H. K., Ozkara, G., Yuksel, B., Kumtepe Colakoglu, Y., Aygun, M., Brualla, A., Erlich, I., Hickman, C., Selimoglu, S., Okten, B., & Kahraman, S. (2022). Simplifying the complexity of

time-lapse decisions with AI: CHLOE (Fairtility) can automatically annotate morphokinetics and predict blastulation (at 30hpi), pregnancy and ongoing clinical pregnancy. *European Society of Human Reproduction and Embryology*.

- Zaninovic, N., Irani, M., & Meseguer, M. (2017). Assessment of embryo morphology and developmental dynamics by time-lapse microscopy: is there a relation to implantation and ploidy? *Fertility and Sterility*, 108(5), 722–729. <https://doi.org/10.1016/j.fertnstert.2017.10.002>
- Zaninovic, N., Sierra, J. T., Malmsten, J. E., & Rosenwaks, Z. (2023). Embryo ranking agreement between embryologists and artificial intelligence algorithms. *F and S Science*, 5(01), 50–57. <https://doi.org/10.1016/j.xfss.2023.10.002>
- Zhan, Q., Ph, D., Sierra, E. T., Ph, D., Malmsten, J., Ph, D., & Ye, Z. (2020). Blastocyst score, a blastocyst quality ranking tool, is a predictor of blastocyst ploidy and implantation potential. *Fertility and Sterility Reports*, 1(2), 133–141.
- Zhan, Q., Ye, Z., Clarke, R., Rosenwaks, Z., & Zaninovic, N. (2016). Direct unequal cleavages: Embryo developmental competence, genetic constitution and clinical outcome. *PLoS ONE*, 11(12), 1–19. <https://doi.org/10.1371/journal.pone.0166398>
- Zhu, J., Wu, L., Liu, J., Liang, Y., Zou, J., Hao, X., Huang, G., & Han, W. (2023). External validation of a model for selecting day 3 embryos for transfer based upon deep learning and time-lapse imaging. *Reproductive BioMedicine Online*, 47(3), 103242. <https://doi.org/10.1016/j.rbmo.2023.05.014>



Valencia, abril 2024