

TRABAJO FIN DE MÁSTER (TFM)

Facultad de Ciencias Biomédicas y de la Salud

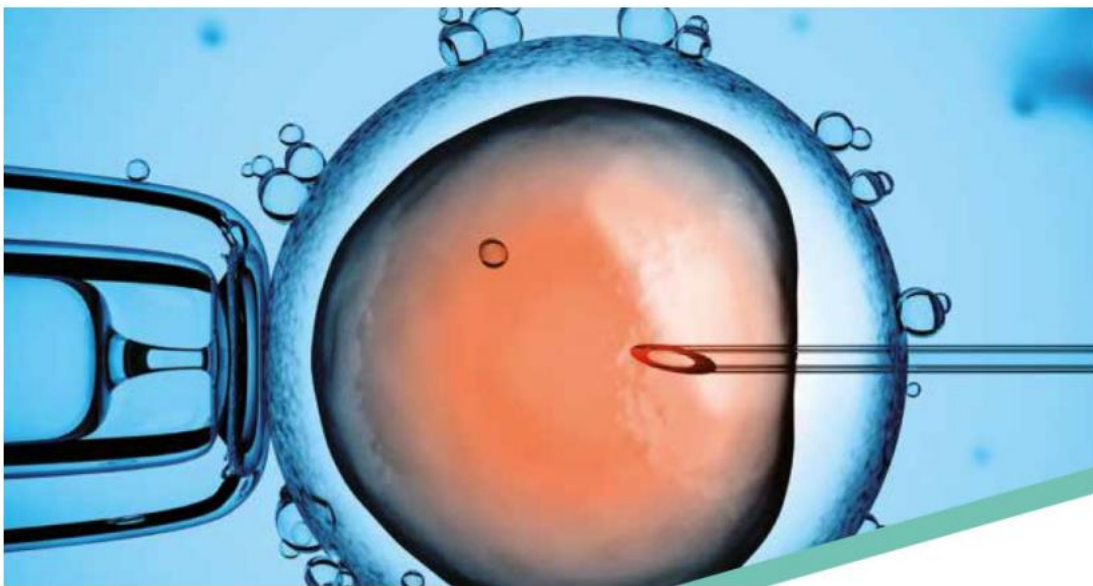
Titulación: Máster Universitario en Biología y Tecnología

Aplicada a la Reproducción Humana Asistida

Curso 2021-2022

Alexa Zepeda Hernández

**“Evaluación de la predicción de euploidía e
implantación de embriones humanos mediante
Inteligencia artificial time-lapse”**



Índice

1. Resumen	3
2. Introducción.....	4
Evaluación embrionaria “clásica”	4
Necesidades en la práctica clínica.....	5
Time-lapse e Inteligencia artificial en TRA	6
3. Objetivos	8
4. Materiales y Métodos	8
4.1 Criterios de selección aplicados a la búsqueda bibliográfica.....	9
4.1.1 Criterios de inclusión.....	9
4.1.2 Criterios de exclusión	9
4.1.3 Estrategia de valoración de información	9
5. Resultados.....	10
5.1 Evaluación de la calidad bibliográfica	10
5.2 Resumen de artículos seleccionados	13
6. Discusión	19
7. Conclusiones.	20
8. Bibliografía.	21

1. Resumen

Título: Evaluación de la predicción de euploidía e implantación de embriones humanos mediante Inteligencia artificial time-lapse. **Tipo de estudio:** revisión sistemática. **Introducción:** La evaluación de la morfología embrionaria es la forma clásica de evaluación de la calidad embrionaria, la variabilidad inter e intra observador está presente, así como un componente subjetivo. Por otro lado, el DGP es el gold estándar para evaluar la euploidía embrionaria. Estos dos procedimientos son la base de la selección embrionaria. Actualmente, la predicción de euploidía e implantación es una necesidad en la práctica clínica de los laboratorios FIV. Diversos modelos predictivos basados en IA sugieren una solución para solventar esta necesidad. Una aplicación práctica de los algoritmos es asistir a los usuarios en la toma de decisiones para brindar a los pacientes predicciones precisas. **Objetivos:** Resumir e identificar la eficacia reportada de los modelos IA para predecir la implantación y euploidía embrionaria en los tratamientos FIV. **Resultados:** Predicción de implantación AUC 0.60-0.82, estadísticamente significativa. Predicción de euploidía AUC 0.74-0.80, estadísticamente significativa. Los scores de los modelos de IA mostraban correlación con la viabilidad embrionaria (euploidía e implantación). **Discusión:** Los estudios demostraban ser predictivos, significativos y mostrar correlación con la implantación y euploidía embrionaria; algunos estaban basados en el entrenamiento del algoritmo, tenían un tamaño muestral bajo y necesitan validación con estudios prospectivos. **Conclusión:** La tecnología de IA es predictiva de implantación y euploidía, de acuerdo con los estudios presentados a la fecha. Sin embargo, se reconocen las limitantes de los algoritmos de IA y se hace énfasis en que deben de ser consideradas como asistentes de embriólogos para mitigar la variabilidad inter e intra evaluador, mas no una tecnología que hoy en día puede sustituir al embriólogo o al DGP.

Abstract

Title: Evaluation of ploidy and implantation prediction of human embryos by time-lapse AI. **Type of study:** Systematic Review. **Introduction:** Classic evaluation of embryo quality is the morphological assessment. Inter and intra-observer variability is present, with a subjective component. PGT is the gold standard to determine embryo ploidy. These two procedures are the basis of embryo selection. Nowadays, ploidy and implantation prediction are an unmet need in the IVF clinical practice. Several AI predictive models are trying to meet this need. AI algorithms can assist embryologists in the decision-making process by providing precise predictions. **Objectives:** To summarize and identify the reported efficacy of AI models to predict implantation and ploidy of human embryos in IVF treatments. **Results:** Implantation prediction AUC 0.60-0.82, statistically significant. Ploidy prediction AUC 0.74-0.80, statistically significant. AI score models showed correlation with implantation and ploidy. **Discussion:** Although, the studies were predictive of implantation and ploidy, some studies were still in the training phase of the algorithm, had a small sample and still need to be validated in prospective studies. **Conclusion:** Throughout the review, statistically significant studies were found to be predictive for ploidy and implantation; outperforming experienced embryologists in comparative studies. Nonetheless, AI's limitations are recognized and it's important to emphasize that AI models can be assistants to embryologists to mitigate inter and intra-observer variability; but to this day the embryologists or PGT cannot be substituted by this technology.

Key Words: Artificial intelligence (AI), assisted reproductive technology, machine learning (ML) models, Time-lapse morphokinetics, preimplantation genetic testing (PGT)

Abreviaturas

AI	Inteligencia artificial
ANN	Artificial neural networks
CNN	Convolutional neural network Red neuronal convolucional
DL	Deep learning
DGP	Diagnóstico genético preimplantacional
FIV	Fertilización in vitro
KID	Known implantation data
ML	Machine learning
NICS	Non invasive chromosome screening
MCI	Masa celular interna
TE	Trofoectodermo
TRA	Tratamiento de reproducción asistida
TL	Time lapse
TLI	Time lapse imaging
VPP	Valor predictivo positivo

2. Introducción

Evaluación embrionaria ‘clásica’

La infertilidad es una patología reproductiva que afecta a 186 millones de personas mundialmente, representando 8-12% de parejas. El tratamiento más común es la fertilización in vitro (FIV). La tecnología FIV ha mejorado considerablemente en los últimos 30 años, sin embargo, la tasa de éxito reportada es considerablemente baja (30%), de acuerdo con un reporte de la CDC (Centros para el control y prevención de enfermedades) en el 2015 [1].

Hoy en día la selección embrionaria para tratamientos de Reproducción asistida (**TRA**) es basada en la evaluación de la morfología y desarrollo embrionario. Para blastocistos, las clasificaciones usualmente toman en consideración la morfología de la masa celular interna y del trofoectodermo, así como la expansión de la cavidad del blastocisto. La variabilidad en la clasificación embrionaria conllevó al desarrollo de clasificaciones embrionarias con el objetivo de mitigar la falta de consenso intercentro. Sin embargo, no existe un estándar alfanumérico internacional para asignar scores que representen un embrión de calidad alta o baja [2].

Las anotaciones manuales son el gold estándar de evaluación embrionaria. Están basadas en una sola imagen estática y representan una evaluación subjetiva e incompleta. La tecnología time lapse permite la observación continua del desarrollo embrionario. No obstante, la

variabilidad inter e intraevaluador realizando anotaciones manuales o anotaciones en TLI está presente en ambas modalidades [1].

La selección embrionaria del mejor embrión de la cohorte de ovocitos fertilizados tiene el objetivo de llegar a un recién nacido vivo. La aneuploidía es una de las principales causas de fallo de tratamientos de reproducción asistida (TRA). El scoring morfológico es la forma “clásica” para evaluar la calidad embrionaria. Sin embargo, podemos encontrar fallos de implantación en embriones morfológicamente excelentes. Las técnicas para evaluar la euploidía embrionaria como el Diagnóstico genético preimplantacional (DGP) son invasivas y costosas. Hoy en día, métodos no invasivos para seleccionar embriones cromosómicamente normales basados en parámetros morfológicos, como time-lapse podrían ser una opción. No obstante, su valor clínico es controversial [3].

Necesidades en la práctica clínica

Actualmente, la predicción de euploidía e implantación es una necesidad en la práctica clínica de los laboratorios FIV. Partiendo de esta necesidad, diversos modelos computacionales predictivos basados en inteligencia artificial (IA) y machine learning (ML) sugieren una solución para solventar esta necesidad y asistir a los embriólogos en el trabajo diario [4].

El desarrollo de la primera neurona artificial fue por McCulloch y Pitts en 1943. La IA está basada en algoritmos computacionales que realizan tareas típicamente asociadas a la inteligencia humana. IA es un término genérico que puede ser subdividido en: artificial neural networks (ANNs), fuzzy logic, algoritmos genéticos, deep learning (DL) y machine learning (ML) [4].

Las herramientas de IA pueden analizar una cantidad masiva de información rápidamente. La técnica ha evolucionado y la percepción general del público también. El DL es una extensión de las técnicas clásicas de redes neuronales, capaces de construir redes neuronales artificiales con una cantidad grande de capas. Esto es posible por el desarrollo computacional [4].

Deep learning puede explorar patrones complejos de información no lineares. Por esta razón es usualmente usado en el análisis de imágenes, considerando que las imágenes son naturalmente complejas y requieren un análisis computacional extensivo. DL es un término que engloba diferentes tipos de algoritmos: redes neuronales recurrentes y la red neuronal convolucional (CNN). El término CNN indica que la red utiliza una operación matemática llamada convolución [4].

Las CNN utilizan nodos que aprenden de acuerdo con la información aportada (input) los cuales suma y pasa a través de una función de activación y se produce un resultado (output). En el análisis de imágenes, la operación matemática de convolución toma una imagen, la filtra y nos aporta un resultado (output). La imagen está compuesta por una matriz de elementos que son analizados por la red neuronal que finalmente son clasificadas para darnos un resultado. Este modelo ha sido utilizado para el conteo de células embrionarias, clasificación de embriones de día 1-5, determinar la fase de desarrollo embrionario, ha sido entrenado en base a la clasificación de Gardner, clasificando la MCI y TE, así mismo, para realizar predicciones de viabilidad embrionaria. Se ha intentado realizar predicción de recién nacido vivo, sin obtener resultados significativos [4].

Time-lapse e Inteligencia artificial en TRA

El enfoque de las técnicas de IA/ML está centrado en el desarrollo de modelos computacionales ayudando a que el tratamiento de FIV sean un proceso más eficiente y predictivo. El objetivo es proveer predicciones precisas, mediante la detección de patrones en las imágenes o videos time-lapse. Con la finalidad de suplementar y apoyar el proceso de selección embrionaria.

Los sistemas time-lapse han sido utilizados en la práctica clínica desde hace casi una década. Permite la observación continua y dinámica sin la necesidad de retirar a los embriones de las condiciones estables del incubador, limitando el daño por cambio de condiciones de cultivo. El uso de imágenes time-lapse para realizar anotaciones de las divisiones celulares es llamada morfocinética. Cada evento en el desarrollo embrionario puede ser medido en horas post inseminación (hpi) o como eventos aislados: tPNa-momento en el que aparecen los pronúcleos, tPNf-momento en el que desaparecen los pronúcleos. t2, t3, t4, t5, t6, t7, t8, t9+ corresponde a los tiempos de división hasta nueve células. tM se define como el estado de mórula, tSB es el momento donde se identifica el blastocelo, tB es el momento en que el blastocisto está totalmente formado, tEB-momento en que se identifica la expansión de la zona pelúcida y se hace referencia a ellos como parámetros de evaluación. Estos eventos morfocinéticos en el desarrollo embrionario han sido vinculados con la blastulación, implantación y contenido cromosómico [5].

Los algoritmos de IA llevan un proceso de aprendizaje basado en experiencia pasada, en este caso videos time-lapse de embriones. Una aplicación práctica de los algoritmos basados en

ML en la medicina es proporcionar asistencia a los usuarios en la toma de decisiones para brindar a los pacientes predicciones precisas [5].

Los algoritmos de IA utilizan diferentes parámetros morfocinéticos para realizar la evaluación embrionaria. Un modelo jerárquico considera t_5 , intervalo t_4-t_3 y t_3-t_2 para el ranking embrionario. En total, 10 clases de embriones son creadas para correlación con la predicción de implantación. Posteriormente, este modelo fue validado en un estudio multicéntrico. Sin embargo, 3 validaciones externas han sido incapaces de repetir los resultados [5].

Otro modelo combina la evaluación morfológica en día 2 con parámetros morfocinéticos (t_8 , t_2 y t_5 en relación a t_{PNf}). Su modelo provee un ranking correspondiente con la tasa de implantación estimada. Por otro lado, otro ranking embrionario está basado en los intervalos de t_2-t_5 , no obstante, no ha sido validado al día de hoy [5]. Los intervalos de división a 4 células son utilizados para predecir la viabilidad embrionaria. Por otro lado, Eeva TL test que toma en cuenta los intervalos de división celular y clasifica a embriones con un Score de TL alto: duración de t_2 9.33–11.45 horas y duración de t_3 0–1.73 horas. TL bajo es considerado como fuera de los rangos mencionados [6].

ML y la IA se han convertidos en uno de los temas mas controversiales de la medicina. Mientras los avances en la tecnología han sido exponenciales, es común que haya poco entendimiento en como los sistemas de IA toman decisiones. Diversos algoritmos de ML son inescrutables, particularmente algoritmos de DL. Esta inescrutabilidad puede afectar la confianza en los sistemas. Las consecuencias son significativas, ya que puede conducir al rechazo de los sistemas. Lo anteriormente mencionado es la base del término “black box” [7].

Para convertir un modelo de black box a un modelo glass box (“caja de cristal”) es necesario que las variables puedan ser interpretadas por humanos y conocer la información que forma parte del input de los modelos de Machine Learning [7].

Cada modelo utiliza diferentes intervalos morfocinéticos para valorar la predicción de implantación y euploidía. Es importante que estos modelos sean transparentes para evitar el fenómeno de black box. A continuación, se realiza una revisión bibliográfica para resumir los hallazgos al día de hoy correspondientes con la predicción de implantación y euploidía por modelos de IA.

3. Objetivos

1. Resumir e identificar la eficacia reportada de los modelos IA para predecir la implantación embrionaria en tratamientos FIV
2. Resumir e identificar la eficacia de los modelos IA para predecir euploidía embrionaria en tratamientos FIV.

4. Materiales y Métodos

Se realizó una búsqueda bibliográfica en Junio del 2022 para identificar artículos relevantes describiendo la predicción de implantación y euploidía por diferentes modelos y algoritmos AI/ML/DL para asistir en el proceso de FIV (Tabla 1).

Se llevó a cabo una búsqueda booleana basada en las siguientes palabras clave:

1. Assisted reproduction technology/ technique/ ART
2. In-vitro fertilization/ IVF
3. Artificial Intelligence/ AI
4. Machine learning/ ML
5. Deep learning/ DL
6. Algorithm
7. Implantation prediction
8. Ploidy prediction

Se realizó la búsqueda bibliográfica en las siguientes bases de datos:

- PubMed / Medline/ NIH/ Embase
- Google/ Google Scholar
- Cochrane Handbook for Systematic Reviews of Interventions
- SciFinder

Se seleccionaron 13 artículos basados en criterios de inclusión y exclusión, así como, la calidad de los estudios. Los artículos seleccionados fueron comparados en base a sus características tecnológicas y clínicas, usabilidad, performance, eficacia y seguridad.

Tabla 1. Búsqueda bibliográfica: búsqueda booleana y bases de datos.

Búsqueda booleana	Base de datos	Período de tiempo
<ul style="list-style-type: none"> • Assisted reproduction technology/ technique/ ART • In-vitro fertilization/ IVF • Artificial Intelligence/ AI • Machine learning/ ML • Deep learning/ DL • Algorithm • Implantation prediction • Ploidy prediction 	<ul style="list-style-type: none"> • PubMed/ Medline/ NIH • Google/ Google Scholar • Cochrane Handbook for Systematic Reviews of Interventions • SciFinder 	2016-2022

4.1 Criterios de selección aplicados a la búsqueda bibliográfica

4.1.1 Criterios de inclusión

- Estudios relacionados con AI/ML/DL
- Artículos con información clínica relacionada con predicción de implantación y euploidía.
- Artículos en inglés

4.1.2 Criterios de exclusión

- Artículos que no involucran embriones humanos
- Artículos usando tecnologías no relevantes
- Comentarios, cartas al editor
- Artículos duplicados

4.1.3 Estrategia de valoración de información

Las siguientes preguntas fueron utilizadas para seleccionar y valorar los artículos encontrados durante la búsqueda bibliográfica:

- ¿La información generada de la muestra de pacientes es representativa de la población de interés?
- ¿La información contiene suficiente información para realizar una evaluación racional y objetiva?
- ¿El diseño del estudio se considera apropiado?
- ¿Las medidas utilizadas para presentar los resultados reflejan el desempeño y seguridad de los algoritmos y modelos?
- ¿El análisis estadístico es adecuado?

5. Resultados

5.1 Evaluación de la calidad bibliográfica

Posterior a la búsqueda bibliográfica se realizó una evaluación de la calidad de los artículos para proceder a la selección bibliográfica. La base de la selección consta en artículos con un adecuado diseño metodológico: prospectivo/retrospectivo, comparativo/controlado, randomizado/no randomizado; reporte del uso de herramientas estadísticas adecuadas para reflejar el funcionamiento de los dispositivos y algoritmos de IA. Así como, justificar la selección en relación específica a la revisión de euploidía e implantación. (Tabla 2,3).

Tabla 2: Criterios de evaluación de calidad bibliográfica.

Contribución clínica	Descripción	Sistema de evaluación	
Tipo de fuente de datos	¿El diseño del estudio era adecuado?	T1	Si (e.g., prospectivo/ retrospectivo, comparativo / controlado, randomizado/no randomizado).
		T2	No (insuficiente o sin información del diseño del estudio).
Medición de resultados	¿Las mediciones de resultados reportadas reflejan el propósito del dispositivo?	O1	Si (i.e., aceptable)
		O2	No (insuficiente o sin información de resultados)
Significancia estadística	¿Tiene un análisis estadístico apropiado de la información y es considerado apropiado?	S1	Si (información con nivel de significancia estadística relacionada a los resultados del estudio y métodos)
		S2	No (insuficiente o sin información de resultados)

Contribución clínica	Descripción	Sistema de evaluación	
Significancia clínica	¿La magnitud del efecto de tratamiento observada fue clínicamente significativa?	C1	Si (i.e. clínicamente significativos relacionados con el estado del arte)
		C2	No (insuficiente o sin información)

Tabla 3. Evaluación de calidad de búsqueda bibliografía por contribución clínica.

	Publicación	Evaluación por contribución clínica				¿Aceptado?
1.	Automated halo identification: a novel predictive feature for IVF success identified through an artificial intelligence (AI) algorithm.	T1	O1	S1	C1	<input checked="" type="checkbox"/> SI <input type="checkbox"/> No
2.	Improved implantation rates of day3 embryo transfers with the use of an automated time-lapse-enabled test to aid in embryo selection	T1	O1	S1	C1	<input checked="" type="checkbox"/> SI <input type="checkbox"/> No
3.	External validation of a time-lapse model; a retrospective study comparing embryo evaluation using a morphokinetic model to standard morphology with live birth as endpoint.	T1	O1	S1	C1	<input checked="" type="checkbox"/> SI <input type="checkbox"/> No
4.	Data-Driven Prediction of Embryo Implantation Probability Using IVF Time-lapse Imaging	T1	O1	S1	C1	<input checked="" type="checkbox"/> SI <input type="checkbox"/> No
5.	Impact of the addition of Early Embryo Viability Assessment to morphological evaluation on the accuracy of embryo selection on day 3 or day 5: a retrospective analysis	T1	O1	S1	C1	<input checked="" type="checkbox"/> SI <input type="checkbox"/> No
6.	Performance of a deep learning based neural network in the selection	T1	O1	S1	C1	<input checked="" type="checkbox"/> SI <input type="checkbox"/> No

	Publicación	Evaluación por contribución clínica				¿Aceptado?
	of human blastocysts for implantation					
7.	The Kidscore™ D5 algorithm as an additional tool to morphological assessment and PGT-A in embryo selection: a time-lapse study	T1	O1	S1	C1	<input checked="" type="checkbox"/> SI <input type="checkbox"/> No
8.	Performance of Day 5 KIDScore™ morphokinetic prediction models of implantation and live birth after single blastocyst transfer	T1	O1	S1	C1	<input checked="" type="checkbox"/> SI <input type="checkbox"/> No
9.	Embryo ranking intelligent classification algorithm (ERICA): artificial intelligence clinical assistant predicting embryo ploidy and implantation	T1	O1	S1	C1	<input checked="" type="checkbox"/> SI <input type="checkbox"/> No
10.	ERICA ranking based on ploidy prediction is strongly correlated with pregnancy outcomes.	T1	O1	S1	C1	<input checked="" type="checkbox"/> SI <input type="checkbox"/> No
11.	Machine learning for prediction of euploidy in human embryos: in search of the best-performing model and predictive features.	T1	O1	S1	C1	<input checked="" type="checkbox"/> SI <input type="checkbox"/> No
12.	End-to-end deep learning for recognition of ploidy status using time-lapse videos	T1	O1	S1	C1	<input checked="" type="checkbox"/> SI <input type="checkbox"/> No
13.	An artificial intelligence model (euploid prediction algorithm) can predict embryo ploidy status based on time-lapse data	T1	O1	S1	C1	<input checked="" type="checkbox"/> SI <input type="checkbox"/> No

Los artículos fueron evaluados y se consideraron aptos para la selección de acuerdo con la contribución clínica, si se cumplía con los siguientes criterios:

- Calidad de recolección de datos - Q1 o Q2;

- Tipo de fuente de datos - T1;
- Objetivos - O1;
- Significancia estadística - S1;
- Significancia clínica - C1.

5.2 Resumen de artículos seleccionados

A continuación, se reportan los siguientes dispositivos de IA: KIDScore, AIVF-EMA, Auxogyn-Eeva, Ubar Embrionics, ERICA, entre otros modelos de IA, en relación con la predicción de implantación y euploidía (Tabla 4).

La implantación fue definida como la confirmación de latido fetal intrauterino a las 6-8 semanas de gestación mediante ultrasonido. La euploidía fue confirmada mediante estudios de PGT-A.

KIDScore es un modelo de implantación diseñado por Embryoscope para apoyar en la selección de embriones viables. Este modelo fue desarrollado a partir de datos de diversas clínicas. El algoritmo sólo utiliza tPNf, t2, t3, t5 y t8. KIDScore provee un ranking KID 1-5. Reportan tasas de implantación de embriones del 54% para KID5, KID4: 49%, KID3: 42%, KID 2: 29%, KID 1: 30%. Demostrando la correlación entre KIDScore e implantación, sin embargo, es necesario combinar su uso con la evaluación morfológica clásica. En otro estudio realizado se encontró una correlación positiva significativa entre el modelo y la tasa de implantación ($r=0.96$ and $r=0.90$, $p= 0.01$). Sin embargo, el poder predictivo de implantación fue significativo pero limitado: (AUC 0.60) [5, 13].

Auxogyn-Eeva describió y validó el primer test no invasivo (Eeva TL-test) que automáticamente mide los intervalos de división celular y genera resultados cuantitativos que clasifican a los embriones de acuerdo con su potencial de desarrollo. Eeva TL test es el primer test de time-lapse aprobado para ayuda a los embriólogos en la selección embrionaria. Así mismo, reportan una mejoría en las tasas de implantación 30.2% vs. 19.0%, TL Test vs morfología respectivamente, en transferencia en día 3 [6]. Kieslinger et al., 2016 fueron incapaces de mejorar el resultado en una validación externa en un estudio prospectivo en dos centros. En otro estudio Auxogyn-Eeva demostró mejorar la selección embrionaria de

embriones de día 3 mejorando la tasa de implantación con resultados significativos, 44.8% en DET-3 ME (morfología + Eeva) vs 30.2% DET-3 M (sólo morfología), $p < 0.02$. El uso de Eeva + morfología en SET de día 5 no se encontraron resultados significativos [8].

AIVF-EMA evalúa la presencia de un "halo" rodeando al núcleo del ovocito fertilizado como medida para predecir la implantación, reporta una Sensibilidad: 42%, especificidad: 83%, VPP: 85% y VPN: 46% [9].

Ubar Embionics reportó una AUC: 0.82 ± 0.07 en comparación con un panel de expertos conformado por 5 embriólogos de 5 países AUC = 0.58 ± 0.04 . PPV: Ubar 93% vs panel expertos $81 \pm 1\%$, NPV: 58% Ubar vs panel de expertos $23 \pm 8\%$. Los resultados del algoritmo muestran un 12% de incremento en el valor predictivo positivo y 29% de incremento en el valor predictivo negativo. Ubar obtuvo una predicción de implantación mayor vs un panel de expertos de 5 países y podría mejorar el potencial clínico de los embriones transferidos [10].

Un modelo de CNN entrenado para evaluar el potencial de implantación usando 97 embriones euploides superó a 15 embriólogos experimentados (75.26% vs. 67.35%, $p < 0.0001$) de 5 centros FIV diferentes. Este modelo está basado en la evaluación de imágenes estáticas a las 113 hpi, a diferencia de los modelos previamente mencionados que utilizan videos time-lapse [11].

En relación a la predicción de euploidía, las transferencias de embriones euploides + el modelo KIDScore D5 tuvieron tasas de implantación y embarazo más altas vs. transferencias de embriones euploides + morfología únicamente. (75.00% vs. 50.00%; $p = 0.002$ and 66.66% vs. 48.83%; $p = 0.037$, respectivamente. Las tasas de implantación fueron significativamente más altas en blastocistos con los KS5 más altos (KS5=6) comparado con blastocistos con KS5 más bajos (KS5=1) (80.00% vs. 49.02%; $p = 0.045$). Las tasas de euploidía fueron significativamente más altas en grupos con KS5=6 que los embriones con KS5=1 (61.88% vs. 48.33%; $p = 0.006$) [12].

A una base de datos con imágenes de blastocisto con KPI, se le aplicó retrospectivamente el algoritmo ERICA (Embryo ranking intelligent classification algorithm). El modelo fue evaluado en cuanto a su capacidad para predecir euploidía, comparar la predicción de euploidía contra embriólogos experimentados y observar si era capaz de hacer un ranking alto en base a

euploidía. 1231 imágenes fueron evaluadas y ERICA fue capaz de predecir euploidía: precisión=0.70 VPP: 0.79. Su métrica para ranking fue mejor que embriólogos $p=0.0242$. ERICA clasificó en primer lugar a blastocistos euploides en 78.9% de los casos y por lo menos un embrión euploide en el rank top 2, mejor que un ranking randomizado y que dos embriólogos. El proceso de clasificación y ranking de 4 blastocistos se llevó a cabo en menos de 25 segundos [14]. En otro estudio con 98 embriones los embriones clasificados como óptimos, tuvieron una tasa de embarazo bioquímico del 51%, $p=0.0378$ [15].

Un algoritmo basado en el modelo de IA RFC (Random Forest classifier) con 539 pacientes reportó una precisión del 71% para predecir euploidía, una AUC 0.75 y sensibilidad: 86%. Posterior al entrenamiento del algoritmo aumentaron a: Precisión: 72%, sensibilidad: 88% [5]. Otro algoritmo, en fase de entrenamiento con 108 pacientes reporta una AUC de 0.74 para predicción de euploidía [16]. Igualmente, otro estudio con 1803 embriones reporta una AUC de 0.80 [17].

Tabla 4. Resumen de resultados de revisión bibliográfica.

#	Publicación	Dispositivo IA	Diseño estudio	n=	Outcomes	Resultados	Conclusiones
1.	Automated halo identification: a novel predictive feature for IVF success identified through an artificial intelligence (AI) algorithm.	AIVF-EMA	Retrospectivo comparativo	123 videos time lapse	Presencia de un "halo" rodeando al núcleo del ovocito fertilizado y predicción de implantación.	Predicción de implantación Sensibilidad: 42% Especificidad: 83% VPP: 85% VPN: 46%	Algoritmo de IA identifica un halo en imágenes de video asociado con un valor predictivo positivo para una implantación exitosa.
2.	Improved implantation rates of day3 embryo transfers with the use of an automated time-lapse-enabled test to aid in embryo selection	Auxogyn-Eeva	Prospectivo	319 pacientes	TL test vs. Evaluación embrionaria con morfología tradicional Implantación embrionaria, embarazo, embarazo múltiple. Subanálisis de implantación embrionaria basado en score automático	Tasas de implantación 30.2% vs. 19.0%, TL Test vs morfología respectivamente. $p=.003$ En el grupo tratado, pacientes que recibieron al menos un embrión con alto TL tuvieron una tasa de implantación mayor que pacientes que recibieron embriones con TL bajo (36.8% vs.	Se muestra mejoría en tasas de implantación en pacientes recibiendo transferencias en día 3 basado en el test Time Lapse y la morfología tradicional. Confirma que el test TL no invasivo aporta valiosa a la clasificación tradicional

#	Publicación	Dispositivo IA	Diseño estudio	n=	Outcomes	Resultados	Conclusiones
					timelapse: score alto, score bajo.	20.6% P=.03)	morfológica.
3.	External validation of a time-lapse model; a retrospective study comparing embryo evaluation using a morphokinetic model to standard	KID Score	Retrospectivo	768 embriones	Modelo de implantación embrionario basado en morfocinética KIDScore comparado con morfología estándar.	Tasas de implantación de embriones del 54% para KID5, KID4: 49%, KID3: 42%, KID 2: 29%, KID 1: 30%.	En general KIDScore se correlaciona con implantación y recién nacido vivo comparado con morfología. Es necesario combinar su uso con la evaluación morfológica clásica
4.	Data-Driven Prediction of Embryo Implantation Probability Using IVF Time-lapse Imaging	Ubar	Retrospectivo	272 embriones	Predicción de implantación embrionaria.	AUC: Ubar 0.82±0.07 vs panel of experts AUC = 0.58±0.04 PPV: Ubar 93% vs panel expertos 81±1% NPV: 58% Ubar vs panel de expertos 23±8% Los resultados del algoritmo muestran un 12% de incremento en el VPP y 29% de incremento en el VPN.	Ubar obtuvo una predicción de implantación mayor vs un panel de expertos de 5 países. Ubar podría mejorar el potencial clínico de los embriones transferidos.
5.	Impact of the addition of Early Embryo Viability Assessment to morphological evaluation on the accuracy of embryo selection on day 3 or day 5.	Auxogyn-Eeva	Retrospectivo	328 pacientes 1818 embriones	Selección embrionaria con Eeva en día 3 y 5 + morfología estándar vs Morfología estándar. Outcomes: tasa de implantación y embarazo.	Tasas de implantación en Transferencia doble en día 3 con morfología + Eeva (DET-3 ME) 44.8% vs. morfología en día 3 (DET-M3) 30.2%, p< 0.02. SET en Dia 5 fue comparable SET-5 M: 49.2% SET-5 ME: 50%	Algoritmo Eeva puede mejorar la selección embrionaria de embriones de día 3 mejorando la tasa de implantación. El impacto en embarazo y recién nacido vivo necesita clarificarse.
6.	Performance of a deep learning based	Extend fertility. CNN AI	Retrospectivo	97 embriones euploides	Predicción de implantación de embriones	La predicción de implantación supero a 15 embriólogos	La predicción de implantación superó a

#	Publicación	Dispositivo IA	Diseño estudio	n=	Outcomes	Resultados	Conclusiones
	neural network in the selection of human blastocysts for implantation	algorithm			euploides	experimentados (75.26% vs. 67.35%, $p < 0.0001$).	embriólogos experimentados. Un tamaño muestral mayor es necesario.
7.	The Kidscore™ D5 algorithm as an additional tool to morphological assessment and PGT-A in embryo selection: a time-lapse study.	KIDScore D5	Prospectivo	912 embriones	Transferencias embrionarias de embriones euploides + KIDScore D5 vs. Transferencias embrionarias de embriones euploides + morfología. Outcomes: tasa de implantación, tasa de embarazo.	Transferencias embrionarias de embriones euploides + KIDScore D5 tuvieron tasas de implantación y embarazo mas altas vs. Transferencias de embriones euploides + morfología. (75.00% vs. 50.00%; $p = 0.002$ and 66.66% vs. 48.83%; $p = 0.037$, respectivamente. Las tasas de euploidia fueron significativamente más altas en grupos con $KS5 = 6$ que los embriones con $KS5 = 1$ (61.88% vs. 48.33%; $p = 0.006$).	La selección embrionaria basada en el algoritmo KS5 mejoro las tasas de implantación de transferencia de embrión único euploide. Los embriones con el KS5 Score más alto tenían mayor probabilidad de ser euploides e implantar.
8.	Performance of Day 5 KIDScore™ morphokinetic prediction models of implantation and live birth after single blastocyst transfer	KIDScore D5	Retrospectivo	210 ciclos ICSI	KIDScore D5 y predicción de implantación.	Una correlación positiva significativa se encontró entre el modelo y la tasa de implantación ($r = 0.96$ and $r = 0.90, p = 0.01$). El poder predictivo de implantación fue significativo pero limitado: (AUC 0.60).	Los modelos predictivos de KIDScore Dia 5 están significativamente asociados a la tasa de implantación de embriones únicos transferidos en día 5. Sin embargo, su poder predictivo puede ser perfeccionable. Necesita ser confirmado por estudios prospectivos randomizados con mayor tamaño muestral.
9.	Embryo ranking intelligent classification algorithm	ERICA	Retrospectivo	1231 imágenes de embriones	Input: ERICA ranking Outcomes:	Predicción de euploidia. precisión=0.70 VPP: 0.79.	Esta herramienta puede asistir a embriólogos en la selección del mejor embrión.

#	Publicación	Dispositivo IA	Diseño estudio	n=	Outcomes	Resultados	Conclusiones
	(ERICA): artificial intelligence clinical assistant predicting embryo ploidy and implantation						Es necesario comprobar reproductibilidad en bases de datos diferentes.
10	ERICA ranking based on ploidy prediction is strongly correlated with pregnancy outcomes.	ERICA	Retrospectivo	98 embriones	Algoritmo basado en predicción de euploidia. Outcome: tasa de embarazo bioquímico.	Los embriones clasificados como óptimos, tuvieron una tasa de embarazo bioquímico del 51%, $p=0.0378$	Los embriones clasificados como óptimos, basados en modelo de predicción de euploidia tuvieron una tasa de implantación con resultados significativos. Es necesario un tamaño muestral mas amplio.
11	Machine learning for prediction of euploidy in human embryos: in search of the best-performing model and predictive features	Algoritmo de IA	Cohorte retrospectivo	128 patients con 539 embriones 45% euploides, 55% aneuploides	Precisión, sensibilidad y AUC del algoritmo para predecir euploidía.	Random forest classifier (RFC) un modelo con una precisión del 71% para predecir euploidía y una AUC (0.75). Sensibilidad: 86% Posterior al entrenamiento del algoritmo aumentaron a: Precisión: 72%, sensibilidad: 88%,	El modelo RFC es capaz de predecir euploidía usando bases de datos de eventos morfocineticos, evaluación estándar, características clínicas y demográficas.
12	End-to-end deep learning for recognition of ploidy status using time-lapse videos	Algoritmo de IA	Retrospectivo comparativo	690 videos time-lapse de 108 pacientes	Outcome: predicción de euploidia de embriones con KID (PGT-A)	AUC= 0.74 en predicción de euploidía	Potencial predictor de euploidía. Estudio de entrenamiento de algoritmo. Necesita validación.
13	An artificial intelligence model (euploid prediction algorithm) can predict embryo ploidy status	Algoritmo de IA	Retrospectivo comparativo	1803 embriones	Outcome: Predicción de euploidía (AUC)	AUC=0.80	Algoritmo en entrenamiento con resultados predictivos de euploidía.

#	Publicación	Dispositivo IA	Diseño estudio	n=	Outcomes	Resultados	Conclusiones
	based on time-lapse data						

6. Discusión

Se encontró una variabilidad de resultados dependiendo del dispositivo de IA. Diversos estudios reportan la capacidad de los dispositivos de predecir implantación, sin sustituir la evaluación morfológica por embriólogos. Esto hace énfasis en que los algoritmos de IA pueden apoyar, mas no sustituir a los embriólogos. Emular a embriólogos experimentados es una misión difícil, ya que es necesario que los modelos reconozcan la variación del grosor de la zona pelúcida, el número de blastómeras, la simetría celular y la fragmentación citoplasmática. Es importante reconocer que esta tecnología depende de que los centros de fertilidad cuenten con tecnología time-lapse y la mayoría de los centros a nivel mundial no cuentan con estos incubadores [11].

KIDScore fue desarrollado a partir de 24 bases de datos en diferentes clínicas, con resultados satisfactorios, no obstante, su diseño es retrospectivo y son necesarios estudios prospectivos para validar resultados [5]. La selección embrionaria basada en el algoritmo KS5 mejoró las tasas de implantación de transferencia de embrión único euploide. Los embriones con el KS5 Score más alto tenían mayor probabilidad de ser euploides e implantar. Los modelos predictivos de KIDScore Dia 5 están significativamente asociados a la tasa de implantación de embriones únicos transferidos en día 5. Sin embargo, su poder predictivo puede ser perfeccionable. Necesita ser confirmado por estudios prospectivos randomizados con mayor tamaño muestral [13].

Auxogyn-Eeva reporta una mejoría en la tasa de implantación en comparación con el grupo control. No obstante, es un algoritmo destinado para transferencia de embriones de día 3. Hoy en día, la practica más usada es la transferencia en día 5. Los algoritmos de día 3 necesitan ajustes para ser capaces de predecir el potencial de implantación para embriones de día 5. Eeva no demostró resultados significativos en predecir implantación en día 5 [8].

El modelo de predicción de implantación basado en CNN, presentó los resultados del entrenamiento del algoritmo. A pesar de haber superado a embriólogos experimentados, es un

algoritmo en entrenamiento y es necesario validar el estudio con un tamaño muestral mayor [11].

El entrenamiento de ERICA para clasificar embriones (ranking) de acuerdo con su predicción de euploidía usando una imagen estática de cada embrión representa un potencial significativo para asistir a los embriólogos en elegir el mejor embrión y reducir tiempo en realizar anotaciones sin requerir estudios de biopsia invasivos. Es necesario evaluar este modelo en bases de datos mayores [14]. En otro estudio, los embriones clasificados como óptimos, basados en modelo de predicción de euploidía tuvieron una tasa de implantación con resultados significativos. Es necesario un tamaño muestral más amplio [15]. Así mismo, 3 algoritmos en fase de entrenamiento reportaron resultados predictivos de euploidia.

7. Conclusiones.

A lo largo de la revisión bibliográfica se encontraron resultados significativos, superando a embriólogos en estudios comparativos en cuanto a la predicción de implantación y euploidía. Es decir, la tecnología de IA es predictiva de implantación y euploidía, de acuerdo con los estudios presentados a la fecha. Sin embargo, se reconocen las limitantes de los algoritmos de IA y se hace énfasis en que deben de ser consideradas como asistentes de embriólogos para mitigar la variabilidad inter e intra evaluador, mas no una tecnología que hoy en día puede sustituir al embriólogo en el laboratorio FIV o a los estudios de DGP.

Entre otras limitantes encontramos que son estudios retrospectivos de tamaño muestral bajo. Es importante validar los algoritmos en entrenamiento con bases de datos diversas en características demográficas y clínicas. Así como realizar estudios prospectivos randomizados. Por otro lado, los algoritmos de día 3 necesitan ajustes para ser capaces de predecir el potencial de implantación para embriones de día 5. Ya que, la práctica clínica habitual es la transferencia en fase de blastocisto.

Se reportaron algunos de los parámetros utilizados por los modelos de IA mencionados en la literatura. No obstante, no se hace profundiza en rangos morfocinéticos y en los inputs que toman en cuenta los algoritmos. Los profesionales en fertilidad necesitan conocer que evalúan estos algoritmos y publicar con transparencia para evitar el efecto de Black box.

8. Bibliografía.

1. Khosravi P, Kazemi E, Zhan Q, Malmsten J, Toschi M, Zisimopoulos P, Sigaras A, Lavery S, Cooper L, 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 Digital Medicine*. 2019; 2(1), pp.21-29.
2. Racowsky C, Vernon M, Mayer J, Ball G, Behr B, Pomeroy K. Standardization of grading embryo morphology. *Journal of Assisted Reproduction and Genetics*. 2010; 27(8):437-439.
3. De Gheselle S, Jacques C, Chambost J, Blank C, Declerck K, De Croo I, Hickman C, Tilleman K. Machine learning for prediction of euploidy in human embryos: in search of the best-performing model and predictive features. *Fertility and Sterility*. 2021;117(4):738-746.
4. Fernandez E, Ferreira A, Cecílio M, Chéles D, De Souza R, Nogueira M. Artificial intelligence in the IVF laboratory: overview through the application of different types of algorithms for the classification of reproductive data. *Journal of Assisted Reproduction and Genetics*. 2020;37(10):2359-2376.
5. Adolfsson E, Porath S, Andershed AN. External validation of a time-lapse model; a retrospective study comparing embryo evaluation using a morphokinetic model to standard morphology with live birth as endpoint. *JBRA Assist Reproduction*. 2018;22(3):205-214.
6. Adamson GD, Abusief ME, Palao L, Witmer J, Palao LM, Gvakharia M. Improved implantation rates of day 3 embryo transfers with the use of an automated time-lapse-enabled test to aid in embryo selection. *Fertility Sterility*. 2016;105(2):369-75.
7. Rai, A. Explainable AI: from black box to glass box. *Journal of the academy marketing science*. 2020; 48, 137–141.
8. Revelli A, Canosa S, Carosso A, Filippini C, Paschero C, Gennarelli G, Delle Piane L, Benedetto C. Impact of the addition of Early Embryo Viability Assessment to morphological evaluation on the accuracy of embryo selection on day 3 or day 5: a retrospective analysis. *Journal Ovarian Research*. 2019;12(1):73.

9. Meseguer M, Maor U, Alegre L, Del Gallego R, Pellicer A, Seidman D, Gilboa D. Automated halo identification: a novel predictive feature for IVF success identified through an artificial intelligence (AI) algorithm. *Fertility and Sterility*. 2019; 112 (3).
10. Silver D, Feder M, Gold-Zamir Y, Polsky A, Rosentraub S, Shachor E, Weinberger A, Mazur P, Zukin V, Bronstein A. Data-Driven Prediction of Embryo Implantation Probability Using IVF Time-lapse Imaging. *MIDL*. 2020; 35 (10): 1-6.
11. Bormann CL, Kanakasabapathy MK, Thirumalaraju P, Gupta R, Pooniwala R, Kandula H, Hariton E, Souter I, Dimitriadis I, Ramirez LB, Curchoe CL, Swain J, Boehnlein LM, Shafiee H. Performance of a deep learning based neural network in the selection of human blastocysts for implantation. *Elife*. 2020; 9 (1): 55301
12. Gazzo E, Peña F, Valdéz F, Chung A, Bonomini C, Ascenzo M, Velit M, Escudero E. The Kidscore™ D5 algorithm as an additional tool to morphological assessment and PGT-A in embryo selection: a time-lapse study. *JBRA Assist Reprod*. 2020;24(1):55-60.
13. Reignier A, Girard JM, Lammers J, Chtourou S, Lefebvre T, Barriere P, Freour T. Performance of Day 5 KIDScore™ morphokinetic prediction models of implantation and live birth after single blastocyst transfer. *Journal of Assisted Reproduction and Genetics*. 2019;36(11):2279-2285.
14. Chavez-Badiola A, Flores-Saiffe-Farías A, Mendizabal-Ruiz G, Drakeley AJ, Cohen J. Embryo Ranking Intelligent Classification Algorithm (ERICA): artificial intelligence clinical assistant predicting embryo ploidy and implantation. *Reproductive Biomedicine Online*. 2020;41(4):585-593.
15. Drakeley A, Flores A, Chavez-Badiola A, Mendizabal G, Reyes-González D, Valencia R, Cohen J. P-244 ERICA's (Embryo Ranking Intelligent Classification Assistant) ranking, based on ploidy prediction, is strongly correlated with pregnancy outcomes. *Human Reproduction*. 2021; 36 (10): 109-115
16. Lee CI, Su YR, Chen CH, Chang TA, Kuo EE, Zheng WL, Hsieh WT, Huang CC, Lee MS, Liu M. End-to-end deep learning for recognition of ploidy status using time-lapse videos. *Journal of Assisted Reproduction and Genetics*. 2021; 38(7):1655-1663.
17. Huang B, Tan W, Li Z. An artificial intelligence model (euploid prediction algorithm) can predict embryo ploidy status based on time-lapse data. *Reproduction Biology Endocrinology*. 2021; (19): 185-191.