RBMO

ARTICLE

Artificial intelligence in the embryology laboratory: a review

BIOGRAPHY

Irene Dimitriadis, MD, is a fellowship-trained reproductive endocrinologist and infertility (REI) specialist at Massachusetts General Fertility Center. A native of Greece, she received her medical degree from the University of Athens and completed her OB/GYN residency at Tufts Medical Center and her REI fellowship at Massachusetts General Hospital in Boston.

Irene Dimitriadis[1], Nikica Zaninovic[2], Alejandro Chavez Badiola[3,4,5], Charles L. Bormann[1,*]

KEY MESSAGE

Artificial intelligence (AI) has the potential to be used as a tool to assist embryologists in daily activities such as performing morphological assessments and in selecting embryos for transfer. AI also has the potential to help clinicians make decisions and help patients achieve their goal of having a healthy baby.

ABSTRACT

The goal of an IVF cycle is a healthy live-born baby. Despite the many advances in the field of assisted reproductive technologies, accurately predicting the outcome of an IVF cycle has yet to be achieved. One reason for this is the method of selecting an embryo for transfer. Morphological assessment of embryos is the traditional method of evaluating embryo quality and selecting which embryo to transfer. However, this subjective method of assessing embryos leads to inter- and intra-observer variability, resulting in less than optimal IVF success rates. To overcome this, it is common practice to transfer more than one embryo, potentially resulting in high-risk multiple pregnancies. Although time-lapse incubators and preimplantation genetic testing for aneuploidy have been introduced to help increase the chances of live birth, the outcomes remain less than ideal. Utilization of artificial intelligence (AI) has become increasingly popular in the medical field and is increasingly being leveraged in the embryology laboratory to help improve IVF outcomes. Many studies have been published investigating the use of AI as an unbiased, automated approach to embryo assessment. This review summarizes recent AI advancements in the embryology laboratory.

[1] Department of Obstetrics and Gynecology, Massachusetts General Hospital, Harvard Medical School, Boston MA, USA
[2] The Ronald O Perelman and Claudia Cohen Center for Reproductive Medicine, Weill Cornell Medicine, New York NY, USA
[3] New Hope Fertility Center, Av. Prado Norte 135, Lomas de Chapultepec, Mexico City, Mexico
[4] IVF 2.0 LTD, 1 Liverpool Rd, Maghull, Merseyside, UK
[5] School of Biosciences, University of Kent Kent, UK

*Corresponding author. E-mail address: cbormann@partners.org (C. L. Bormann). https://doi.org/10.1016/j.rbmo.2021.11.003 1472-6483/
Declaration: The authors report no financial or commercial conflicts of interest.

KEYWORDS
Artificial intelligence
Embryo selection
IVF automation
IVF laboratory
Oocyte
Spermatozoa

INTRODUCTION

Since the late 1970s, when the first 'test-tube baby' was born in England, there have been many advances in the field of reproductive endocrinology and infertility. However, despite many attempts to create prediction models, it is still a struggle to accurately predict the outcome of an IVF cycle.

Initially, prediction models were based on well-known statistical models (*Bancsi et al., 2004*; *Hunault et al., 2002*; *Jurisica et al., 1998*; *van Weert et al., 2008*). More recently, the emerging technologies of time-lapse incubators and preimplantation genetic testing (PGT) were introduced as important achievements in the field, with the potential to produce a more objective method of selecting embryos with the best implantation probability. However, at present, there is insufficient evidence to recommend the routine use of these techniques for the sole purpose of improving single embryo transfer live birth rates (*Khosravi et al., 2019*; *Tiitinen, 2019*). Within the last decade, machine learning, more specifically convolutional neural networks (CNN), have been used to assist with medical imaging in a variety of fields, such as ophthalmology (*Abràmoff et al., 2016*), dermatology (*Esteva et al., 2017*), radiology (*Hosny et al., 2018*) and pathology (*Khosravi et al., 2018*) (TABLE 1). This technology has also been applied in the embryology laboratory, aiming to improve the selection of a single embryo with the best implantation potential to achieve the ultimate goal of fertility treatment: the birth of a healthy baby (*Khorsavi et al., 2019*). Since artificial intelligence (AI) has found a place in IVF, its potential use in nearly every aspect of infertility patient care has been investigated, including for identifying empty or oocyte-containing follicles; predicting embryo cell stages, blastocyst formation from oocytes, and live birth from blastocysts; assessing sperm morphology and human blastocyst quality; improving embryo selection; developing optimal IVF stimulation protocols; and quality control (*Bormann et al., 2021a*; *Curchoe and Bormann, 2019*). The goal of this review is to summarize recent advancements using AI technology in the embryology laboratory (TABLE 2).

AI LEARNING ALGORITHMS

AI is a general concept made up of diverse mathematical approaches with the capacity to make predictions based on complex pattern recognition by incorporating the processing power of computers (*Malik et al., 2021*). The selected algorithm(s) and the weight distribution attributed to its parameters define an AI model (*Burkov, 2019*).

The selection of a machine learning model is determined by the intended task (e.g. classification versus regression versus ranking), the dataset's characteristics (e.g. size, labelled/unlabelled data, structured versus unstructured data), and the planned learning approach (e.g. supervised, unsupervised). Based on these variables, scientists can choose among several different approaches to build algorithms or blocks (pipelines) of algorithms with different learning capabilities (i.e. shallow or deep learning). Examples of learning algorithms include artificial neural networks (ANN), support vector machines (SVM) and decision trees, among others (*Jordan and Mitchell, 2015*). Selecting machine learning models is difficult, which explains why sometimes several architectures can be tested at once (*Burkov, 2019*). *Chavez-Badiola et al. (2020b)* presented one such example as a proof of concept when five different algorithms were trained and tested on two datasets to assess their generalization capabilities to predict embryo implantation. This study presents an example of how this approach could guide scientists during the selection of a model toward clinical implementation. In this study, however, the limited size of the datasets could explain the poor performance of ANN, making a real comparison against ANN potentially inadequate.

Several other studies have tested multiple architectures (*Miyagi et al., 2019*; Morales *et al*., 2008), including the description by *VerMilyea et al. (2020)* of how different model architectures and hyperparameters (i.e. loss function and optimization methods) were considered before building their final architecture. Overall, results from these studies illustrate how different algorithms, even when trained on identical datasets, result in different performances, underlining the essential importance of a well-designed mathematical and computational approach.

AI algorithm training and validation

As problem complexity scales, most learning algorithms begin to show their inherent limits. One outstanding exception is ANN. ANN are designed to solve challenging classification problems and process large amounts of complex (non-linear) features simultaneously (*Lancashire et al., 2009*), which in turn tend to benefit from large training datasets. Disadvantages of ANN include their tendency to overfit and the 'black box' nature of their hidden layers (*Tu, 1996*).

ANN are a family of algorithms that includes CNN, which stand out for image analysis due to their ability to perform numerical matrix analysis, in contrast to non-CNN, which allow other information as input (e.g. age). As expected, CNN have become a common recourse for embryo analysis based on static images and time-lapse videos, as confirmed by the recent number of publications describing their implementation as either stand-alone solutions (*Bormann et al., 2020a*; *Chen et al., 2019*) or part of a pipeline of algorithms allowing for efficient image analysis (*Chavez-Badiola et al., 2020a*; *Kragh et al., 2019*).

The next step after selecting a learning algorithm is its training. This involves adjusting the model to minimize the error of the output using the values of the data provided as a ground truth (i.e. training), and a second step where the trained model is exposed to 'unseen' data to assess its performance (i.e. validation). The relevance of a high-quality dataset cannot be overestimated, because problems related to training on suboptimal datasets are numerous. One example is the result of training on an unbalanced dataset, which can lead to unreliable results (*Chawla et al., 2004*), which may have been the case in a study by *Tran et al. (2019)*. In this study, the high proportion of embryos with negative outcomes outweighed those with positive outcomes, resulting in a deeply unbalanced dataset, perhaps not representative of the problem, which in turn led to an almost unrealistic performance (area under the ROC curve of 0.93) (*Kan-Tor et al., 2020a*; *Tran et al., 2019*).

The size of a dataset is also relevant. However, encountering high-quality and large datasets is uncommon in the field of reproductive medicine due to a lack of standardization in data collection and storage, the routine use of manual annotations, and the challenges

TABLE 1 GLOSSARY OF ARTIFICIAL INTELLIGENCE (AI) TERMINOLOGY

AI terminology	Abbreviation	Definition	References
Adaptive adversarial neural networks	AANN	Method of deep learning that can be used with unlabelled data from unseen domain-shifted datasets to adapt pretrained supervised networks to new distributions, even when data from the original distribution are not available.	*Kanakasabapathy et al. (2021)*
Adversarial machine learning	AML	A machine learning technique that attempts to fool models by supplying deceptive input.	*Kianpour and Wen, 2020*
Artificial intelligence	AI	Any intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans and other animals.	*Malik et al. (2021)*; *Poole et al. (1998)*
Artificial neural network	ANN	A highly abstracted and simplified model compared to the mammalian brain, used in machine learning. A set of units receives input data, performs computations on them, and passes them to the next layer of units. The final layer represents the answer to the problem.	*Curchoe and Bormann (2019)*
Black box	–	The calculations performed by some deep learning systems between input and output are not easy (and potentially impossible) for humans to understand.	*Curchoe and Bormann (2019)*
Computer vision	CV	An interdisciplinary scientific field that deals with how computers can be made to gain high-level understanding from digital images or videos.	*Sonka et al. (2008)*
Convolutional neural network	CNN (or ConvNet)	In deep learning, a class of deep neural networks, mostly applied to analysing visual imagery.	*Curchoe and Bormann (2019)*
Data augmentation	–	In data analysis, techniques used to increase the amount of data. It helps reduce overfitting when training a machine learning.	*Shorten and Khoshgoftaar (2019)*
Decision tree	–	A flow chart-like structure in which each internal node represents a 'test' on an attribute, each branch represents the outcome of the test, and each leaf node represents a class label. The paths from root to leaf represent classification rules.	*Kamiński et al. (2017)*
Deep learning	DL	A specific subfield of deep learning. It is a process by which a neural network becomes sensitive to progressively more abstract patterns. Hundreds of successive layers of data representations are learned automatically through exposure to training data.	*Curchoe and Bormann (2019)*
Feature extraction	–	In machine learning, a feature is an individual measurable property or characteristic of a phenomenon. Features are intended to be informative and non-redundant, facilitating the subsequent learning and generalization steps, and in some cases leading to better human interpretations.	*Bishop (2006)*
Generative adversarial network	GAN	Two neural networks contest with each other in a game (in the form of a zero-sum game, where one agent's gain is another agent's loss).	*Goodfellow et al. (2014)*
Ground truth	–	Information that is known to be real or true, provided by direct observation and measurement (i.e. empirical evidence) as opposed to information provided by inference.	*Lemoigne and Caner (2006)*
Hidden layers	–	An internal layer of neurons in an artificial neural network, not dedicated to input or output.	*Uzair and Jamil (2020)*
Image segmentation	–	The process of partitioning a digital image into multiple segments (sets of pixels, also known as image objects). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyse.	*Shapiro and Stockman (2001)*
Machine learning	ML	Algorithms that find patterns in data without explicit instructions. Machine learning is a single contributing entity for AI technology.	*Curchoe and Bormann (2019)*
Overfitting	–	The production of an analysis that corresponds too closely or exactly to a set of data and may therefore fail to fit additional data or predict future observations reliably.	*Chicco (2017)*
Prediction models	–	Uses statistics to predict outcomes. Most often the event one wants to predict is in the future, but predictive modelling can be applied to any type of unknown event, regardless of when it occurred.	*Geisser (1993)*
Reinforcement learning	RL	An area of machine learning concerned with how software agents ought to take actions in an environment to maximize some notion of cumulative reward.	*Kaelbling et al. (1996)*
Shallow learning	–	A type of machine learning where we learn from data described by predefined features.	*Bengio et al. (2013)*
Supervised learning	SL	The machine learning task of learning a function that maps an input to an output based on example input–output pairs. It infers a function from labelled training data consisting of a set of training examples.	*Hinton and Sejnowski (1999)*; *Mohri et al. (2012)*
Support vector machines	SVM	In machine learning, support vector machines are supervised learning models with associated learning algorithms that analyse data used for classification and regression analysis.	*Cortes and Vapnik (1995)*
Synthetic data	–	Any production data applicable to a given situation that are not obtained by direct measurement.	*Patki et al. (2016)*
Test dataset	–	The sample of data used to provide an unbiased evaluation of a final model fit on the training dataset.	*Curchoe and Bormann (2019)*
Training dataset	–	The sample of data used to fit the model. The actual dataset that we use to train the model (weights and biases in the case of neural networks). The model sees and learns from this data.	*Curchoe and Bormann (2019)*
Transfer learning	TL	A technique in machine learning where the algorithm learns one task, and builds on that knowledge while learning a different, but related, task. Transfer learning is an alternative approach to help mitigate the large, manually annotated datasets needed for training an AI.	*Curchoe and Bormann (2019)*
Underfitting	–	Occurs when a statistical model cannot adequately capture the underlying structure of the data.	*Chicco (2017)*
Unsupervised learning	UL	A type of self-organized learning that helps find previously unknown patterns in datasets without pre-existing labels. It is also known as self-organization and allows modelling probability densities of given inputs.	*Hinton and Sejnowski (1999)*
Validation dataset	–	The sample of data used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyperparameters. The evaluation becomes more biased as skill on the validation dataset is incorporated into the model configuration.	*Curchoe and Bormann (2019)*

TABLE 2 LIST OF KEY ADVANCEMENTS IN THE AUTOMATION OF IVF LABORATORY PROCEDURES WITH THE AID OF AI

Cell type	ART procedure	Summary of advancement	References
Spermatozoa	Sperm count	Automated calculation of sperm concentration on a handheld device.	*Kanakasabapathy et al. (2017)*
	Sperm motility assessment	Automated calculation of sperm motility on a handheld device.	*Kanakasabapathy et al. (2017)*
	Forward progression score	Automated measurement of sperm velocity and classification of individual sperm forward progression score.	*Goodsen et al. (2017)*; *Kanakasabapathy et al. (2017)*
	DNA fragmentation assay	Automated measurement of sperm DNA fragmentation on a handheld device.	*Dimitriadis et al. (2019a)*
	Sperm viability assessment	Automated differential count of live–dead sperm staining.	*Dimitriadis et al. (2019a)*
	Sperm morphology measurement	Automated classification and measurement of normal and abnormal sperm morphology forms.	*Mirsky et al. (2017)*; *Thirumalaraju et al. (2019a)*
Oocyte	Oocyte morphology classification	Identification and classification of oocyte morphological features.	*Dickinson et al. (2020)*; *Manna et al. (2013)*; *Targosz et al. (2021)*
	Oocyte quality assessment	Association of oocyte morphology with pronuclear development and subsequent embryo development.	*Kanakasabapathy et al. (2020a)*; *Manna et al. (2013)*; *Sacha et al. (2021)*
	Oocyte maturation assessment	Automated identification of extruded polar body in metaphase II oocytes.	*Dickinson et al. (2020)*
	Alignment of oocyte for ICSI	Identification of proper location to inject spermatozoa into oocytes during ICSI.	*Dickinson et al. (2020)*
Pronuclear stage	Fertilization assessment	Automated fertilization assessment 14-18 h post-insemination.	*Dimitriadis et al. (2019b)*; *Kanakasabapathy et al. (2020b)*
	Pronuclear stage morphology classification	Segmentation and classification of pronuclear stage morphologic features.	*Zhao et al. (2021)*
	Pronuclear stage quality assessment	Prediction of embryo development at the pronuclear stage based on cytoplasmic movement.	*Coticchio et al. (2021)*
	Assessment of ICSI performance	Automated monitoring of individual embryologists performing ICSI using deep-learning enabled fertilization assessment.	*Thirumalaraju et al. (2019b)*
Cleavage stage	Predict day 5 embryo development	Prediction of blastocyst-stage development on Day 3 of development using extracted features, static images and time-lapse imaging data from cleavage-stage embryos.	*Bortoletto et al. (2019)*; *d'Estaing et al. (2021)*; *Kanakasabapathy et al. (2020a)*; *Liao et al. (2021)*; *Wang et al. (2018)*
	Predict implantation potential	Cleavage-stage prediction of embryo implantation using extracted features in a decision tree model and from direct learning using static images.	*Bormann et al. (2021a)*; *Carrasco et al. (2017)*
	Monitor embryo culture environment	Development of a KPI that associates the development prediction of cleavage-stage embryos with implantation outcomes.	*Bormann et al. (2021a)*
	Predict ploidy status of embryo	Non-invasive embryo ploidy prediction using static cleavage-stage embryo images.	*Meyer et al. (2020)*
	Identify correct location to perform assisted hatching	Identification of proper location to perform laser-assisted hatching based on cleavage-stage embryo morphology.	*Kelly et al. (2020)*
	Embryo identification and witnessing	Utilization of a CNN to assess cleavage-stage embryo quality and develop a unique key specific to each embryo for purposes of tracking and witnessing them throughout culture.	*Bormann et al. (2021b)*
Blastocyst stage	Blastocyst-stage classification	Classification and grading of blastocyst-stage embryos based on morphology and implantation outcome.	*Bormann et al. (2020b)*; *Khosravi et al. (2019)*; *Malmsten et al. (2020)*; *Leahy et al. (2020)*; *Thirumalaraju et al. (2021)*; *VerMilyea et al. (2020)*
	Vitrification and embryo biopsy decision-making	Use of static images to determine whether a blastocyst meets developmental criteria for vitrification and/or trophectoderm biopsy.	*Bormann et al. (2020b)*; *Souter et al. (2020)*
	Select embryo(s) for transfer	Prediction and selection of blastocyst-stage embryos for transfer based on static images, developmental size, trophectoderm expansion and proteomics.	*Bori et al. (2020*a, 2020b); *Bormann et al. (2020a)*; *Fitz et al. (2021)*; *Huang et al. (2021)*; *Louis et al. (2021)*; *Tran et al. (2019)*
	Predict ploidy status of embryo	Non-invasive embryo ploidy prediction using static blastocyst-stage embryo images and patient characteristics.	*Chavez-Badiola et al. (2020a)*; *Jiang et al. (2021)*; *Meyer et al. (2020)*; *Pennetta et al. (2018)*
	Quality assurance monitoring of laboratory procedures	Use of implantation prediction models to assess embryo selection, vitrification, warming and transfer competencies of embryologists and physicians.	*Dimitriadis et al. (2021)*
	Embryo identification and witnessing	Utilization of a CNN to assess blastocyst-stage embryo quality and develop a unique key specific to each embryo for purposes of tracking and witnessing them throughout culture.	*Kanakasabapathy et al. (2020c)*

AI = artificial intelligence; ART = assisted reproductive technology; CNN = convolutional neural network; ICSI = intracytoplasmic sperm injection; KPI = key performance indicator.

related to data sharing (*Curchoe, 2021*; *Hickman et al., 2020*). There are, however, strategies to optimize a dataset's size. Examples include the recourse to data augmentation made by *VerMileya et al. (2020)*, where images in their training set were subjected to manipulations (e.g. rotations, reflections, jitter) (*Kanakasabapathy et al., 2021*; *VerMileya et al., 2020*), allowing the training examples to multiply without a real increase in the size of the dataset. In this context, the use of synthetic data seems a promising tool to generate large, diverse, representative and balanced datasets without the constraint of accessing analogue clinical data. Still, understanding its inherent challenges will become paramount to making best use of this attractive approach (*Chen et al., 2021*). Another proposed solution to approach a limited-sized dataset is a top-down feature extraction (*Chavez-Badiola et al., 2020b*), which relies on the use of customized feature extractors designed with knowledge of the problem, as opposed to CNN, which require a lot of data to determine the feature extractors to use (bottom-up approach).

In brief, most training datasets used in AI protocols are labelled data, i.e. supervised learning. Labelling is performed by humans and thus is very subjective. In addition, if clinical outcome data are used, humans are selecting the embryos for transfer. The requirement for heterogeneous diverse training data, including an ethnically and racially diverse population of patients, is essential. A balanced set of data is also important to eliminate bias in AI learning (*Swain et al., 2020*). Unsupervised learning is an attractive alternative that needs to be explored.

Clinical training and validation

Validation as a part of the training process should be separated from the validation of a system in a clinical setting (*Curchoe et al., 2020*). AI should be built to become robust enough to perform beyond its training dataset. But as described recently by *Meseguer and Valera (2021)*, when a system is deployed in real life, specific conditions from new datasets, including the wide range of characteristics that surround clinical and laboratory procedures, may lead to an AI system's suboptimal and sometimes even erratic performance, a common machine learning problem known as under-specification (*D'Amour et al., 2020*).

Most current AI models for embryo selection rely on expert human supervision (supervised learning). One notable exception is the study by *Kanakasabapathy et al. (2021)*, where the authors present an adaptive adversarial neural network (AANN), which uses a form of unsupervised learning called adversarial learning. In this study, AANN performance was tested when using different microscopes on a variety of samples including human embryos, spermatozoa and blood cells. The authors compared a supervised learning model against their AANN and showed how the latter managed to maintain performance despite profound variations in image quality, suggesting AANN could overcome training bias and task-irrelevant feature information incorporated into the model. By training neural networks to focus on relevant features alone, AI might show better performance when deployed through different laboratory settings (*Kanakasabapathy et al., 2021*). Because this study only discriminates between blastocyst and no blastocysts, its clinical application during the embryo selection process is still to be tested. This, however, presents an example of how AI's training could be designed to become self-supervised.

Learning algorithms are attractive because they are expected to continuously improve performance as the available dataset grows. However, the brute force of a large dataset alone does not guarantee improved performance; if further training is not carefully undertaken, it risks performance degradation (*Lavin et al., 2021*) and the threat of data poisoning (Schwarzchild *et al*., 2021), whether intentional or not. Understanding the risks associated with further training is key to assessing a model's robustness (e.g. internal validation, external validation). Moreover, continuously evaluating its performance after tuning according to individual practices through a standardized quality assurance process is paramount, or at least highly desirable when considering the clinical readiness of an AI system (*Curchoe et al., 2020*; *Mahadevaiah et al., 2020*).

AI APPLICATION IN ASSISTED REPRODUCTIVE MEDICINE

Both invasive and non-invasive methods are used to select competent, healthy gametes for combination during assisted reproductive technology (ART) procedures. Every stage of ART treatment (fertilization, embryo development, implantation, healthy clinical pregnancy) depends on high-quality, mature, genetically normal spermatozoa and oocytes. Morphology of oocytes (cumulus–oocyte complex, polar body and ooplasm defects) and motility characteristics of spermatozoa (swim-up, gradient centrifugation or laminar flow microchannels on chip, and polyvinylpyrrolidone challenge) combined with morphology (vacuoles, head shape, and midpiece and tail defects) are routinely used to select gametes for insemination. Unfortunately, developmentally incompetent oocytes may exhibit the same morphology as competent ones. In addition, even high-powered microscopy, such as that used for intracytoplasmic morphologically selected sperm injection (IMSI), cannot detect DNA fragmentation in spermatozoa.

AI application on spermatozoa

In reproductive urology, early AI applications focused on semen parameters, but the technology has advanced to include the development of automated sperm detection and semen analyses. AI technology for semen analysis, sperm viability and DNA integrity has even been bridged with external hardware devices and smartphone (mobile) applications (*Dimitriadis et al., 2019a*; *Kanakasabapathy et al., 2017*).

Goodson et al. (2017) classified single spermatozoa as progressive, intermediate, hyperactivated, slow or weakly motile using SVM with 89.9% accuracy. Mirsky and colleagues employed interferometric phase microscopy along with SVM to develop a model to assess sperm morphology and classify spermatozoa into 'good' or 'bad' morphology with over 88% accuracy (*Mirsky et al., 2017*). *Thirumalaraju et al. (2019a)* used smartphone microscopy in conjunction with deep transfer learning to develop an inexpensive system that can accurately measure sperm morphology based on the Kruger strict criteria in the fifth edition of the WHO manual.

Ovarian stimulation management

Infertility is a multifactorial disease, which makes diagnosis and treatment complicated. *Liao et al. (2020)* have shown that a machine learning-derived

algorithm is useful to help clinicians make an efficient and accurate initial judgement on the condition of patients with infertility. In their study, the medical records of more than 60,000 infertile couples were evaluated using a grading system that classified patients into five grades, ranging from A to E. The worst grade, E, represented a 0.90% pregnancy rate, while the pregnancy rate in the A grade was 53.8%. The cross-validation results showed that the stability of the system was 95.9%.

Letterie and MacDonald (2020) evaluated a computer decision support system for day-to-day management of ovarian stimulation during IVF following key decisions made during an IVF cycle: (i) stop stimulation or continue stimulation; if the decision was to stop, then the next automated decision was to (ii) trigger or cancel. If the decision was to continue stimulation, then the next key decisions were (iii) the number of days to follow-up and (iv) whether any dose adjustment was needed. The authors used data derived from an electronic medical records system of a female population undergoing IVF cycles and oocyte cryopreservation to include the patient demographics, past medical history and infertility evaluation, including diagnosis, laboratory testing for ovarian reserve, and any radiological studies pertinent to a diagnosis of infertility. The four key decisions during the process of ovarian stimulation and IVF were compared to expert decisions across 12 providers; they were found to have a sensitivity of 0.98 for trigger and 0.78 for cycle cancellation.

AI application on oocytes

Ovarian stimulation yields oocytes at various stages of meiotic maturity. Identification of metaphase II (MII) (extruded polar body), metaphase I (MI) (no polar body), germinal vesicle (GV) (germinal vesicle indicative of prophase I), giant MII oocytes and other abnormalities is primarily performed by embryologists; however, nuclear and cytoplasmic maturity cannot be assessed. Non-invasive AI methods to evaluate oocyte competency could become an important selection and prediction tool to reduce the number of embryos created and wasted (of paramount importance in countries that restrict supernumerary embryos), to reduce the number of embryos for trophectoderm (TE) biopsy and PGT, and to provide the prognosis of the success of an IVF cycle. In the case of donor egg cycles, a tool to objectively assess oocyte quality and subsequent fertilization potential may be very valuable to intended parents for psycho-social reasons. Additionally, experimental and research procedures like in-vitro maturation (IVM) of oocytes, somatic cell nuclear transfer and reprogramming, in-vitro gametogenesis (IVG), and more would benefit from prediction and selection AI systems.

In *2011, Setti* and colleagues performed a meta-analysis to identify the relationship between oocyte morphology and ICSI outcomes (*Setti et al., 2011*). Their study demonstrated that the presence of a large first polar body and a large perivitelline space and the inclusion of refractile bodies or vacuoles are associated with decreased oocyte fertilization. In a later study, *Manna et al. (2013)* performed texture analysis of 269 oocyte images and tracked the corresponding embryo development. Texture features were used with a neural network to predict the outcome of a given cycle, meaning that multiple transfers were present in the data used, for an AUC of 0.80. In 2021, Targosz and colleagues tested 71 deep neural network models for semantic oocyte segmentation (*Targosz et al., 2021*). They trained their algorithm to classify the following oocyte morphological features: clear cytoplasm, diffuse cytoplasmic granularity, smooth endoplasmic reticulum cluster, dark cytoplasm, vacuoles, first polar body, multipolar body, fragmented polar body, perivitelline space, zona pellucida, cumulus cells and the GV. In this study, the top training accuracy (ACC) reached about 85% for training patterns and 79% for validation.

In 2020, Kanakasabapathy and colleagues trained a CNN to predict fertilization potential (two-pronuclear [2PN] or non-2PN [pronuclear formation]) from oocyte images and to identify oocytes with the highest fertilization potential >86% of the time (*Kanakasabapathy et al., 2020a*). Results from this study allow for the development of novel quality assurance tools used to monitor oocyte stimulation regimens, assess ICSI performance, maintain optimal fertilization and embryo culture conditions, and evaluate oocyte vitrification and warming procedures. This oocyte quality algorithm was helpful in identifying an association between oocyte morphology and subsequent embryo development (*Sacha et al., 2021*).

Dickinson et al. (2020) used deep CNN to locate the first extruded polar body, which allowed them to distinguish mature, MII oocytes from MI and GV stage oocytes. Pinpointing the location of the extruded polar body also allowed this algorithm to identify the correct location on the oocyte to inject spermatozoa for ICSI. In this study, over 14,000 images of MII oocytes were used for training, validation and testing. The deep learning CNN was able to correctly identify the location of the polar body and the corresponding location for sperm injection for a test set of 3888 oocytes with 98.9% accuracy with a 95% confidence interval (CI) ranging between 98.5% and 99.2% (*Dickinson et al., 2020*).

AI application on pronuclear-stage embryos

Normal fertilization follows a definite course of events. Oocytes show circular waves (*Payne et al., 1997*) of granulation within the ooplasm after ICSI. During this granulation phase, the sperm head decondenses and the second polar body is extruded. This is followed by the formation of the male pronucleus. At about the same time, the female pronucleus forms and is drawn toward the male pronucleus until apposition is achieved. Both pronuclei then increase in size, and their nucleoli move around and arrange themselves near the common junction. Only zygotes with two distinct pronuclei are considered normal and appropriate for transfer. It is critical that embryologists assess fertilization status correctly, as there is only a small window of time in which pronuclei can be properly counted.

Fertilization checks and embryo quality assessments require manual examination, status recording and embryo development scoring. These processes are labour-intensive and subjective. In 2019, Dimitriadis and colleagues described the development of a CNN that can distinguish between 2PN and non-2PN zygotes at 18 h post-insemination with >90% accuracy (*Dimitriadis et al., 2019b*). This system can be used as an embryologist aid to help confirm the fertilization assessment of each oocyte. It can also be used to monitor individual embryologists performing ICSI in a clinical setting for

advanced quality assurance to improve patient outcomes (*Bormann et al., 2021a*; *Thirumalaraju et al., 2019b*).

Several studies have shown that morphological features specific to the pronuclear-stage embryo can be used to assess embryo quality and developmental potential. These grading systems factor in the size, shape and alignment of pronuclei. They also factor in the number and distribution of nucleoli and the overall appearance of the cytoplasm (*Scott and Smith, 1998*; *Scott et al., 2000*; *Tesarik and Greco, 1999*). These morphological grading systems have also been shown to help aid embryologists in selecting embryos with high implantation potential (*Lan et al., 2003*; *Zollner et al., 2003*). Manually scoring zygotes is a labour-intensive and subjective activity. As such, few practices continue to assess this critical stage of development. However, with the use of AI, these predictive features may be readily incorporated into an embryo selection algorithm.

In 2021, Zhao and colleagues used CNN for segmentation of pronuclear-stage embryos. They examined the morphokinetic patterns of the zygote cytoplasm, zona pellucidae and pronuclei. Their manually annotated test set had precision of >97% for the cytoplasm, 84% for the pronuclei and approximately 80% for the zona pellucida. The authors concluded that their CNN system has the potential to be incorporated in clinical practice for pronuclear-stage segmentation as a powerful tool with high precision, reproducibility and speed (*Zhao et al., 2021*). Early parameters of zygotic (cytoplasmic movement) development, analysed by AI-powered methods, have been shown to be predictive of blastocyst development. Compared to human evaluation and prediction using morphological parameters, AI-based methods using cytoplasmic kinetics showed on average 10% higher accuracy (*Coticchio et al., 2021*).

AI application on cleavage-stage embryos

Embryo transfers are generally performed at the cleavage or blastocyst stage of development. Cleavage-stage embryos are generally selected for transfer based on only three features: blastomere cell count, percentage of overall cytoplasmic fragmentation and degree of asymmetry between blastomeres (*Prados et al., 2012*). These grades are assigned by visual examination of the embryos and have been shown to be highly subjective in nature.

The introduction of time-lapse imaging (TLI) technology has allowed for both automated and manual assessments of embryo development at precise times and under controlled environments (*Azzarello et al., 2012*; *Cruz et al., 2012*; *Hlinka et al., 2012*; *Lechniak et al., 2008*; *Lemmen et al., 2008*). However, most of the TLI algorithms have only shown promising results in identifying embryos with low developmental potential. The incorporation of TLI systems to standard manual embryo assessments did not improve overall clinical outcomes, nor did they decrease the amount of time embryologists spent assessing embryo morphology (*Chen et al., 2017*; *Conaghan et al., 2013*; *Kaser et al., 2016*; *Kirkegaard et al., 2015*).

Dimitriadis et al. (2017) demonstrated a fast and simple cohort embryo selection (CES) method for selecting cleavage-stage embryos that will develop into high-quality blastocysts. This study demonstrated the ability of embryologists to quickly identify high-quality cleavage-stage embryos when all embryos in the cohort were simultaneously compared in a single image. This method of selection outperformed traditional methods of cleavage-stage embryo ranking based on both morphology and adjunctive morphokinetic TLI parameters. This method is excellent at identifying high-quality embryos from a cohort; however, this method of selection is subjective and lacks consistency between operators.

Computer vision technology has been proposed as a solution to overcome the labour constraints and subjective nature of assessing and selecting embryos based on morphology and morphokinetic measurements. Kanakasabapathy and colleagues used deep learning CNN to train and validate embryo assessments on Day 3 embryo images based on embryo developmental outcomes recorded on Day 5 of culture. This algorithm was trained to make the following Day 5 developmental predictions: embryo arrest, morula, early blastocyst, full blastocyst and high-quality blastocyst. Using a test set of 748 embryos, the accuracy of the algorithm in predicting blastocyst development at 70 hpi was 71.9% (95% CI 68.4–75.2%) (*Bortoletto et al., 2019*; *Kanakasabapathy et al., 2020c*).

To evaluate the potential improvement in predictive power, *Kanakasabapathy et al. (2020c)* also compared the accuracy of predictions by embryologists in identifying embryos that will eventually develop into blastocysts when presented with embryo morphology imaged on Days 2 and 3 of development. Additionally, their performance was evaluated with and without the use of the Eeva three-category TLI algorithm that uses P2 (duration of the 2-cell stage) and P3 (duration of the 3-cell stage) to predict blastocyst development (*VerMilyea et al., 2014*). The neural network significantly outperformed the embryologists in identifying embryos that will develop into blastocysts correctly ($P < 0.0001$) and the overall accuracy in prediction, regardless of the evaluated methodology ($P < 0.0001$). This was the first AI-based system for predicting the developmental fate of cleavage-stage embryos (*Kanakasabapathy et al., 2020c*).

Bormann et al. (2021a) described an early warning system for using cleavage-stage embryos and statistical process controls for detecting clinically relevant shifts due to laboratory conditions. This study presented a novel key performance indicator (KPI) for monitoring embryo culture conditions at the cleavage stage of development. This AI-based KPI predicted the percentage of cleavage-stage embryos that would develop into high-quality blastocysts on Day 5 of development. When compared with five established cleavage-stage KPI, this AI-based KPI for predicting high-quality blastocyst formation had the highest association with ongoing pregnancy rates ($R^2 = 0.906$). This is the first AI-based cleavage-stage KPI demonstrated to detect changes in a culture environment that resulted in a shift in pregnancy outcomes.

Carrasco et al. (2017) used 800 cleavage-stage embryo images with decision tree methods and statistical analysis of features to determine the implantation potential of cleavage-stage embryos. *Wang et al. (2018)* extracted features from textures from 206 micrographs of early embryos (2 h of development). SVM were used (10-fold cross-validation) to achieve 77.7% accuracy and 0.78 of AUC

to predict the early embryo development stage (initial and Days 1, 2, 3 and 4).

Using CNN, *Meyer et al. (2020)* were able to classify Day 3 cleavage-stage embryo images as aneuploid or euploid with a high specificity and thus were able to sufficiently identify 85.5% of aneuploid embryos. These results demonstrate the ability of CNN to identify non-invasive markers for detecting genetically abnormal embryos. Collectively, these studies show that a variety of AI techniques can be used to extract unique features from cleavage-stage embryos, which may be used for classification, assessment ranking or to aid in clinical decision-making.

Kelly and colleagues used CNN to identify safe regions on a cleavage-stage embryo to perform laser-assisted hatching. This study utilized more than 13,000 annotated images of cleavage-stage embryos to develop an algorithm that identified the largest perivitelline space region or atretic/fragmented blastomeres. These regions of the cleavage-stage embryos were considered the safest at which to perform laser-assisted hatching. The AI-trained network was tested on almost 4000 cleavage-stage images and had 99.4% accuracy with a 95% CI ranging between 99.1% and 99.6% (*Kelly et al., 2020*).

Embryo witnessing is a critical step in the embryo transfer process. Traditionally, embryo identification is performed by two embryologists to ensure the correct embryo has been selected for transfer. However, as gametes and embryos are moved from one dish to another during an ART cycle, there is still the possibility of misidentification. Bormann and colleagues used CNN to classify images of embryos captured on Day 3 of development at 60 and 64 h post-insemination. The algorithm processed embryo images for each patient and produced a unique key that was associated with the patient ID at the initial evaluation. At the later time, images were captured and CNN were used to match the embryo morphology with the initial image. The accuracy of the CNN in correctly matching embryos at the different time periods on Day 3 was 100% (95% CI 99.1–100%, n = 412) (*Bormann et al., 2021b*). This technology offers a robust witnessing step based on unique morphological features that are specific to each individual embryo.

AI application on blastocyst-stage embryos

A key question about blastocyst assessment needs to be answered: When do we evaluate blastocysts? Because blastocyst development is a dynamic process, do we evaluate and grade blastocysts when they are exhibiting the 'best' appearance? Or should we evaluate them at a particular time? This question has yet to be answered by existing AI applications, which have used both fixed and flexible time-based methods of evaluation.

Another issue with blastocyst assessment involves grading. For instance, the problem with using Gardner-type blastocyst grading to assess embryo quality is that it is subjective and does not include quantitative parameters. It is a visual estimate of the number, size and morphology of the inner cell mass (ICM) and TE cells. On the other hand, blastocyst expansion can be easier to standardize if measurement tools and volume ratios are used. The quality of the ICM is estimated by the number and compaction of the cells. However, the minimum number of ICM cells necessary to develop into a viable human fetus is unknown. In addition, the ICM is a cocktail of pluripotent (epiblast) and primitive endoderm (hypoblast) cells. The size of the ICM alone does not indicate the composition of the cells within.

Assessing TE cells is more challenging, as the cell number, shape, nuclear content and position in the expanding blastocyst are not standardized. AI methods that use segmentation of the blastocyst will enable us to objectively score TE complement. It is easier to judge the compaction of the ICM than it is to assess TE quality.

The bigger question is, do we need to assess blastocysts at a particular time point? We know that Day 5 and Day 6 blastocysts have different outcomes, even when using fresh or frozen embryo transfer cycles (*Irani et al., 2018*). This is especially important to consider when developing AI algorithms that use a single 2D blastocyst image. We must consider the speed and timing of developmental events, particularly compaction and blastulation.

For successful implantation, both blastocyst cell types (ICM and TE) are required. Because current blastocyst grading systems are very simple, it is no surprise that they are not very informative when used to predict implantation. More complex and detailed blastocyst grading systems correlate very well with implantation potential and ploidy assessment. In their recent paper, *Zhan et al. (2020b)* converted alphanumeric blastocyst grades into a numeric score for use in statistical analysis and correlations. By using AI, it might be possible to strengthen the correlation between blastocyst assessment and outcome in a more objective manner. Also, the ability of AI blastocyst applications predicted by early developmental versus later developmental events needs to be explored.

Time-lapse microscopy (TLM) image analysis

AI algorithms can be applied to 'raw' TLM images. In a recently described image analysis system (*Tran et al., 2019*), supervised AI training using previously labelled images was developed. The labels used included blastocyst and morphokinetic annotations with positive or negative implantation results. One of the drawbacks of the system was its reliance on humans to create the labels, introducing biased observations and scores. The other problematic practice was the use of non-viable, non-fertilized or discarded material for negative training groups to increase the training dataset. The rationale behind this was the establishment of a completely automatic system that would also be able to recognize these negative embryos. The question remains, Will the developed algorithms perform equally well after removing the discarded group? And are they superior to the blastocyst grading system (*Kan-Tor et al., 2020b*)?

In another recent study, a different approach was used to predict blastocyst development. It used TLM data up to Day 3 of embryo development. Two different AI algorithms were developed: an automatic morphokinetic data model (temporal) and a TLM embryo image model (spatial). Both models have comparable predictive power (~0.7). When combined, the different weights were used to optimize blastocyst prediction. Interestingly, more weights were given to the morphokinetic data compared to the images. When compared to embryologists, the AI model performed better in terms of

sensitivity and specificity (*Liao et al., 2021*). In another TLM study, blastocyst prediction was accomplished by using morphokinetic TLM data from the first 3 days of development. Interestingly, by applying a self-improvement (reinforcement) strategy, the predictive power of the AI system improved (*d'Estaing et al., 2021*).

One unique approach to assessing blastocyst quality is to evaluate a quantitative standard expansion assay (qSEA) using AI. This measures the kinetics of blastocyst expansion and correlates to outcome, where faster-expanding blastocysts exhibit higher implantation potential (*Huang et al., 2021*).

The following novel embryo parameters have been proposed by *Bori et al. (2020*, 2021), to be included in AI selection models: pronuclear kinetics, blastocyst measurements, the size of the ICM and the cell cycle length of the TE cells. To verify the general utilization of their proposed model (donor oocytes), the authors' algorithm will need to be evaluated on the IVF patient population. The same group presented a novel model utilizing AI to predict embryo implantation. Using AI image analysis combined with the embryo proteomic profile of PGT euploid embryo spent culture media, the authors were able to demonstrate very high implantation prediction. Although the study is preliminary, it demonstrates the power of AI to combine different data points (proteins and morphology) (*Bori et al., 2020*, *2021*).

Static image analysis of blastocysts

The objective of a study by *Khosravi et al. (2019)* was to establish an AI deep learning model that can evaluate blastocyst quality. In this AI-based prediction model, the blastocyst expansion was an important parameter, followed by ICM and TE quality. The precise time point used for the AI evaluation (110 h) demonstrated the importance of embryo developmental kinetics for embryo prediction. In a 2020 study, a single image from the TLM image pool at 113 h was used for analysis (*Bormann et al., 2020a*). A CNN system was used to classify blastocysts based on the presence of the cavity and the morphological quality of the ICM and TE. Similar to *Khosravi et al. (2019)*, Bormann's group demonstrated that the accuracy of this system for classifying blastocysts versus non-blastocysts was very high (91%). By using the genetic algorithm, the authors established a blastocyst ranking system called the 'BL score'. The evaluation of the AI blastocyst selection method, using implantation outcomes of the blastocysts selected by humans for transfer, showed over 50% positive outcomes. It will be necessary to perform a comparative prospective study to identify the (dis)agreement in blastocyst selection for transfer between AI models and embryologists. The emerging question is how different the blastocyst selection for embryo transfer is between embryologists using the Gardner blastocyst grading system (*Gardner et al., 2000*) and AI model selection.

Bormann and colleagues demonstrated that the high degree of variability seen among embryologists making decisions on vitrification and embryo biopsy based on standard morphological assessments can be dramatically improved using deep neural networks (*Bormann et al., 2020b*). Souter *et al*. (2019) further demonstrated that deep learning CNN can be used to accurately identify which Day 3 assisted-hatched embryos met Day 5 criteria for TE biopsy and cryopreservation with 93.7% sensitivity and 96.3% specificity. This validation study was the first of its kind to demonstrate that an embryo decision-making algorithm could be successfully applied to embryos that had been artificially breached to promote premature herniation of TE cells for blastocyst-stage biopsy (Souter *et al*., 2019).

How many times will the AI choose a different blastocyst for embryo transfer than the embryologist within the cohort of available embryos? There is a lot of disagreement among embryologists grading blastocysts, but how many times is the best blastocyst chosen for embryo transfer? There are no standards in choosing an AI system for embryo evaluation. They depend on the type of data, the size of the dataset and the output queries (*Fernandez et al., 2020*). It will be helpful to compare multiple AI models on the same dataset.

Other AI models do not use a specific time point for image analysis. In the model by *VerMilyea et al. (2020)*, the 'viability' of the embryos was categorized based on the embryologist-given Gardner BL grade, where a '3BB' blastocyst was a cut-off for viable and non-viable classes using fetal heart measurements. Using computer vision image processing and deep learning, the authors achieved an overall accuracy of over 60% and an average accuracy improvement of 24% over embryologist grading.

Numerous complex neural network architectures have been proposed for image recognition and performance of these architectures are highly dependent on the requested task. *Thirumalaraju et al. (2021)* compared the use of eight different architectures to classify blastocyst-stage embryo images captured on a variety of imaging platforms. This study showed that Xception performed best in learning categorical embryo data and was able to accurately classify blastocysts based on their morphological quality. Xception correctly classified >99.5% of the highest quality blastocysts, which is of critical importance, clinically, when identifying embryos suited for transfer (*Thirumalaraju et al., 2021*).

Automated annotation of blastocysts

One of the potentially confounding factors that can affect AI protocols is the fact that the morphokinetic annotations are done by humans and are subjective. It will be necessary to develop AI models that can recognize abnormal karyokinetic (nuclear) and cytokinetic abnormalities (direct divisions 1–3, cell fusion) for optimal automatic annotation.

Most machine learning methods for embryo assessment and selection have used 'computer vision methods' using visual data (TLM or microscopic images). CNN is a method of choice to process visual information. It can be used for automatic cell annotation (*Malmsten et al., 2020*), cell detection and tracking (*Leahy et al., 2020*), blastocyst-stage identification and witnessing (*Kanakasabapathy et al., 2020b*), embryo grading and selection, and blastocyst and implantation prediction (*Louis et al., 2021*). Furthermore, *Dimitriadis et al. (2021)* used an AI implantation prediction model as a novel and unbiased morphology-based evaluation tool to assess the competencies of embryologists selecting embryos, performing vitrification and warming and of embryologists and physicians performing embryos transfers. It is important to note that these studies were done on retrospective data under

experimental settings. The clinical application of AI still requires prospective studies.

Implantation prediction

In a recent study, *Fitz et al. (2021)* sought to determine whether embryologists could improve their ability to select euploid embryos with the highest implantation potential with the aid of an AI-trained implantation algorithm. In this two-part study, embryologists from five separate laboratories were asked to select the top embryo for transfer from an image set of two embryos (n = 200 image sets). Next, they were provided with the same image set and a notation of which embryo was predicted to implant using AI. Embryologists were told that the AI implantation algorithm had a 75% accuracy, which could be incorporated into their embryo selection decision. All 14 embryologists participating in this study improved their ability to select the top-quality embryo when incorporating AI, with a mean improvement of 11.1% (range 1.4–15.5%) (*Fitz et al., 2021*). One limitation of this study is its retrospective nature.

In studies using AI to predict embryo implantation potential on static or TLM images, secondary factors such as laboratory conditions or other human factors have not been analysed or included in the models. Culture conditions and human expertise are important factors that influence embryo development and quality. To achieve a useful and objective prediction, these factors will need to be included in models. In addition, it is known that successful implantation and live birth depend on other factors not inherent to the embryo. Predicting implantation solely on embryo quality is an incomplete assessment. The focus of AI embryo prediction models should be the ranking of the embryos within the patient cohort rather than on implantation prediction. The variation in success rates among IVF centres and laboratories prevents the establishment of universal AI models for implantation prediction (*Zaninovic and Rosenwaks, 2020*).

How can AI-based models be used in the clinical laboratory setting and within laboratory workflows in a prospective way? First, AI models need to be evaluated in parallel with standard laboratory embryo selection practice. Second, prospective studies of embryo selection by machine and human need to be performed.

AI for non-invasive ploidy screening

PGT for aneuploidies (PGT-A) remains the most objective way to assess an embryo. However, its invasive nature, cost and the assumption of diagnostic accuracy limit a more widespread use. It is no surprise that non-invasive approaches to embryo selection, including time-lapse morphokinetic evaluation (*Campbell et al., 2013*), morphology assessment (*Capalbo et al., 2014*; *Zhan et al., 2020a*) and AI systems (*Pennetta et al., 2018*; *Meyer et al., 2020*) have aimed to compare PGT-A outcomes against their findings. However, it is still difficult to find studies presenting AI systems for embryo ranking that are trained against ploidy status as their ground truth.

The first published study of this kind was most likely by *Chavez-Badiola et al. (2020a)*, in which the authors trained and tested an AI model called ERICA to rank embryos based on its ability to predict euploidy, using a single static blastocyst image as the only source of information. Following training and validation on 1231 images from three IVF centres, the ERICA device showed significantly better prediction capabilities (70% overall accuracy for euploidy prediction) than chance and the embryologists involved in the study. It is important to acknowledge that despite seniority and experience, conclusions on the device's superiority cannot be drawn based on a comparison against the performance of only two embryologists. As the authors acknowledge, a larger testing set, as well as a larger number of embryologists with different levels of experience and seniority, would be required to confirm the study's results. At this point, however, the results are encouraging enough to suggest that ERICA has the potential to assist embryologists and clinicians during embryo selection in a non-invasive fashion (*Chavez-Badiola et al., 2020a*).

We can anticipate that other similar full-paper publications will follow shortly, presenting new approaches aimed at embryo selection based on ploidy. These studies will perhaps target time-lapse sequences (*Barnes et al., 2020*) and incorporate omics (*Bori et al., 2021*), patient and cycle characteristics (*Jiang et al., 2021*), non-invasive chromosome screening tests (*Chavez-Badiola et al., 2020c*), as well as new AI approaches. Building high-quality datasets from diverse settings – while managing hype (*VerMilyea et al., 2019*) and expectations – are challenges that will remain.

CONCLUSION

AI has long been utilized in other industries and has recently found a place in medical imaging; however, it is just beginning to have an impact on the clinical practice of reproductive medicine, a field familiar to rapid advancements and open to using new technologies to achieve the ultimate goal of a healthy baby.

Because there are over 2 million IVF cycles performed annually throughout the world, and with IVF being a medical procedure globally registered, one can only hope that the data collection from throughout the years will help to develop AI systems that are widely applicable across clinics and independent of differences in protocols and populations. Barriers to achieving this include health record privacy terms, paper records and variations in electronic medical record systems.

AI systems developed thus far for the field of reproductive medicine have focused primarily on the use of embryo imaging and have been summarized here. However, AI has the potential to assist in other areas of reproductive medicine as well, including endometrial receptivity, uterine function, fertility impact of diseases such as endometriosis and adenomyosis, recurrent implantation failure, and recurrent pregnancy loss (*Curchoe, 2021*). In summary, AI has the potential to be used as a promising tool to resolve many longstanding challenges in the field of reproductive medicine, as well as to help clinicians make decisions and achieve the ultimate goal of a healthy live-born baby. However, at present, AI has not established its role in the world of reproductive medicine, and it is important to keep in mind that its use in improving outcomes is not, as of yet, proven in the literature. Further studies, ideally randomized controlled, are required, to identify indicated use of this very promising tool.

REFERENCES

Abràmoff, M.D., Lou, Y., Erginay, A., Clarida, W., Amelon, R., Folk, J.C., Niemeijer, M **Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning.** Invest. Ophthalmol. Vis. Sci. 2016; 57: 5200–5206

Azzarello, A., Hoest, T., Mikkelsen, A.L. **The impact of pronuclei morphology and dynamicity on live birth outcome after time-lapse culture.** Hum. Reprod. 2012; 27: 2649–2657. doi:10.1093/humrep/des210

Bancsi, L.F., Broekmans, F.J., Looman, C.W., Habbema, J.D., te Velde, E.R. **Impact of repeated antral follicle counts on the prediction of poor ovarian response in women undergoing in vitro fertilization.** Fertil. Steril. 2004; 81: 35–41

Barnes, J., Malmsten, J., Zhan, O., Hajirasouliha, I., Elemento, O., Sierra, J., Zaninovic, N., Rosenwaks, Z. **Noninvasive detection of blastocyst ploidy (euploid vs. aneuploid) using artificial intelligence (AI) with deep learning methods.** Fertil. Steril. 2020; 114

Bengio, Y., Courville, A., Vincent, P. **Representation learning: a review and new perspectives.** IEEE Trans. Pattern Anal. Mach. Intell. 2013; 35: 1798–1828. doi:10.1109/tpami.2013.50

Bishop, C. 2006 **Pattern Recognition and Machine Learning.** Springer Berlin

Bori, L., Dominguez, F., Fernandez, E.I., Del Gallego, R., Alegre, L., Hickman, C., Quiñonero, A., Nogueira, M.F.G., Rocha, J.C., Meseguer, M. **An artificial intelligence model based on the proteomic profile of euploid embryos and blastocyst morphology: a preliminary study.** Reprod. Biomed. Online 2021; 42: 340–350. doi:10.1016/j.rbmo.2020.09.031

Bori, L., Paya, E., Alegre, L., Viloria, T.A., Remohi, J.A., Naranjo, V., Meseguer, M. **Novel and conventional embryo parameters as input data for artificial neural networks: an artificial intelligence model applied for prediction of the implantation potential.** Fertil. Steril. 2020; 114: 1232–1241

Bormann, C.L., Curchoe, C.L., Thirumalaraju, P., Kanakasabapathy, M.K., Gupta, R., Pooniwala, R., Kandula, H., Souter, S., Dimitriadis, I., Shafiee, H. **Deep learning early warning system for embryo culture conditions and embryologist performance in the ART laboratory.** J. Assist. Reprod. Genet. 2021; 38: 1641–1646. doi:10.1007/s10815-021-02198-x

Bormann, C.L., Kanakasabapathy, M.K., Thirumalaraju, P., Dimitriadis, I., Souter, I., Hammer, K., Shafiee, H. **Development of an artificial intelligence embryo witnessing system to accurately track and identify patient specific embryos in a human IVF laboratory.** Hum. Reprod. 2021; 36. doi:10.1093/humrep/deab126.050

Bormann, C.L., Kanakasabapathy, M.K., Thirumalaraju, P., Gupta, R., Pooniwala, R., Kandula, H., Hariton, E., Souter, I., Dimitriadis, I., Ramirez, L.B., Curchoe, C.L., Swain, J., Boehnlein, L.M., Shafiee, H. **Performance of a deep learning based neural network in the selection of human blastocysts for implantation.** eLife 2020; 9: e55301. doi:10.7554/eLife.55301

Bormann, C.L., Thirumalaraju, P., Kanakasabapathy, M.K., Kandula, H., Souter, I., Dimitriadis, I., Gupta, R., Pooniwala, R., Shafiee, H. **Consistency and objectivity of automated embryo assessments using deep neural networks.** Fertil. Steril. 2020; 113: 781–787

Bortoletto, P., Kanakasabapathy, M.K., Thirumalaraju, P., Gupta, R., Pooniwala, R., Souter, I., Dimitriadis, I., Dickinson, K., Hariton, E., Bormann, C.L., Shafiee, H. **Predicting blastocyst formation of day 3 embryos using a convolutional neural network (CNN): a machine learning approach.** Fertil. Steril. 2019; 112: e272–e273

Burkov, A., 2019. The Hundred-Page Machine Learning Book. Available from: http://ema.cri-info.cm/wp-content/uploads/2019/07/2019BurkovTheHundred-pageMachineLearning.pdf

Campbell, A., Fishel, S., Bowman, N., Duffy, S., Sedler, M., Hickman, C.F. **Modelling a risk classification of aneuploidy in human embryos using non-invasive morphokinetics.** Reprod. BioMed. Online 2013; 26: 477–485

Capalbo, A., Rienzi, L., Cimadomo, D., Maggiulli, R., Elliott, T., Wright, G., Nagy, Z.P., Ubaldi, F.M. **Correlation between standard blastocyst morphology, euploidy and implantation: an observational study in two centers involving 956 screened blastocysts.** Hum. Reprod. 2014; 29: 1173–1181

Carrasco, B., Arroyo, G., Gil, Y., Gómez, M.J., Rodríguez, I., Barri, P.N., Veiga, A., Boada, M. **Selecting embryos with the highest implantation potential using data mining and decision tree based on classical embryo morphology and morphokinetics.** J. Assist. Reprod. Genet. 2017; 34: 983–990. doi:10.1007/s10815-017-0955-x

Chavez-Badiola, A., Flores-Saiffe-Farías, A., Mendizabal-Ruiz, G., Drakeley, A.J., Cohen, J. **Embryo Ranking Intelligent Classification Algorithm (ERICA): artificial intelligence clinical assistant predicting embryo ploidy and implantation.** Reprod. Biomed. Online 2020; 41: 585–593. doi:10.1016/j.rbmo.2020.07.003

Chavez-Badiola, A., Flores-Saiffe Farias, A., Mendizabal-Ruiz, G., Garcia-Sanchez, R., Drakeley, A.J., Garcia-Sandoval, J.P. **Predicting pregnancy test results after embryo transfer by image feature extraction and analysis using machine learning.** Sci. Rep. 2020; 10: 4394. doi:10.1038/s41598-020-61357-9

Chavez-Badiola, A., Zhang, J., Flores-Saiffe-Farías, A., Mendizabal-Ruiz, G., Drakeley, A.J., Olcha, M. **Non-invasive chromosome screening and its correlation against ranking prediction made by ERICA, a deep-learning embryo ranking algorithm.** Fertil. Steril. 2020; 114: e436–e437

Chawla, N.V., Japkowicz, N., Kotcz, A. **Editorial: special issue on learning from imbalanced data sets.** ACM SIGKDD Explorations Newsletter 2004; 6: 1–6

Chen, M., Wei, S., Hu, J., Yuan, J., Liu, F. **Does time-lapse imaging have favorable results for embryo incubation and selection compared with conventional methods in clinical in vitro fertilization? A meta-analysis and systematic review of randomized controlled trials.** PLoS One 2017; 12e0178720

Chen, R.J., Lu, M.Y., Chen, T.Y., Williamson, D.F.K., Mahmood, F **Synthetic data in machine learning for medicine and healthcare.** Nat. Biomed. Eng. 2021; 5: 493–497. doi:10.1038/s41551-021-00751-8

Chen, T.J., Zheng, W., Liu, C.H., Huang, I., Lai, H., Liu, M. **Using deep learning with large dataset of microscope images to develop an automated embryo grading system.** Fertil. Reprod. 2019; 01: 51–56

Chicco, D. **Ten quick tips for machine learning in computational biology.** BioData Mining 2017; 10: 35. doi:10.1186/s13040-017-0155-3

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. **Improving embryo selection using a computer-automated time-lapse image analysis test plus day 3 morphology: results from a prospective multicenter trial.** Fertil. Steril. 2013; 100: 412–419. doi:10.1016/j.fertnstert.2013.04.021

Cortes, C., Vapnik, V.N. **Support-vector networks.** Machine Learning 1995; 20: 273–297. doi:10.1007/BF00994018

Coticchio, G., Fiorentino, G., Nicora, G., Sciajno, R., Cavalera, F., Bellazzi, R., Garagna, S., Borini, A., Zuccotti, M. **Cytoplasmic movements of the early human embryo: imaging and artificial intelligence to predict blastocyst development.** Reprod. Biomed. Online 2021; 42: 521–528

Cruz, M., Garrido, N., Herrero, J., Perez-Cano, I., Munoz, M., Meseguer, M. **Timing of cell division in human cleavage-stage embryos is linked with blastocyst formation and quality.** Reprod. Biomed. Online 2012; 25: 371–381. doi:10.1016/j.rbmo.2012.06.017

Curchoe, C.L. **The paper chase and the big data arms race.** J. Assist. Reprod. Genet. 2021; 38: 1613–1615. doi:10.1007/s10815-021-02122-3

Curchoe, C.L., Bormann, C.L. **Artificial intelligence and machine learning for human reproduction and embryology presented at ASRM and ESHRE 2018.** J. Assist. Reprod. Genet. 2019; 36: 591–600. doi:10.1007/s10815-019-01408-x

Curchoe, C.L., Flores-Saiffe Farias, A., Mendizabal-Ruiz, G., Chavez-Badiola, A. **Evaluating predictive models in reproductive medicine.** Fertil. Steril. 2020; 114: 921–926. doi:10.1016/j.fertnstert.2020.09.159

D'Amour, A., Heller, K., Moldovan, D., Adlam, B., Alipanahi, B., Beutel, A., Chen, C., Deaton, J., Eisenstein, J., Hoffman, M.D., Hormozdiari, F., Houlsby, N., Hou, S., Jerfel, G., Karthikesalingam, A., Lucic, M., Ma, Y., McLean, C., Mincu, D., Mitani, A., Montanari, A., Nado, Z., Natarajan, V., Nielson, C., Osborne, T.F., Raman, R., Ramasamy, K., Sayres, R., Schrouff, J., Seneviratne, M., Sequeira, S., Suresh, H., Veitch, V., Vladymyrov, M., Wang, X., Webster, K., Yadlowsky, S., Yun, T., Zhai, X., Sculley, D. Underspecification presents challenges for credibility in modern machine learning. 2020; arXiv:2011.03395v2

d'Estaing, S.G., Labrune, E., Forcellini, M., Edel, C., Salle, B., Lornage, J., Benchaib, M. **A machine learning system with reinforcement capacity for predicting the fate of an ART embryo.** Syst. Biol. Reprod. Med. 2021; 67: 64–78

Dickinson, J., Meyer, A., Kelly, N., Thirumalaraju, P., Kanakasabapathy, M., Kartik, D., Bormann, C.L., Shafiee, H. **Advancement in the future automation of ICSI: use of deep convolutional neural networks (CNN) to identify precise location to inject sperm in mature human oocytes.** Hum. Reprod. 2020; 35: 70–71

Dimitriadis, I., Bormann, C.L., Kanakasabapathy, M.K., Thirumalaraju, P., Kandula, H., Yogesh, V., Gudipati, N., Natarajan, V., Petrozza, J.C., Shafiee, H. **Automated smartphone-based**

system for measuring sperm viability, DNA fragmentation, and hyaluronic binding assay score. PLoS One 2019; 14e0212562. doi:10.1371/journal.pone.0212562

Dimitriadis, I., Bormann, C.L., Kanakasabapathy, M.K., Thirumalaraju, P., Gupta, R., Pooniwala, R., Souter, I., Rice, S.T., Bhowmick, P., Shafiee, H. **Deep convolutional neural networks (CNN) for assessment and selection of normally fertilized human embryos.** Fertil. Steril. 2019; 112: e272

Dimitriadis, I., Christou, G., Dickinson, K., McLellan, S., Brock, M., Souter, I., Bormann, C.L. **Cohort embryo selection (CES): a quick and simple method for selecting cleavage stage embryos that will become high quality blastocysts (HQB).** Fertil. Steril. 2017; 108: e162–e163. doi:10.1016/j.fertnstert.2017.07.488

Dimitriadis, I., Kanakasabapathy, M.K., Thirumalaraju, P., Vagios, S., Souter, I., Shafiee, H., Bormann, C.L. **Artificial intelligence assistance for the quality assessment of embryo vitrification, warming and transfers in the IVF laboratory.** ASRM 2021 Scientific Congress & Expo 2021

Esteva, A., Kuprel, B., Novoa, R. **Dermatologist-level classification of skin cancer with deep neural networks.** Nature 2017; 542: 115–118. doi:10.1038/nature21056

Fernandez, E.I., Ferreira, A.S., Miquelão Cecílio, M.H., Chéles, D.S., Milanezi de Souza, R.C., Gouveia Nogueira, M.F., Rocha, J.C. **Artificial intelligence in the IVF laboratory: overview through the application of different types of algorithms for the classification of reproductive data.** J. Assist. Reprod. Genet. 2020; 37: 2359–2376

Fitz, V.W., Kanakasabapathy, M.K., Thirumalaraju, P. **Should there be an "AI" in TEAM? Embryologists selection of high implantation potential embryos improves with the aid of an artificial intelligence algorithm.** J. Assist. Reprod. Genet 2021; 38: 2663–2670. doi:10.1007/s10815-021-02318-7

Gardner, D.K., Lane, M., Stevens, J., Schlenker, T., Schoolcraft, W.B. **Blastocyst score affects implantation and pregnancy outcome: towards a single blastocyst transfer.** Fertil. Steril. 2000; 73: 1155–1158

Geisser, S. 1993 **Predictive Inference: An Introduction.** Chapman & Hall/CRCBoca Raton, FL, USA

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y. **Generative adversarial nets.** Proceedings of the International Conference on Neural Information Processing Systems 2014: 2672–2680

Goodson, S.G., White, S., Stevans, A.M., Bhat, S., Kao, C.Y., Jaworski, S., Marlowe, T.R., Kohlmeier, M., McMillan, L., Zeisel, S.H., O'Brien, D.A. **CASAnova: a multiclass support vector machine model for the classification of human sperm motility patterns.** Biol. Reprod. 2017; 97: 698–708. doi:10.1093/biolre/iox120

Hickman, C.F.L., Alshubbar, H., Chambost, J., Jacques, C., Pena, C.A., Drakeley, A., Freour, T. **Data sharing: using blockchain and decentralized data technologies to unlock the potential of artificial intelligence: what can assisted reproduction learn from other areas of medicine?** Fertil. Steril. 2020; 114: 927–933. doi:10.1016/j.fertnstert.2020.09.160

Hinton, J., Sejnowski, .T 1999 **Unsupervised Learning: Foundations of Neural Computation.** MIT Press Cambridge, MA

Hlinka, D., Kaľatová, B., Uhrinová, I., Dolinská, S., Rutarová, J., Rezáčová, J., Lazarovská, S., Dudáš, M. **Time-lapse cleavage rating predicts human embryo viability.** Physiol. Res. 2012; 61: 513–525

Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L.H., Aerts, H. **Artificial intelligence in radiology.** Nat. Rev. Cancer 2018; 18: 500–510. doi:10.1038/s41568-018-0016-5

Huang, T.T.F., Kosasa, T., Walker, B., Arnett, C., Huang, C.T.F., Yin, C., Harun, M.Y., Ahn, H.J., Ohta, A. **Application of a deep learning neural network for the analysis of human blastocyst expansion from defined time-lapse image files.** Reprod. BioMed. Online 2021

Hunault, C.C., Eijkemans, M.J., Pieters, M.H., te Velde, E.R., Habbema, J.D., Fauser, B.C. **A prediction model for selecting patients undergoing in vitro fertilization for elective single embryo transfer.** Fertil. Steril. 2002; 77: 725–732

Irani, M., O'Neill, C., Palermo, G.D., Xu, K., Zhang, C., Qin, X., Zhan, Q., Clarke, R.N., Ye, Z., Zaninovic, N., Rosenwaks, Z. **Blastocyst development rate influences implantation and live birth rates of similarly graded euploid blastocysts.** Fertil. Steril. 2018; 110: 95–102

Jiang, V.S., Kanakasabapathy, M.K., Thirumalaraju, P., Kandula, H., Souter, I., Dimitriadis, I., Bormann, C.L., Shafiee, H. **The use of voting ensembles and patient characteristics to improve the accuracy of deep neural networks as a non-invasive method to classify embryo ploidy status.** ASRM 2021 Scientific Congress & Expo 2021 Oct 19

Jordan, M.I., Mitchell, T.M. **Machine learning: trends, perspectives, and prospects.** Science 2015; 349: 255–260

Jurisica, I., Mylopoulos, J., Glasgow, J., Shapiro, H., Casper, R.F. **Case-based reasoning in IVF: prediction and knowledge mining.** Artif. Intell. Med. 1998; 12: 1–24

Kaelbling, L.P., Littman, M.L., Moore, A.W. **Reinforcement learning: a survey.** J. Artif. Intell. Res. 1996; 4: 237–285

Kamiński, B., Jakubczyk, M., Szufel, P. **A framework for sensitivity analysis of decision trees. Cent.** Eur. J. Oper. Res. 2017; 26: 135–159. doi:10.1007/s10100-017-0479-6

Kanakasabapathy, M.K., Bormann, C.L., Thirumalaraju, P., Banerjee, R., Shafiee, H. **Improving the performance of deep convolutional neural networks (CNN) in embryology using synthetic machine-generated images.** Hum. Reprod. 2020; 35

Kanakasabapathy, M.K., Hammer, K.C., Dickinson, K., Veiga, C., Kelly, F., Thirumalaraju, P., Bormann, C.L., Shafiee, H. **Using artificial intelligence to avoid human error in identifying embryos.** Fertil. Steril. 2020; 113: e45

Kanakasabapathy, M.K., Sadasivam, M., Singh, A., Preston, C., Thirumalaraju, P., Venkataraman, M., Bormann, C.L., Draz, M.S., Petrozza, J.C., Shafiee, H. **An automated smartphone-based diagnostic assay for point-of-care semen analysis.** Sci. Transl. Med. 2017; 9: 382. doi:10.1126/scitranslmed.aai7863

Kanakasabapathy, M.K., Thirumalaraju, P., Bormann, C.L., Gupta, R., Pooniwala, R., Kandula, H., Souter, I., Dimitriadis, I., Shafiee, H. **Deep learning mediated single time-point image-based prediction of embryo developmental outcome at the cleavage stage.** arXiv 2020; 08346: 5–21

Kanakasabapathy, M.K., Thirumalaraju, P., Kandula, H., Doshi, F., Sivakumar, A.D., Kartik, D., Gupta, R., Pooniwala, R., Branda, J.A., Tsibris, A.M., Kuritzkes, D.R., Petrozza, J.C., Bormann, C.L., Shafiee, H. **Adaptive adversarial neural networks for the analysis of lossy and domain-shifted datasets of medical images.** Nat. Biomed. Eng. 2021; 5: 571–585

Kan-Tor, Y., Ben-Meir, A., Buxboim, A. **Can deep learning automatically predict fetal heart pregnancy with almost perfect accuracy?** Hum. Reprod. 2020; 35: 1473 doi.org/10.1093/humrep/deaa083

Kan-Tor, Y., Zabari, N., Erlich, I., Szeskin, A., Amitai, T., Richter, D., Or, Y., Shoham, Z., Hurwitz, A., Har-Vardi, I., Gavish, M., Ben-Meir, A., Buxboim, A. **Automated evaluation of human embryo blastulation and implantation potential using deep-learning.** Adv. Intell. Syst. 2020; 2: 1–12

Kaser, D.J., Bormann, C.L., Missmer, S.A., Farland, L.V., Ginsburg, E.S., Racowsky, C. **Eeva™ pregnancy pilot study: a randomized controlled trial of single embryo transfer (SET) on day 3 or day 5 with or without time-lapse imaging (TLI) selection.** Fertil. Steril. 2016; 106: e312. doi:10.1016/j.fertnstert.2016.07.886

Kelly, N., Bormann, C.L., Dickinson, J., Meyer, A.D., Kanakasabapathy, M.K., Thirumalaraju, P., Shafiee, H. **Future of automation: use of deep convolutional neural networks (CNN) to identify precise location to perform laser assisted hatching on human cleavage stage embryos.** Fertil. Steril. 2020; 114: e144

Khosravi, P., Kazemi, E., Imielinski, M., Elemento, O., Hajirasouliha, I. **Deep convolutional neural networks enable discrimination of heterogeneous digital pathology images.** EBioMedicine 2018; 27: 317–328

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., Elemento, O., Zaninovic, N. **Hajirasouliha, I. Deep learning enables robust assessment and selection of human blastocysts after in vitro fertilization.** NPJ Digit. Med. 2019; 2: 21

Kianpour, M., Wen, S.F. **Timing attacks on machine learning: state of the art.** Bi Y., Bhatia R., Kapoor S. Intelligent Systems and Applications. IntelliSys 2019. Advances in Intelligent Systems and Computing, Vol. 1037 2020: 111–125

Kirkegaard, K., Ahlström, A., Ingerslev, H.J., Hardarson, T. **Choosing the best embryo by time lapse versus standard morphology.** Fertil. Steril. 2015; 103: 323–332. doi:10.1016/j.fertnstert.2014.11.003

Kragh, M.F., Rimestad, J., Berntsen, J., Karstoft, H. **Automatic grading of human blastocysts from time-lapse imaging.** Comp. Biol. Med. 2019; 115103494. doi:10.1016/j.compbiomed.2019.103494

Lan, K.C., Huang, F.J., Lin, Y.C., Kung, F.T., Hsieh, C.H., Huang, H.W., Tan, P.H., Chang, S.Y. **The predictive value of using a combined Z-score and day 3 embryo morphology score in the assessment of embryo survival on day 5.** Hum. Reprod. 2003; 18: 1299–1306

Lancashire, L.J., Lemetre, C., Ball, G.R. **An introduction to artificial neural networks**

in bioinformatics – application to complex microarray and mass spectrometry datasets in cancer studies. Brief Bioinform 2009; 10: 315–329. doi:10.1093/bib/bbp012

Lavin, A., Gilligan-Lee, C.M., Visnjic, A., Ganju, S., Newman, D., Ganguly, S., Lange, D., Güneş Baydin, A., Sharma, A., Gibson, A., Gal, Y., Xing, E.P., Mattmann, C., Parr, J. **Technology readiness levels for machine learning systems.** arXiv Jan. 2021; 11

Leahy, B.D., Jang, W.D., Yang, H.Y., Struyven, R., Wei, D., Sun, Z., Lee, K.R., Royston, C., Cam, L., Kalma, Y., Azem, F., Ben-Yosef, D., Pfister, H., Needleman, D. **Automated measurements of key morphological features of human embryos for IVF.** Med. Image Comput. Assist. Interv. 2020; 12265: 25–35

Lechniak, D., Pers-Kamczyc, E., Pawlak, P. **Timing of the first zygotic cleavage as a marker of developmental potential of mammalian embryos.** Biol. Reprod. 2008; 8: 23–42

Lemmen, J.G., Agerholm, I., Ziebe, S. **Kinetic markers of human embryo quality using time-lapse recordings of IVF/ICSI-fertilized oocytes.** Reprod. Biomed Online 2008; 17: 385–391

Lemoigne, Y., Caner, A. Molecular Imaging: Computer Reconstruction and Practice. 2006

Letterie, G., MacDonald, A. **Artificial intelligence in in vitro fertilization: a computer decision support system for day-to-day management of ovarian stimulation during in vitro fertilization.** Fertil. Steril. 2020; 114: 1026–1031. doi:10.1016/j.fertnstert.2020.06.006

Liao, Q., Zhang, Q., Feng, X., Huang, H., Xu, H., Tian, B., Liu, J., Yu, Q., Guo, N., Liu, Q., Huang, B., Ma, D., Ai, J., Xu, S., Li, K. **Development of deep learning algorithms for predicting blastocyst formation and quality by time-lapse monitoring.** Commun. Biol. 2021; 4: 415. doi:10.1038/s42003-021-01937-1

Liao, S., Pan, W., Dai, W.Q., Jin, L., Huang, G., Wang, R., Hu, C., Pan, W., Tu, H. **Development of a dynamic diagnosis grading system for infertility using machine learning.** JAMA Netw. Open. 2020; 3e2023654. doi:10.1001/jamanetworkopen.2020.23654

Louis, C.M., Erwin, A., Handayani, N., Polim, A.A., Boediono, A., Sini, I. **Review of computer vision application in in-vitro fertilization: the application of deep learning-based computer vision technology in the world of IVF.** J. Assist. Reprod. Genet. 2021; 38: 1627–1639

Mahadevaiah, G., RV, P., Bermejo, I., Jaffray, D., Dekker, A., Wee, L. **Artificial intelligence-based clinical decision support in modern medical physics: selection, acceptance, commissioning, and quality assurance.** Med. Phys. 2020; 47: e228–e235

Malik, A., Patel, P., Ehsan, L., Guleria, S., Hartka, T., Adewole, S., Sana, S. **Ten simple rules for engaging with artificial intelligence in biomedicine.** PLoS Comput. Biol. 2021; 17e1008531. doi:10.1371/journal.pcbi.1008531

Malmsten, J., Zaninovic, N., Zhan, Q., Rosenwaks, Z., Shan, J. **Automated cell division classification in early mouse and human embryos using convolutional neural networks.** Neural Comput. Applic. 2020: 1–12

Manna, C., Nanni, L., Lumini, A., Pappalardo, S. **Artificial intelligence techniques for embryo and oocyte classification.** Reprod. Biomed. Online 2013; 26: 42–49. doi:10.1016/j.rbmo.2012.09.015

Meseguer, M., Valera, M.Á. **The journey toward personalized embryo selection algorithms.** Fertil. Steril. 2021; 115: 898–899. doi:10.1016/j.fertnstert.2021.01.039

Meyer, A., Dickinson, J., Kelly, N., Kandula, H., Kanakasabapathy, M., Thirumalaraju, P., Bormann, C.L., Shafiee, H. **Can deep convolutional neural network (CNN) be used as a non-invasive method to replace preimplantation genetic testing for aneuploidy (PGT-A)?** Hum. Reprod. 2020; 35

Mirsky, S.K., Barnea, I., Levi, M., Greenspan, H., Shaked, N.T. **Automated analysis of individual sperm cells using stain-free interferometric phase microscopy and machine learning.** Cytometry Part A 2017; 91A: 893–900. doi:10.1002/cyto.a.23189

Miyagi, Y., Habara, T., Hirata, R., Hayashi, N. **Feasibility of artificial intelligence for predicting live birth without aneuploidy from a blastocyst image.** Reprod. Med. Biol. 2019; 18: 204–211. doi:10.1002/rmb2.12267

Mohri, M., Rostamizadeh, A., Talwalkar, A. 2012 **Foundations of Machine Learning.** MIT Press Cambridge, MA

Patki, N., Wedge, R.,, Veeramachaneni, K. **The Synthetic Data Vault.** 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA) 2016: 399–410. doi:10.1109/DSAA.2016.49

Payne, D., Flaherty, S.P., Barry, M.F., Matthews, C.D **Preliminary observations on polar body extrusion and pronuclear formation in human oocytes using time-lapse video cinematography.** Hum. Reprod. 1997; 12: 532–541

Pennetta, F., Lagalla, C., Borini, A. **Embryo morphokinetic characteristics and euploidy.** Curr. Opin. Obstet. Gynecol. 2018; 30: 185–196. doi:10.1097/GCO.0000000000000453

Poole, D., Mackworth, A., Goebel, R. 1998 **Computational Intelligence: A Logical Approach.** Oxford University Press New York

Prados, F.J., Debrock, S., Lemmen, J.G., Agerholm, I. **The cleavage stage embryo.** Hum. Reprod. 2012; 27: i50–i71. doi:10.1093/humrep/des224

Sacha, C.R., Vagios, S., Souter, I., Kanakasabapathy, M.K., Thirumalaraju, P., Shafiee, H., Bormann, C.L. **Maturity of oocyte cohort impacts blastocyst development as classified by artificial intelligence (AI).** ASRM 2021 Scientific Congress & Expo 2021 Oct 19

Scott, L.A., Smith, S. **The successful use of pronuclear embryo transfers the day following oocyte retrieval.** Hum. Reprod. 1998; 13: 1003–1013

Scott, L., Alvero, R., Leondires, M., Miller, B. **The morphology of human pronuclear embryos is positively related to blastocyst development and implantation.** Hum. Reprod. 2000; 15: 2394–2403

Setti, A.S., Figueira, R.C., Braga, D.P., Colturato, S.S., Iaconelli, A.Jr., Borges, E. Jr. **Relationship between oocyte abnormal morphology and intracytoplasmic sperm injection outcomes: a meta-analysis.** Eur. J. Obstet. Gynecol. Reprod. Biol. 2011; 159: 364–370. doi:10.1016/j.ejogrb.2011.07.031

Shapiro, L.G., Stockman, G.C. 2001 **Computer Vision.** Prentice-HallNew Jersey: 279–325

Shorten, C., Khoshgoftaar, T.M. **A survey on Image Data Augmentation for Deep Learning.** J. Big Data 2019; 6: 60. doi:10.1186/s40537-019-0197-0

Sonka, M., Hlavac, V., Boyle, R. 2008 **Image Processing, Analysis, and Machine Vision.** Thomson

Souter, I., Vagios, S., Sacha, C., Kanakasabapathy, MK., Thirumalaraju, P., Bormann, CL., Shafiee, H. **Validation of a deep convolutional neural network trained to assess whether an embryo meets criteria for biopsy/cryopreservation using a test set of PGT embryos.** Hum. Reprod. 2020; 35: 1257–1258

Swain, J., VerMilyea, M.T., Meseguer, M., Ezcurra, D. **Fertility AI Forum Group. AI in the treatment of fertility: key considerations.** J. Assist. Reprod. Genet. 2020; 37: 2817–2824

Targosz, A., Przystałka, P., Wiaderkiewicz, R., Mrugacz, G. **Semantic segmentation of human oocyte images using deep neural networks.** Biomed. Eng. Online 2021; 20: 40

Tesarik, J., Greco, E. **The probability of abnormal preimplantation development can be predicted by a single static observation on pronuclear stage morphology.** Hum. Reprod. 1999; 14: 1318–1323

Thirumalaraju, P., Kanakasabapathy, M.K., Bormann, C.L., Gupta, R., Pooniwala, R., Kandula, H., Souter, I., Dimitriadis, I., Shafiee, H. **Evaluation of deep convolutional neural networks in classifying human embryo images based on their morphological quality.** Heliyon 2021; 7: E06298

Thirumalaraju, P., Kanakasabapathy, M.K., Bormann, C.L., Kandula, H., Kota, S., Pavan, S., Yarravarapu, D., Shafiee, H. **Human sperm morphology analysis using smartphone microscopy and deep learning.** Fertil. Steril. 2019; 112: e41

Thirumalaraju, P., Kanakasabapathy, M.K., Gupta, R., Pooniwala, R., Kandula, H., Souter, I., Dimitriadis, I., Bormann, C.L., Shafiee, H. **Automated quality assessment of individual embryologists performing ICSI using deep learning-enabled fertilization and embryo grading technology.** Fertil. Steril. 2019; 112: e71

Tiitinen, A. **Single embryo transfer: why and how to identify the embryo with the best developmental potential.** Best Pract. Res. Clin. Endocrinol. Metab. 2019; 33: 77–88

Tran, D., Cooke, S., Illingworth, P.J., Gardner, D.K. **Deep learning as a predictive tool for fetal heart pregnancy following time-lapse incubation and blastocyst transfer.** Hum. Reprod. 2019; 34: 1011–1018. doi:10.1093/humrep/dez064

Tu, J.V. **Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes.** J. Clin. Epidemiol. 1996; 49: 1225–1231. doi:10.1016/S0895-4356(96)00002-9

Uzair, M., Jamil, N. **Effects of hidden layers on the efficiency of neural networks.** IEEE 23rd International Multitopic Conference (INMIC) 2020: 1–6. doi:10.1109/INMIC50486.2020.9318195

van Weert, J.M., Repping, S., van der Steeg, J.W., Steures, P., van der Veen, F., Mol, B.W **A prediction model for ongoing pregnancy after in vitro fertilization in couples with male subfertility.** J. Reprod. Med. 2008; 53: 250–256

VerMilyea, M., Hall, J.M.M., Diakiw, S.M., Johnston, A., Nguyen, T., Perugini, D., Miller, A., Picou, A., Murphy, A.P., Perugini, M. **Development of an artificial intelligence-based assessment model for prediction of embryo viability using static images captured**

by optical light microscopy during IVF. Hum. Reprod. 2020; 35: 770–784. doi:10.1093/humrep/deaa013

VerMilyea, M., Hall, J.M.M., Perugini, D., Murphy, A.P., Nguyen, T., Rios, C. **Artificial intelligence: non-invasive detection of morphological features associated with abnormalities in chromosomes 21 and 16.** Fertil. Steril. 2019; 112: e237–e238

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. **Computer-automated time-lapse analysis results correlate with embryo implantation and clinical pregnancy: a blinded, multi-centre study.** Reprod. Biomed. Online 2014; 29: 729–736. doi:10.1016/j.rbmo.2014.09.005

Wang, L.H., Fu, Z.X., Ye, S.Z., Ke, D.G. **Pattern recognition of IVF's early embryo images based on support vector machines and texture features.** Int. J. Eng. Trends Technol 2018; 66: 7–11

Zaninovic, N., Rosenwaks, Z. **Artificial intelligence in human in vitro fertilization and embryology.** Fertil. Steril. 2020; 114: 914–920

Zhan, Q., Sierra, ET., Malmsten, J., Ye, Z., Rosenwaks, Z., Zaninovic, N. **Blastocyst score, a blastocyst quality ranking tool, is a predictor of blastocyst ploidy and implantation potential.** F&S Reports 2020; 1: 133–141

Zhan, Q., Tabora, J., Malmsten, J., Rosenwaks, Z., Zaninovic, N. **The blastocyst score, a quantitative measurement of blastocyst quality, is a predictor of blastocyst ploidy and implantation potential.** F&S Reports 2020; 1: 52–60

Zhao, M., Xu, M., Li, H., Alqawasmeh, O., Chung, J.P.W., Li, T.C., Lee, T.L., Tang, P.M., Chan, D.Y.L. **Application of convolutional neural network on early human embryo segmentation during in vitro fertilization.** J. Cell Mol. Med. Mar 2021; 25: 2633–2644. doi:10.1111/jcmm.16288

Zollner, U., Zollner, K.P., Steck, T., Dietl, J. **Pronuclear scoring. Time for international standardization.** J. Reprod. Med. 2003; 48

Received 22 June 2021; received in revised form 7 October 2021; accepted 4 November 2021.