Journal of Assisted Reproduction and Genetics (2023) 40:301–308
https://doi.org/10.1007/s10815-022-02707-6

ASSISTED REPRODUCTION TECHNOLOGIES

The use of voting ensembles to improve the accuracy of deep neural networks as a non-invasive method to predict embryo ploidy status

Victoria S. Jiang[1] · Hemanth Kandula[2] · Prudhvi Thirumalaraju[2] · Manoj Kumar Kanakasabapathy[2] · Panagiotis Cherouveim[1] · Irene Souter[1] · Irene Dimitriadis[1] · Charles L. Bormann[1] · Hadi Shafiee[2]

Received: 16 June 2022 / Accepted: 23 December 2022 / Published online: 14 January 2023

Abstract

Purpose To determine if creating voting ensembles combining convolutional neural networks (CNN), support vector machine (SVM), and multi-layer neural networks (NN) alongside clinical parameters improves the accuracy of artificial intelligence (AI) as a non-invasive method for predicting aneuploidy.

Methods A cohort of 699 day 5 PGT-A tested blastocysts was used to train, validate, and test a CNN to classify embryos as euploid/aneuploid. All embryos were analyzed using a modified FAST-SeqS next-generation sequencing method. Patient characteristics such as maternal age, AMH level, paternal sperm quality, and total number of normally fertilized (2PN) embryos were processed using SVM and NN. To improve model performance, we created voting ensembles using CNN, SVM, and NN to combine our imaging data with clinical parameter variations. Statistical significance was evaluated with a one-sample *t*-test with 2 degrees of freedom.

Results When assessing blastocyst images alone, the CNN test accuracy was 61.2% (±1.32% SEM, $n = 3$ models) in correctly classifying euploid/aneuploid embryos ($n = 140$ embryos). When the best CNN model was assessed as a voting ensemble, the test accuracy improved to 65.0% (AMH; $p = 0.1$), 66.4% (maternal age; $p = 0.06$), 65.7% (maternal age, AMH; $p = 0.08$), 66.4% (maternal age, AMH, number of 2PNs; $p = 0.06$), and 71.4% (maternal age, AMH, number of 2PNs, sperm quality; $p = 0.02$) ($n = 140$ embryos).

Conclusions By combining CNNs with patient characteristics, voting ensembles can be created to improve the accuracy of classifying embryos as euploid/aneuploid from CNN alone, allowing for AI to serve as a potential non-invasive method to aid in karyotype screening and selection of embryos.

Keywords Artificial intelligence · Non-invasive genetic testing · Embryo ploidy prediction · Machine learning · Preimplantation genetic testing (PGT) · Assisted reproductive technology (ART)

Introduction

Embryo morphology was initially the only method for assessing embryo quality and viability. Identifying and selecting a high-quality embryo for transfer is a key component of improving pregnancy rates; however, morphology alone cannot guarantee a euploid embryo or a successful implantation event. Paired morphological and cytogenetic assessment of cleavage and blastocyst embryos revealed that chromosomal abnormalities had little effect on morphology up to day 3 of development, with many chromosomally abnormal embryos still advancing to the highest morphological grade [1]. As time-lapse imaging platforms became increasingly preferred within clinics, multiple studies [2] have assessed the impact of morphologic time points and

✉ Charles L. Bormann
cbormann@partners.org

✉ Hadi Shafiee
hshafiee@bwh.harvard.edu

[1] Division of Reproductive Endocrinology and Infertility, Obstetrics and Gynecology, Massachusetts General Hospital, Harvard Medical School, 55 Fruit Street, Suite 10A, VincentBoston, MA 02114, USA

[2] Division of Engineering in Medicine, Brigham and Women's Hospital, Harvard Medical School, 65 Landsdowne Street, Cambridge, MA 02139, USA

time to development in correlation with aneuploidy rates, with morphokinetic parameter assessment failing to improve the likelihood of selecting euploid embryos [3]. Additionally, subjectivity in embryo morphologic grading continues to be a major challenge in training and standardization of morphologic-based decision making among embryologists.

The advent of next-generation sequencing and the advancement of preimplantation genetic testing for aneuploidy (PGT-A) has led to a genetic renaissance that will continue to grow and impact the future of the field. While transferring PGT-A tested euploid embryos has been shown in randomized controlled trials to improve the ongoing pregnancy rate of women between 35 and 40 compared to morphology alone [4], this narrow niche of indication and application is not the reality of the widespread use emphasizing couples' desire to improve outcomes and avoid miscarriage. PGT-A testing, however, is a significant financial burden, with an average add-on cost of $5000 to the already surmounting cost of IVF in the add-on market. Also, importantly, there is lost time for the patient while awaiting results, increases in the embryology workflow, and continual concerns behind the long-term effects of trophectoderm sampling and assisted hatching procedures on the embryo. Given the time needed to perform PGT-A testing, utilization necessitates either fresh transfer of an untested embryo or delayed frozen embryo transfer, leading to more medications, ultrasounds, and subsequent financial strain upon the patients. Recently, PGT-A testing has also come under scrutiny as embryo mosaicism continues to muddy results, challenging physicians and patients alike in the storage, management, and use of these embryos. When considering the implications of possible errors in sequencing, insufficient sampling with possible need for re-biopsy, and the concerns of mosaicism, PGT-A testing is far from perfect.

In reaction to these imperfect invasive biopsy techniques, the field is seeing a rapid expansion of non-invasive genetic testing methods such as spent culture media (SCM) testing, blastocoel fluid sampling (BFS), and the use of artificial intelligence (AI) to be able to bridge the gap in assisting patients and physicians in clinical management. SCM testing involves collecting the spent media from extended embryo culture and using cell-free DNA testing techniques to assess the ploidy status of the embryos. While the non-invasive technique preserves embryo integrity, the technique currently lacks diagnostic uniformity, with a reported concordance with trophectoderm biopsy or whole embryo sequencing ranging from 30.4–90% [5], with standard trophectoderm biopsy consistently outperforming SCM in direct comparison [6]. BFS serves as an alternative, non-cellular source of genetic material which involves needle aspiration of the blastocoel fluid. While less invasive than trophectoderm biopsy, this invasive technique posed significant technical challenges in both sample collection and analysis, with inferior predictive results compared to SCM [7]. The accuracy and efficiency of both blastocoel fluid sampling and SCM testing are mainly limited by technical challenges associated with low quality and quantity of DNA [8], making these techniques far from being considered safe or effective for commercial use. These results are widely inconsistent between techniques, and with this range of accuracy, these methods are not a reliable form of testing when considering choosing embryos for transfer, vitrification, or discard.

AI is the use of machine learning and computational statistics to perform tasks that require complex interpretation or processing, previously only attainable through human intelligence. AI has been a growing interest in the field of REI, with particular promise in both embryology quality assurance/quality control such as combating subjectivity in embryology morphology grading [9–11] and improving clinical pregnancy outcomes such as predicting implantation [10, 12]. Previous image-based AI models [13, 14] have shown promising results in euploidy prediction when analyzing blastocyst images alongside patient metadata such as age and lab characteristics. In this study, we describe the use of an AI system that combines blastocyst images with patient characteristics into a voting ensemble composed of convolutional neural networks (CNN), support vector machines (SVM), and multi-layer neural networks (NN) to improve the accuracy of predicting embryo ploidy status (Fig. 1).

Materials and methods

Data collection and handling

Data was collected at the Massachusetts General Hospital (MGH) Fertility Center in Boston, Massachusetts, following approval through the Institutional Review Board (IRB#2019P001000 and 2022P002955). Sperm quality was defined as 1 "Excellent" if the total motile count (TMC) from the raw specimen was > 15 million and the sperm concentration, motility, and strict morphology were within the normal range as defined by the WHO 5th edition [15]. Sperm quality was classified as 2 "Good" if the TMC was > 15 million and if sperm concentration, motility, or strict morphology were outside the normal range. Sperm quality was classified as 3 "Fair" if the TMC was between 5 and 15 million. Sperm quality was classified as 4 "Poor" if the TMC was below 5 million. Time-lapse imaging videos of embryos were recorded using a commercial time-lapse imaging system (EmbryoScope, Vitrolife). The imaging system used a Leica 20 × objective that collected images at 10 min intervals under illumination from a single 635 nm LED. Each embryo was exported as a video (.avi) using the imaging system software. Videos were processed into each respective image frame for all timepoints post

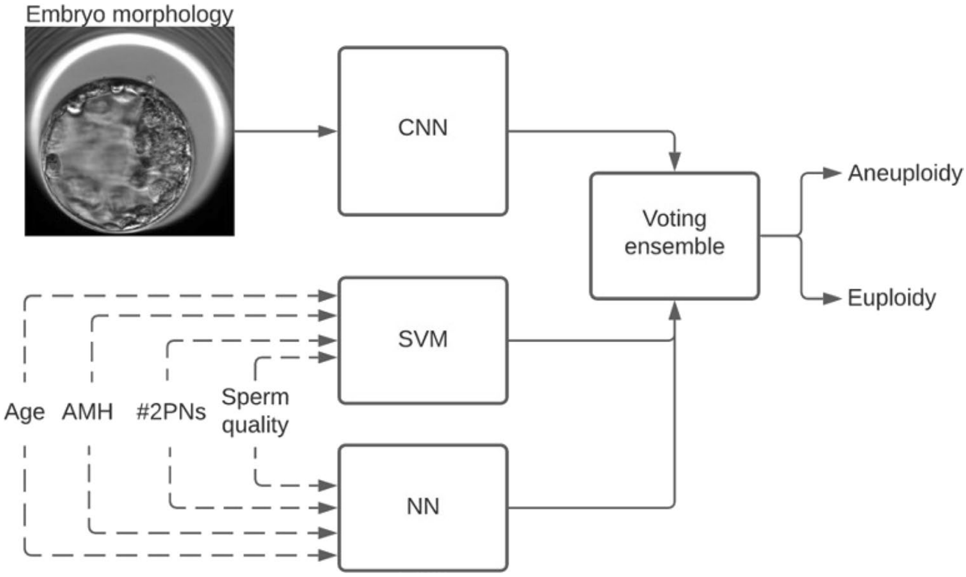

Fig. 1 Utilization of non-morphological parameters in ploidy prediction. The scheme shows the dataflow of the different parameters through the voting ensemble. The dotted lines show optional parameters that were used in the comparisons

insemination. Each extracted image was 250 × 250 pixels and subsequently cropped to 210 × 210 pixels to remove any potential identifiers present within the frame. Out-of-focus images were included in the datasets and used for both testing and training. Only completely non-discernable images of embryos were removed from the study. Since image collection timepoints across all patients were not consistent we binned them into groups of around 18 min intervals. In total, imaging data from 699 embryos from 248 patients were utilized in this study.

Voting ensemble neural network development

Voting ensembles are a type of machine learning model that uses information gathered from multiple machine learning models to improve the overall performance for a given task, ideally achieving better performance than any single model used in the ensemble. Three models were used as part of a soft voting ensemble namely, a convolutional neural network (CNN), a support vector machine (SVM), and a multi-layer neural network (NN).

For the CNN, we used our previously developed models for blastocyst classification [12, 16]. Embryos were evaluated on day 5 blastocyst stage, prior to trophectoderm biopsy and vitrification. The CNN was used alone and in conjunction with a successive add-in of patient parameters analyzed in SVM and multilayer NN models to generate a voting ensemble to assess embryo ploidy status. SVMs construct a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks like outlier's detection. SVMs can also be adapted to efficiently perform a non-linear classification, implicitly mapping their inputs into high-dimensional feature spaces. A nonlinear SVM with the radial bias function kernel was trained to predict embryo ploidy status. A grid search is performed to find the best combination of parameters, C and gamma for radial basis function (RBF) kernels and C for linear kernels, in respect of the area under the curve (AUC). In RBF kernel, gamma was varied between 0.001 and 30 and C between 0.01 and 100. The class weight was set to "balanced."

NN consists of multiple layers of nodes that include an input layer, multiple intermediate hidden layers, and an output layer. Except for the input nodes, each node is a neuron that uses a sigmoid activation. This network was trained with a stochastic gradient descent optimizer for predicting embryo ploidy status. SVM and NN model processes following patient characteristics such as maternal age, AMH level, paternal sperm quality (1–4: 1 = Excellent, 4 = Poor), and total number of normally fertilized (2PN) embryos. A total of 699 embryos were used in training and optimizing the network models. The CNN, SVM, and multi-layer NN models were trained with the Keras framework, CNN, and multi-layer NN uses with TensorFlow backend and were trained on NVidia 1080Ti GPUs.

Study design and statistical analysis

A total of 6828 embryos were manually assessed and graded by trained embryologists according to the Gardner grading system, as previously described [17]. "Euploid" status was included "euploid" classification on PGT-A testing (Invitae). "Non-euploid" status included "aneuploid" and "indeterminate" classifications on PGT-A testing (Invitae). PGT-A results ("euploid" and "non-euploid" as mentioned above) were further registered for each associated embryo grade (embryo stage (3–6), inner-cell mass (ICM, A-C), and trophectoderm grade (A–C)). Manual grading accuracy in correctly classifying embryos as "euploid" or "non-euploid" was calculated using neural network and logistic regression models both before and after accounting for clinical parameters (maternal age, anti-Mullerian hormone (AMH), number of 2 pronuclei embryos (2PNs), sperm quality).

The CNN was initially tested alone using day 5 blastocyst time-lapse images exclusively, without associated patient information. We tested the system's ability to identify embryo ploidy status correctly using a unique, independent test set of embryos (140 embryos) which were not previously used and did not overlap with the training data set. The CNN model was tested over three repetitions of running models. When the system identified the embryo ploidy status correctly, a "pass" was noted. If the system identified the embryo ploidy status incorrectly, a "fail" was noted. If the system cannot arrive at any consistent decision, it was considered as an error and removed from the analysis. Patient characteristics were subsequently collected and combined into a SVM and multilayer NN.

The absolute error between the AI-predicted ploidy status and the documented PGT-A testing results was used by the software to calculate the accuracy rate. Two-tailed *t*-tests and chi-squared tests were used to compare patient demographics between groups, and to compare manual morphology grades to euploid status. Differences in accuracy among each AI model were analyzed and statistical significance was evaluated with a one-sample *t*-test with 2 degrees of freedom. Statistical significance was met when the *p*-value was less than 0.05.

Results

Of the 699 day 5 blastocyst embryos within the CNN testing set, 360 embryos were aneuploid and 339 embryos were euploid on PGT-A testing. Demographic characteristics in this study are described in Table 1. Within the entire

Table 1 Demographics of the embryo cohorts

Demographics	All Embryos	Aneuploid	Euploid	*p*-value
Age (SD)	37.30 (3.6)	38.28 (3.5)	36.26 (3.4)	3.01E-14
BMI (kg/m2) (SD)	24.53 (4.4)	24.55 (4.4)	24.52 (4.5)	0.90
AMH (SD)	3.00 (2.4)	2.87 (2.4)	3.14 (2.5)	0.12
Day 3 FSH (SD)	7.29 (2.1)	7.33 (2.2)	7.26 (2.0)	0.66
Total oocytes retrieved (SD)	12.76 (4.7)	12.33 (4.7)	13.22 (4.7)	0.02
# of 2PNs (SD)	8.30 (3.6)	7.74 (3.4)	8.89 (3.7)	7.16E-05
# of HQB (SD)	5.09 (2.7)	4.68 (2.6)	5.53 (2.7)	3.27E-05
Race/ethnicity				
White, *n* (%)	527 (75.4)	280 (77.8)	247 (72.9)	0.18
Black, *n* (%)	13 (1.9)	5 (1.4)	8 (2.4)	0.68
Asian, *n* (%)	105 (15.0)	48 (13.3)	57 (16.8)	0.19
Hispanic/Latino, *n* (%)	6 (0.9)	4 (1.1)	2 (0.6)	0.76
Other, *n* (%)	19 (2.7)	9 (2.5)	10 (2.9)	0.07
Declined, *n* (%)	13 (1.9)	3 (0.8)	10 (2.9)	0.27
Unavailable, *n* (%)	16 (2.3)	11 (3.1)	5 (1.5)	1.00
SART diagnosis				
Male factor, *n* (%)	243 (34.8)	113 (31.4)	130 (38.4)	0.09
Endometriosis, *n* (%)	27 (3.9)	14 (3.9)	13 (3.8)	1.00
DOR, *n* (%)	97 (13.9)	68 (18.9)	29 (8.6)	2.08E-4
Tubal factor, *n* (%)	85 (12.2)	39 (10.8)	46 (13.6)	0.32
Uterine factor, *n* (%)	79 (11.3)	40 (11.1)	39 (11.5)	0.94
Unexplained, *n* (%)	97 (13.9)	48 (13.3)	49 (14.5)	0.82
PCOS, *n* (%)	56 (8.0)	21 (5.8)	35 (10.3)	0.03
RPL, *n* (%)	44 (6.3)	21(5.8)	23 (6.8)	0.66
Fertility preservation, *n* (%)	96 (13.7)	50 (13.9)	46 (13.6)	1.00
Ovulation disorder, *n* (%)	165 (23.6)	91 (25.3)	74 (21.8)	0.43
Other, *n* (%)	62 (8.9)	41 (11.4)	21 (6.2)	0.03
Sperm class				
1, Excellent, *n* (%)	326 (46.6)	172 (47.8)	154 (45.4)	0.65
2, Good, *n* (%)	273 (39.1)	145 (40.3)	128 (37.8)	0.47
3, Fair, *n* (%)	85 (12.2)	36 (10.0)	49 (14.5)	0.10
4, Poor, *n* (%)	15 (2.2)	7 (1.9)	8 (2.4)	0.87

Table 2 Manual embryo morphology with associated Gardner grade and PGT-A testing results

Embryo grade	# Euploid embryos	# Non-euploid embryos	Total # embryos analyzed	% euploid
3AB	524	667	1191	44.0%
3BB	114	184	298	38.3%
3BC	74	148	222	33.3%
3CB	4	11	15	26.7%
4AB	332	324	656	50.6%
4BB	130	120	250	52.0%
4BC	63	109	172	36.6%
4CB	3	7	10	30.0%
5AB	136	88	224	60.7%
5BA	822	848	1670	49.2%
5BB	556	484	1040	53.5%
5BC	266	364	630	42.2%
5CB	16	18	34	47.1%
6AB	117	91	208	56.3%
6BB	77	40	117	65.8%
6BC	32	43	75	42.7%
6CB	8	8	16	50.0%
Total	3274	3554	6828	47.9%

cohort, embryos were from predominantly White (75.4%) and Asian (15.0%) patients with an average maternal age of 37.3 ± 3.6 years, an average BMI of 24.5 ± 4.4 kg/m^2, an average serum AMH of 3.00 ± 2.4 ng/mL, and average day 3 FSH of 7.3 ± 2.1 IU/mL. The most common infertility diagnoses among the entire cohort included male factor (34.8%), unexplained infertility (13.9%), and diminished ovarian reserve (DOR, 13.9%). Sperm parameters were predominantly of "Excellent" (Class 1, 46.6%) and "Good" (Class 2, 39.1%) quality. When comparing the aneuploid embryos to the euploid embryos, aneuploid embryos came from older patients (38.3 ± 3.5 vs 36.3 ± 3.4 years, p=3.01E-14, aneuploid vs euploid, respectively) with fewer oocytes (12.3 ± 4.7 vs 13.2 ± 4.7 oocytes, p=0.02, aneuploid vs euploid, respectively), fewer number of 2PNs (#2PN, 7.7 ± 3.4 vs 8.9 ± 3.7 years, p=7.16E-5, aneuploid vs euploid, respectively), and fewer number of high-quality blastocysts (HQB, 4.7 ± 2.6 vs 5.5 ± 2.7, p=3.01E-14, aneuploid vs euploid, respectively). Infertility diagnoses of DOR (18.9% vs 8.6%, p=2.08E-4, aneuploid vs euploid, respectively) and other (11.4% vs 6.2%, $p=0.03$) were higher among aneuploid embryos compared to euploid embryos. No other statistically significant differences were noted among the aneuploid vs euploid embryo cohorts.

Manual embryo morphology with associated Gardner grade was analyzed among 6828 blastocysts with associated PGT-A testing results (Table 2). There was a total of 3274 blastocysts classified as "euploid," representing 47.9% of embryos analyzed. Based on these data, euploid classification was associated with multiple morphologic markers by the Gardner grading, with advanced blastulation stage (stage 6 > 3) and higher inner cell mass and trophectoderm grades (grade A > C) associated with euploid classification ($p < 0.001$). Furthermore, accuracy based solely on manual embryo grade morphologies was 47.9% when using a neural network model and 60.0% when applying a logistic regression model in correctly classifying embryos as euploid or aneuploid without additional clinical information. After factoring in patient characteristics (maternal age, AMH, number of 2PNs, sperm quality), accuracy increased to 60.0% and 62.1% when using a neural network and logistic regression models, respectively.

Accuracy rates of each AI model with associated statistical comparisons are listed in Table 3. When used to assess embryoscope blastocyst images alone, the CNN had a test accuracy of 61.2% (SEM: 1.32, 3 replicates) in correctly classifying embryos as euploid or aneuploid without additional patient information. When the best CNN model was assessed as a voting ensemble, the test accuracy improved after incorporating different clinical parameters to 65.0% (AMH; $p=0.1$), 66.4% (maternal age; $p=0.06$), 65.7% (maternal age, AMH; $p=0.08$), 66.4% (maternal age, AMH, number of 2PNs; $p=0.06$), and 71.42% (maternal age, AMH, number of 2PNs, sperm quality; $p=0.02$) ($n=140$ embryos), as listed in Table 3.

Table 3 Accuracy of the artificial intelligence models used to predict the ploidy status of a day 5 blastocysts with addition of patient-specific information

AI model	Accuracy (%)	Δ Accuracy (%)	95% CI	*P*-value
CNN only	61.19	-	-	-
CNN, AMH	65.00	3.81	– 1.89, 9.51	0.1028
CNN, maternal age	66.42	5.23	– 0.46, 10.94	0.0585
CNN, AMH, maternal age	65.71	4.52	– 1.18, 10.23	0.0762
CNN, AMH, maternal age, #2PNs	66.42	5.23	– 0.46, 10.94	0.0585
CNN, AMH, maternal age, #2PNs, sperm quality	71.42	10.23	4.53, 15.94	0.0164

Discussion

Through this study, we show the novel use of voting ensembles to improve the accuracy of image-based CNN processing alone. Combining patient characteristics with the image-based algorithm not only significantly improved our system by over 10%, but also further enforced these findings are greater than pure chance. Creating non-invasive genetic analytic methods for embryo karyotype screening is incredibly beneficial to improve patient outcomes, particularly in the context of the growing emphasis on prioritizing single embryo transfer.

The use of voting ensembles to incorporate additional testing variables into CNN-only AI is an important first step to improving predictive accuracy. Manual morphology alone for ploidy prediction was aligned with the accuracy of our CNN-only AI model, showing the limited utility of morphology images alone in euploid classification. By incorporating additional variables within the model, our combined voting ensemble system achieved a predictive accuracy of 71.4%, a significant improvement of accuracy from CNN-only AI, manual morphology alone, and combined with clinical parameters (62.1%) (Fig. 2). This novel use of voting ensembles to simultaneously incorporate continuous, categorical, and image-based data serves as an important stepping stone to improve AI predictive power, particularly to assist with embryo selection in the absence of PGT-A testing results.

AI serves as a promising new addition to the field of REI, especially given the field uniquely combines embryo imaging with patient characteristics. This AI system we have created is able to assess blastocyst quality and can serve as an important tool to be able to aid in embryo selection, especially when there are more than 2 HBQ in the absence of PGT-A testing. Our AI system would be easily integrated into any clinic, as many clinics already have advanced imaging apparatuses and high-grade light microscopy to capture embryo images. This AI system can be utilized on the day of transfer, returning immediate results that can impact clinical care and fresh transfer. Although training and installation of the software will be an initial hurdle to integration, this system serves as a low-cost adjunct to care to aid in embryo selection for fresh transfers or in the absence of PGT-A results.

The major limitation of the study includes the training set. The embryos utilized to train the CNN were composed only of embryos that were imaged with the embryoscope. While this allowed for uniformity of image quality for training the CNN in this novel utilization of voting ensembles, this may

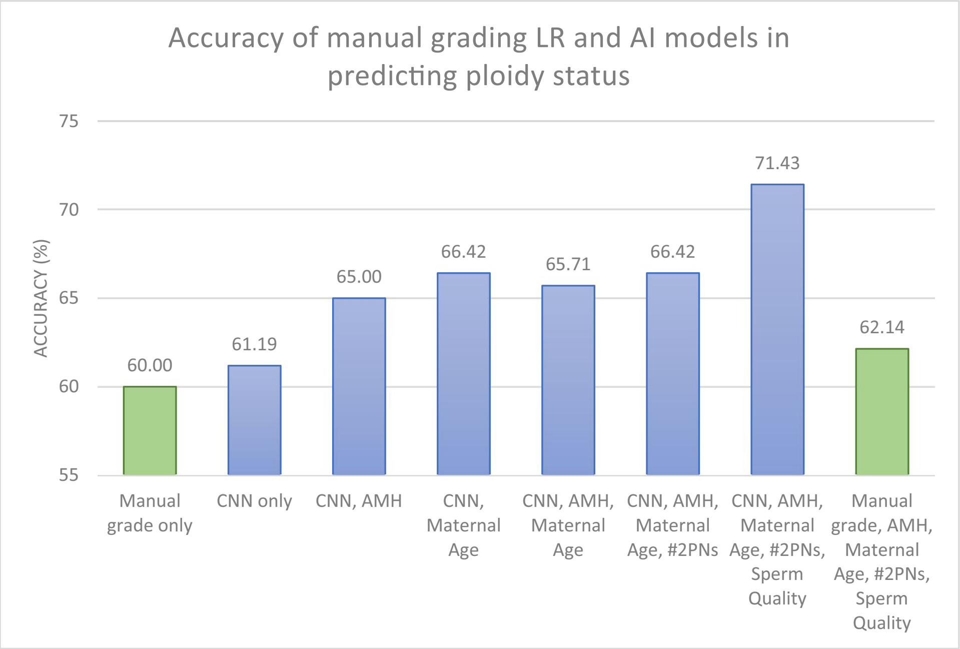

Fig. 2 Accuracy of manual grading logistic regression (LR) and AI models in predicting the ploidy status of day 5 blastocysts. Green columns (first and last column, respectively) represent logistic regression accuracy results when trying to predict the ploidy status of the embryo based on manual embryo grading before and after taking into account clinical characteristics (maternal age, AMH, number of 2PNs, and sperm quality); blue columns represent AI models accuracy results when trying to predict ploidy status of the embryo based on the CNN image before and after stepwise taking into account clinical characteristics (maternal age, AMH, number of 2PNs embryos, and sperm quality); AI artificial intelligence, CNN convolutional neural network, AMH anti-Mullerian hormone, 2PN 2 pronuclei

potentially limit the clinics that can integrate these machine learning systems, given not all laboratories have the same imaging platforms. Though there have been advancements made in applying AI algorithms to images captured on different imaging devices [18], the current algorithm has not been validated across different systems. All biopsied embryos were classified as being either euploid or aneuploid. This algorithm did not factor in high- or low-level mosaics, which can result in positive clinical outcomes [19]. Additionally, the limited sample size may lead to unintended bias within the machine learning algorithms and may limit the generalizability of the data. For example, Black and Hispanic patients are underrepresented in our population, composing 1.9% and 0.9% of our embryo population, respectively, while nationally, they represent 13.4% and 18.5% of the US population, respectively, according to the 2021 US Census. To address these limitations, future studies should include a more diverse, representative embryo cohort with a larger sample for training, validation, and testing to improve generalizability and reduce unintended bias. Also, we only incorporated 4 patient characteristics into our current system, and exploring additional patient characteristics may improve the accuracy of the model.

Conclusions

By combining image-based CNNs with patient characteristics, the use of voting ensembles is a novel method to improve the accuracy of classifying embryos as euploid/aneuploid over image-based CNN alone. This AI system can be easily integrated into clinics, and serves to improve accessibility, affordability, and feasibility of in-clinic, immediate genetic screening for those who may not have initially qualified or opted for PGT-A. As accuracy, imaging, and data acquisition continue to improve, AI may serve as a powerful non-invasive method that aids in the karyotype screening of embryos. Future studies should include analysis for mosaicism and larger embryo cohorts.

Author contribution All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Charles Bormann, Victoria Jiang, Manoj Kumar Kanakasabapathy, Prudhvi Thirumalaraju, and Panagiotis Cherouveim. The first draft of the manuscript was written by Victoria Jiang and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This work was partially supported by the Brigham Precision Medicine Developmental Award and Innovation Evergreen Fund (Brigham Precision Medicine Program, Brigham and Women's Hospital), Partners Innovation Discovery Grant (Partners Healthcare), and R01EB033866, R01AI138800, R33AI140489, and R61AI140489 (National Institute of Health).

Data Availability The datasets, R Studio statistical code, and machine learning algorithms used and/or analyzed during the current study are available from the corresponding author on reasonable request and under a data transfer agreement with Mass General Brigham.

Declarations

Ethics approval and consent to participate Informed consent was obtained from each individual before participation. Study protocols were approved by the Institutional Review Board (IRB#2019P001000 and 2022P002955) at Massachusetts General Hospital and Brigham and Women's Hospital.

Consent for publication Not applicable.

Competing interests Authors Dr. Hadi Shafiee, Dr. Charles Bormann, Prudhvi Thirumalaraju, and Manoj Kumar Kanakasabapathy wish to disclose a patent, currently licensed by a commercial entity, on the use of AI for embryology (US11321831B2). The rest of the authors declare that they have no competing interests.

References

1. Fragouli E, Alfarawati S, Spath K, Wells D. Morphological and cytogenetic assessment of cleavage and blastocyst stage embryos. Mol Hum Reprod. 2014;20:117–26.
2. Swain JE. Could time-lapse embryo imaging reduce the need for biopsy and PGS? J Assist Reprod Genet. 2013;30:1081–90.
3. Zhang J, Tao W, Liu H, Yu G, Li M, Ma S, et al. Morphokinetic parameters from a time-lapse monitoring system cannot accurately predict the ploidy of embryos. J Assist Reprod Genet. 2017;34:1173–8.
4. Munné S, Kaplan B, Frattarelli JL, Child T, Nakhuda G, Shamma FN, et al. Preimplantation genetic testing for aneuploidy versus morphology as selection criteria for single frozen-thawed embryo transfer in good-prognosis patients: a multicenter randomized clinical trial. Fertil Steril. 2019;112:1071-1079.e7.
5. Belandres D, Shamonki M, Arrach N. Current status of spent embryo media research for preimplantation genetic testing. J Assist Reprod Genet. 2019;36:819–26.
6. Yin B, Zhang H, Xie J, Wei Y, Zhang C, Meng L. Validation of preimplantation genetic tests for aneuploidy (PGT-A) with DNA from spent culture media (SCM): concordance assessment and implication. Reprod Biol Endocrinol. 2021;19:41.
7. Shi W, Zhao Z, Xue X, Li Q, Yao Y, Wang D, et al. Ploidy testing of blastocoel fluid for screening may be technically challenging and more invasive than that of spent cell culture media. Front Physiol. 2022;13:794210.
8. Leaver M, Wells D. Non-invasive preimplantation genetic testing (niPGT): the next revolution in reproductive genetics? Hum Reprod Update. 2020;26:16–42.
9. Bormann CL, Thirumalaraju P, Kanakasabapathy MK, Kandula H, Souter I, Dimitriadis I, et al. Consistency and objectivity of automated embryo assessments using deep neural networks. Fertil Steril. 2020;113:781-787.e1.
10. Fitz VW, Kanakasabapathy MK, Thirumalaraju P, Kandula H, Ramirez LB, Boehnlein L, et al. 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;38:2663–70.
11. Kanakasabapathy MK, Thirumalaraju P, Bormann CL, Kandula H, Dimitriadis I, Souter I, et al. Development and evaluation of inexpensive automated deep learning-based imaging systems for embryology. Lab Chip. 2019;19:4139–45.

12. Thirumalaraju P, Kanakasabapathy MK, Bormann CL, Gupta R, Pooniwala R, Kandula H, et al. Evaluation of deep convolutional neural networks in classifying human embryo images based on their morphological quality. Heliyon. 2021;7:e06298.
13. Chavez-Badiola A, Flores-Saiffe-Farías A, Mendizabal-Ruiz G, Drakeley AJ, Cohen J. Embryo ranking intelligent classification algorithm (ERICA): artificial intelligence clinical assistant predicting embryo ploidy and implantation. Reprod Biomed Online. 2020;41:585–93.
14. Huang B, Tan W, Li Z, Jin L. An artificial intelligence model (euploid prediction algorithm) can predict embryo ploidy status based on time-lapse data. Reprod Biol Endocrinol. 2021;19:185.
15. World Health Organization. WHO laboratory manual for the examination and processing of human semen. 5th ed. Geneva: World Health Organization; 2010. p. 271.
16. Bormann CL, Kanakasabapathy MK, Thirumalaraju P, Gupta R, Pooniwala R, Kandula H, et al. Performance of a deep learning based neural network in the selection of human blastocysts for implantation. eLife. 2020;9:55301.
17. Alpha Scientists in Reproductive Medicine and ESHRE Special Interest Group of Embryology. The Istanbul consensus workshop on embryo assessment: proceedings of an expert meeting. Hum Reprod. 2011;26:1270–83.
18. Diakiw SM, Hall JMM, VerMilyea MD, Amin J, Aizpurua J, Giardini L, et al. Development of an artificial intelligence model for predicting the likelihood of human embryo euploidy based on blastocyst images from multiple imaging systems during IVF. Hum Reprod. 2022;37:1746–59.
19. Viotti M, Victor AR, Barnes FL, Zouves CG, Besser AG, Grifo JA, et al. Using outcome data from one thousand mosaic embryo transfers to formulate an embryo ranking system for clinical use. Fertil Steril. 2021;115:1212–24.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.