Received: 17 April 2019 | Revised: 21 May 2019 | Accepted: 2 June 2019

DOI: 10.1002/rmb2.12284

ORIGINAL ARTICLE

Reproductive Medicine and Biology

WILEY

Feasibility of predicting live birth by combining conventional embryo evaluation with artificial intelligence applied to a blastocyst image in patients classified by age

Yasunari Miyagi[1,2] | Toshihiro Habara[3] | Rei Hirata[3] | Nobuyoshi Hayashi[3]

[1]Medical Data Labo, Okayama City, Japan

[2]Department of Gynecologic Oncology, Saitama Medical University International Medical Center, Hidaka City, Japan

[3]Okayama Couple's Clinic, Okayama City, Japan

Correspondence
Yasunari Miyagi, Medical Data Labo, 289-48 Yamasaki, Naka Ward, Okayama City, Okayama Prefecture, Japan 703-8267.
Email: ymiyagi@mac.com

Abstract

Purpose: To identify the multivariate logistic regression in a combination (combination method) involving artificial intelligence (AI) classifiers in images of blastocysts along with a conventional embryo evaluation (CEE) to predict the probability of accomplishing a live birth in patients classified by maternal age.

Methods: Retrospectively, a total of 5691 blastocysts were enrolled. Images captured 115 hours or 139 hours if not yet sufficiently large after insemination were classified according to age as follows: <35, 35-37, 38-39, 40-41, and ≥42 years old. The classifiers for each category were created by using convolutional neural networks associated with deep learning. Next, the feasibility of a method combining AI with multivariate logistic model functions by CEE was investigated.

Results: The values of the area under the curve (AUC) and the accuracies to predict live birth achieved by the CEE/AI/combination methods were 0.651/0.634/0.655, 0.697/0.688/0.723, 0.771/0.728/0.791, 0.788/0.743/0.806 and 0.820/0.837/0.888, and 0.631/0.647/0.616, 0.687/0.675/0.671, 0.725/0.697/0.732, 0.714/0.776/0.801, and 0.910/0.866/0.784 for age categories of <35, 35-37, 38-39, 40-41, and ≥42 years old, respectively.

Conclusions: Though there were mostly no significant differences regarding the AUC and the sensitivity plus specificity in all age categories, the combination method seemed to be the best.

KEYWORDS

artificial intelligence, blastocyst, deep learning, live birth, neural network

1 | INTRODUCTION

The goal of assisted reproductive technology (ART) is securing a live birth. Development failure of embryo, or miscarriage, results in gloomy psychological results, the loss of time, and inflicted costs for the patients.

Morphological structures, such as smooth endoplasmic reticulum clusters (sERCs), vacuoles or refractile bodies, have been studied. None of these structures have been found to be prognostic with respect to developmental ability of oocytes.[1] Conventional morphological evaluation has had confined success to identify aneuploid embryos.[2-6] Though time-lapse information has been reported to be predictive of aneuploidy, the available evidence may

be not enough to ensure introduction of time-lapse microscopy[2]. Suboptimal embryos can be euploid, while embryos of good morphological quality may be aneuploid.[2,7,8] Pre-implantation genetic testing for aneuploidy (PGT-A)[9,10] is another approach to examine chromosomal profiles, but it is an invasive method for the embryo and is relating to ethical debates. Some countries forbid the transfer of an embryo after biopsy. The profile of the chromosome of the biopsied sample does not always indicate the profile of the rest of the sample because the embryo might have genetic heterogeneity. Mosaicism in the trophectoderm (TE) was found. A single biopsy for TE may not represent the complete TE.[11] There is a report that classifiers using artificial intelligence applied toward an image of a blastocyst implanted later had a potential to predict the probability of live birth.[12] Thus, there have been no established procedures to predict live birth.

Age is one of the most critical factors in sterility.[13,14] The number and quality of oocyte decrease as age advances. Patients older than 35 years should receive fast evaluation for the reason of sterility.[15] Women older than 40 years should ensure more immediate evaluation[16]. An aged oocyte shows dysfunction of cellular organelles and increase in chromosomal abnormality.[17] Advanced age is a risk factor for female sterility, miscarriage, and stillbirth.[18] The delivery rate categorized by age (<35, 35-37, 38-39, 40-41, 42-45 years old) affects the developmental speed of the embryo significantly ($P < 0.0001$).[19] The live birth rates associated with ART of patients in <35, 35-37, 38-39, 40-41, and ≥42 years old were 0.20, 0.17, 0.12, 0.08, and 0.01, respectively, by the Japan Society of Obstetrics and Gynecology in 2015.[20] Thus, age is one of the most critical factors in fertility, and there are no standard procedures to treat blastocysts or patients by age. Thus, we reported our system with an application of deep learning in a convolutional neural network[21-24] with artificial intelligence (AI), which was applied to blastocyst images classified by age to explore a means of meeting this study by noninvasively predicting live births.[25] Deep learning becomes popular among machine learning methods including logistic regression,[26] naive Bayes,[27] nearest neighbor,[28] neural network,[29] random forest,[30] and deep learning. AI can generate the confidence score that is a probability showing the estimated value of belonging to the live birth category. The score can be recognized as a blastocyst ranking. This approach will make it easier for physicians and embryologists to select better blastocysts.

We here demonstrate the retrospective predictions of live birth by using the multivariate regression function in combination with a conventional embryo evaluation (CEE) method that includes clinical information, observation, and grading of the morphological features of blastocysts and is applied with AI to blastocysts images categorized by age. In this article, we demonstrate the advanced outcome by introducing the multivariate regression function defined as the combination method and by investigating the optimal cutoff points of the receiver operator characteristic (ROC) curves generated by AI, CEE, and the combination method, respectively. We present the feasibility of the combination method to predict the probability of accomplishing a live birth.

2 | MATERIALS AND METHODS

2.1 | Patients and data preparation

This study was approved by the Institutional Review Board (IRB) at Okayama Couples' Clinic (IRB no. 18000128-05) and was performed with explanations to the patients and a Web site with additional information with an opt-out option. A total of 5691 blastocysts obtained from patients from January 2009 to April 2017 with fully deidentified data were enrolled. All of blastocysts were tracked. Whether the outcome was a live birth or a non-live birth was confirmed. All data were divided into training datasets and test datasets randomly at a ratio of 4 to 1 (Figure 1).

2.2 | CEE

Every blastocyst with the morphological features and clinical information, such as maternal age, body mass index, time of in vitro fertilization, time of embryo transfer, FSH value, anti-Müllerian hormone value, blastocyst grade on day 3, embryo cryopreservation day, blastomere number on day 3 after insemination, grade of TE, grade of inner cell mass, antral follicle count, average diameter of the blastocyst, existence of oviduct infertility, existence of endometriosis, existence of immune sterility, insemination procedures, ovarian stimulation methods, refractile body, sERC grade, degree of blastocyst expansion, existence of a vacuole, male body mass index, and male age, were pursued to assess the outcome of live birth or non-live birth. The information above is defined as CEE in this study.[25] This information was provided by embryologists and doctors who were engaged in clinical practice for at least twenty years and were thought to be specialists who carried out the standardized laboratory practice related to embryo morphology assessment according to the international consensus meeting in 2011.[31]

The relationships between live birth and each factor in the CEE were examined, and then, we obtain univariate regression functions. The significant factors without multicollinearity, indicating a state of strong correlations among the independent variables, were chosen in the multivariate analysis. Then, a multivariate regression function for the CEE to predict live birth was obtained.

2.3 | Blastocyst images

An image of the blastocyst was captured on about at 115 hours or 139 hours if the blastocyst was not yet large enough after insemination. The image was saved and deidentified, containing no data, which could be used to identify the person. The images were sent to the AI system offline. The images were classified by maternal age into five categories of patients who were <35, 35-37, 38-39, and 40-41, or ≥42 years old. The numbers of non-live births and live births were 1389 and 876, respectively, in the <35 group; 863 and

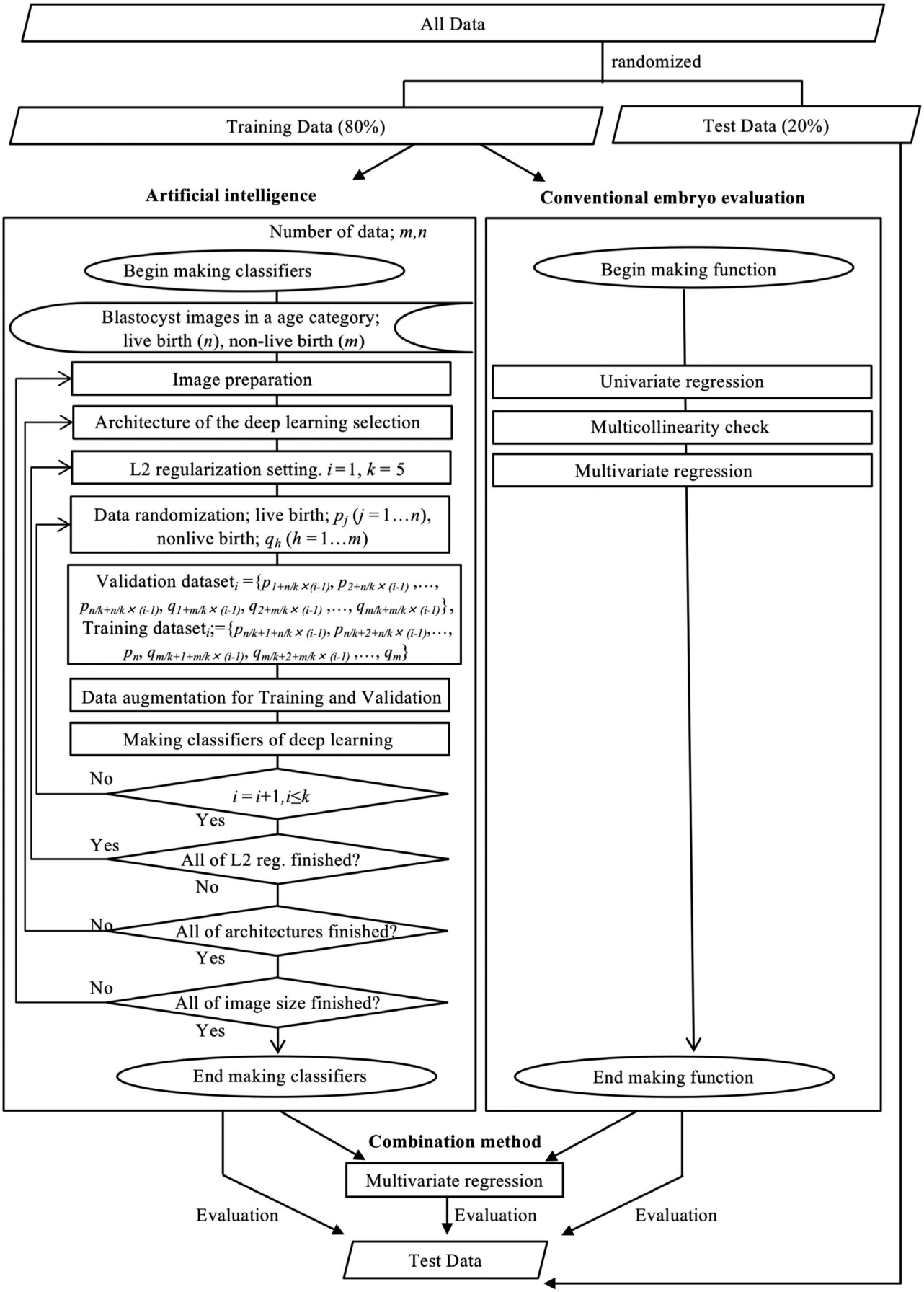

FIGURE 1 A flowchart for making classifiers

381, respectively, in the 35-37 group; 545 and 164, respectively, in the 38-39 group; 674 and 130, respectively, in the 40-41 group; and 633 and 36, respectively, in the ≥42 group. The probabilities of live birth for the age groups <35, 35-37, 38-39, 40-41, and ≥42 years old were 0.387, 0.306, 0.231, 0.162, and 0.054, respectively. The images of blastocysts which resulted in live births and those of blastocysts which resulted in miscarriages and resulted in non-live births were used to create the AI classifiers.[25]

2.4 | Preparation for AI

All of images that were deidentified were transferred to the AI system offline. Each image was cropped to a square and then saved in 100 × 100 pixel size. Eighty percent of the training dataset was used as the AI training dataset. The rest of the dataset was defined as the validation dataset. Thus, the AI training dataset, validation dataset, and test dataset were not overlapped. The AI classifier was trained by an AI training dataset with simultaneous validation and then tested with the test dataset. The training datasets were augmented, because the blastocyst image processing of the arbitrary any degrees of rotation can yield to images resulting in the different vector data of the same category.[25]

2.5 | AI classifier

AI classifier programs in each age category were developed. The classifiers consisted of convolutional neural networks[32] that attempted to imitate the visual cortex of the brain of the mammals[21-24,32-37] and L2 regularization[38,39] to acquire the probability for predicting either live birth or non-live birth.[25] We conducted deep learning with a convolutional neural network of 11 layers consisting of convolution layers with various kernel sizes[40-42] and output channels, pooling layers,[43-46] flattened layers,[47] linear layers,[48,49] rectified linear unit layers,[50,51] and one softmax layer.[52,53] The softmax layer generated confidence score that was the probability of a live birth (Table 1). We used cross-validation.[54-56] The suitable number of images for the training data was studied by evaluating variances and accuracy using the fivefold cross-validation method (Figure 1): First, the test data consisted of the initial twenty percent of the data collected in each category, and a classifier was trained. Next, the test data were changed to the next twenty percent of the data. We repeated this procedure five times to encompass all data. The number of augmented training data was investigated until the accuracy and the variance of accuracies were likely to show the maximum and minimum values, respectively. This procedure reveals the suitable number of training data to validate the prediction performance more accurately by combining five times. Then, the best classifier showing the smallest variance and the best accuracy was chosen by varying the architecture of the neural network and by varying hyperparameters and an image size (40 × 40, 50 × 50, 75 × 75, and 100 × 100 pixels). In case of the accuracies were not clearly different, the best classifier was selected based on the values of the summation of the specificity and the sensitivity. The AI classifiers were obtained for each age category.

TABLE 1 Architectures of the best classifier that showed the best accuracy for each age category

		Age (years)					
Layers		<35	35-37	38-39	40-41	≥42	All ages
1. Convolution layer	Output channels	50	40	50	64	50	50
	Kernel size	5 × 5	5 × 5	5 × 5	5 × 5	5 × 5	5 × 5
2. ReLU							
3. Pooling layer	Kernel size	2 × 2	2 × 2	2 × 2	2 × 2	2 × 2	2 × 2
4. Convolution layer	Output channels	64	64	64	64	64	64
	Kernel size	5 × 5	5 × 5	5 × 5	5 × 5	5 × 5	5 × 5
5. ReLU							
6. Pooling layer	Kernel size	2 × 2	2 × 2	2 × 2	2 × 2	2 × 2	2 × 2
7. Flatten layer							
8. Linear layer size		2^{10}	2^{10}	2^{10}	2^{10}	2^{10}	2^{10}
9. ReLU							
10. Linear layer size		2	2	2	2	2	2
11. Softmax layer							

Notes: The proper convolutional neural network structures, which consisted of eleven layers in convolutional deep learning, were obtained. The numbers of output channels in the first convolution layer were different.
Abbreviation: ReLU, Rectified linear units.

2.6 | Live birth prediction function of AI and CEE

By using both the value, x_1, calculated by the multivariate regression function for the CEE, and the confidence score, x_2, derived from the AI classifier, the multivariate logistic model functions, $y = 1/(1 + \mathrm{Exp}(\beta_0 + \beta_1 x_1 + \beta_2 x_2))$; β_0, β_1, β_2: coefficients, were developed to predict the probability of live birth. The area under the receiver operator characteristic curve (AUC); the optimal cut-point value corresponding to the point with the lowest distance to the upper-left corner of the ROC curve[57]; and the sensitivity, specificity, and accuracy of AI, CEE, and the combination of AI and CEE were obtained.

2.7 | Development environment

The environment for development used in this study was as follows: Intel Core i5, 3.30 GHz, 32 GB (Santa Clara, California, USA) and NVIDIA GeForce GTX 1080 Ti (Santa Clara, California, USA), Windows 10 (Redmond, Washington, USA) and Mathematica 11.3 (Wolfram Research, Champaign, IL, USA).

2.8 | Statistics

Mathematica 11.3 was used for statistical analyses.

3 | RESULTS

The live birth ratio of all data was 0.279. The live birth ratios for the age groups <35, 35-37, 38-39, 40-41, and ≥42 years old were 0.387 (876/2265), 0.306 (381/1244), 0.231 (164/709), 0.162 (130/804), and 0.054 (36/669), respectively.[25]

Univariate regression functions and the multivariate regression function of the CEE for predicting the probability of live birth are shown in Tables 2 and 3, respectively. Ten independent variables for the multivariate function with no multicollinearity were detected. The variables demonstrated in Table 3 were acquired using the formulae in Table 2. The results demonstrated that the age seemed to be the most important because the *P*-value of the age was the minimum, as shown in Table 3. When the ten values were substituted into the multivariate logistic regression function, the calculated value was the predicted probability of live birth by the CEE.

Prior to creating the best AI classifier, the best numbers of the training dataset were determined to be 18 120, 17 910, 8505, 7716, and 12 840 in the groups <35, 35-37, 38-39, 40-41, and ≥42 years old, respectively, in this study.[25] Then, the best values in the L2 regularization were determined to be 0.01, 0.0005, 0.01, 0.0001, and 0.00015 for the groups <35, 35-37, 38-39, 40-41, and ≥42 years old, respectively. Next, the best AI classifiers were obtained. The best image size was 50 × 50 pixels. It took 0.15 seconds/image to classify and generate the confidence score.

Next, the multivariate regression functions in combination with CEE and with AI applied to the blastocyst images of patients categorized by age, which were defined as combination methods, were obtained as shown in Table 4. The AUC values for predicting live birth accomplished by the CEE/AI/combination methods were 0.651 ± 0.027/0.634 ± 0.027/0.655 ± 0.027, 0.697 ± 0.037/0.688 ± 0.

TABLE 2 Univariate regression functions of the CEE parameters for predicting the probability of live birth

Independent variables	Formulae	Coefficients
Age	$k/(1 + \mathrm{Exp}(\beta_0 + \beta_1 x))$	$\beta_0 = -10.742 \pm 4.106$ ($P = 0.0089$)
		$\beta_1 = 0.284 \pm 0.109$ ($P = 0.0088$)
		k = 0.451
Times of embryo transfer	$1/(1 + \mathrm{Exp}(\beta_0 + \beta_1 x))$	$\beta_0 = 0.635 \pm 1.158$ ($P = 0.584$)
		$\beta_1 = 0.156 \pm 0.123$ ($P = 0.204$)
Anti-Müllerian Hormone (ng/ml)	$1/(1 + \mathrm{Exp}(\beta_0 + \beta_1 x))$	$\beta_0 = 1.282 \pm 2.640$ ($P = 0.627$)
		$\beta_1 = 0.062 \pm 0.139$ ($P = 0.678$)
Blastomere number on Day 3	$k/(2\pi\sigma^2)^{1/2}$ $\mathrm{Exp}(-(x-m)^2/(2\sigma^2))$	$\sigma = 4.668 \pm 0.773$ ($P = 4.179 \times 10^{-5}$)
		$m = 11.624 \pm 0.663$ ($P = 1.969 \times 10^{-10}$)
		k = 4.643 ± 0.611 ($P = 3.91 \times 10^{-6}$)
Grade on Day 3	$k/(1 + \mathrm{Exp}(\beta_0 + \beta_1 x))$	$\beta_0 = -7.967 \pm 8.012$ ($P = 0.320$)
(Class A = 1, B = 2, C = 3, D = 4)		$\beta_1 = 2.584 \pm 2.582$ ($P = 0.317$)
		k = 0.319
Embryo cryopreservation day	$\beta_0 + \beta_1 x$	$\beta_0 = 0.435$
(Day 5 = 1, Day 6 = 2)		$\beta_1 = -0.131$
Inner Cell Mass	$\beta_0 + \beta_1 x$	$\beta_0 = 0.479 \pm 0.037$ ($P = 0.049$)
(A = 1, B = 2, C = 3)		$\beta_1 = -0.131 \pm 0.017$ ($P = 0.083$)
Trophectoderm	$\beta_0 + \beta_1 x$	$\beta_0 = 0.526 \pm 0.002$ ($P = 0.0026$)
(A = 1, B = 2, C = 3)		$\beta_1 = -0.124 \pm 0.001$ ($P = 0.005$)
Averaged diameter (μm)	$1/(1 + \mathrm{Exp}(\beta_0 + \beta_1 x))$	$\beta_0 = 2.623 \pm 5.312$ ($P = 0.621$)
		$\beta_1 = -0.011 \pm 0.030$ ($P = 0.723$)
Body mass index (kg/m^2)	$1/(1 + \mathrm{Exp}(\beta_0 + \beta_1 x))$	$\beta_0 = -0.631 \pm 0.844$ ($P = 0.454$)
		$\beta_1 = 0.079 \pm 0.035$ ($P = 0.026$)

Notes: Independent variables, which were related to live birth and were also used in the multivariate regression, are presented. Each formula is determined to fit the data distribution. Coefficients are shown as the mean ± SE.

TABLE 3 Multivariate logistic regression function, $1/(1 + \mathrm{Exp}(\beta_0 + \beta_1 x_1 + \ldots + \beta_{10} x_{10}))$, of the CEE for predicting live birth

Independent variables	Coefficients	*P*-value	Odds ratio
Constant (β_0)	$\beta_0 = 6.642 \pm 0.500$	2.81×10^{-40}	-
Age value (β_1)	$\beta_1 = -4.047 \pm 0.368$	3.68×10^{-28}	57.24
Average diameter value (β_2)	$\beta_2 = -4.628 \pm 0.749$	6.359×10^{-10}	102.34
TE value (β_3)	$\beta_3 = -2.164 \pm 0.462$	2.832×10^{-6}	8.71
Embryo cryopreservation day value (β_4)	$\beta_4 = -3.202 \pm 0.862$	2.230×10^{-4}	24.59
ET times value (β_5)	$\beta_5 = -2.652 \pm 0.817$	1.158×10^{-3}	14.19
ICM value (β_6)	$\beta_6 = -0.837 \pm 0.537$	0.119	2.31
AMH value (β_7)	$\beta_7 = -1.078 \pm 0.767$	0.160	2.94
Blastomere number value (β_8)	$\beta_8 = -0.618 \pm 0.593$	0.298	1.85
Body mass index value (β_9)	$\beta_9 = -0.819 \pm 0.838$	0.328	2.27
Grade on day 3 value (β_{10})	$\beta_{10} = 0.266 \pm 0.93$	0.778	1.31

Notes: The values of independent variables except constant β_0 are calculated values by univariate regression functions, shown in Table 1. Multicollinearity is not observed between any two independent variables. Coefficients are shown as the mean ± SE.
AMH, Anti-Müllerian hormone; Embryo transfer; ICM, Inner cell mass; Trophoectoderm.

TABLE 4 Coefficients of logistic regression, $y = 1/(1 + \mathrm{Exp}(\beta_0 + \beta_1 x_1 + \beta_2 x_2))$, that show the probability of live birth as a function of the CEE score and of the confidence score that is the AI-generated predicted probability for live birth from a blastocyst image

Patient age (years)	β_0 (±SE)	β_1 (±SE)	β_2 (±SE)
Age < 35	4.326 (±1.686)	−4.150 (±1.010)	−5.634 (±4.648)
35 ≤ age < 38	3.628 (±0.672)	−4.266 (±1.685)	−5.508 (±2.408)
38 ≤ age < 40	7.314 (±2.346)	−9.123 (±2.896)	−18.988 (±11.408)
40 ≤ age < 42	5.044 (±0.833)	−10.929 (±4.119)	−4.031 (±2.593)
42 ≤ age	8.087 (±1.995)	−19.341 (±7.444)	−20.985 (±12.226)

Abbreviations: β_0, β_1, β_2, Coefficients; SE, Standard error; x_1, Score of the CEE; x_2, Confidence score of the blastocyst; *y*, Probability of live birth.

036/0.723 ± 0.036, 0.771 ± 0.052/0.728 ± 0.054/0.791 ± 0.050, 0.788 ± 0.063/0.743 ± 0.066/0.806 ± 0.061, and 0.820 ± 0.106/0.837 ± 0.103/0.888 ± 0.089 (mean ± SE) for <35, 35-37, 38-39, 40-41, and ≥42 years old, respectively, as shown in Figure 2 and Table 5. The AUC values of the CEE/AI/combination methods were 0.745 ± 0.069/0.726 ± 0.075/0.773 ± 0.088 (mean ± SD), respectively. There were no significant differences between CEE, AI, and combination methods in each age category. The AUC value significantly increased as a function of age in all methods because the slopes of the linear regression were more than zero in the CEE/AI/combination methods with *P*-values of 0.00425/0.00580/0.00219, respectively, by linear regression analysis. Furthermore, the optimal cutoff points from the ROC curve of the CEE/AI/combination methods were 0.420/0.414/0.388, 0.279/0.314/0.281, 0.190/0.226/0.219, 0.190/0.268/0.142, and 0.184/0.118/0.037 for <35, 35-37, 38-39, 40-41, and ≥42 years old, respectively (Table 5).

The sensitivities at the optimal cut-point of the CEE/AI/combination methods were 0.580/0.530/0.652, 0.714/0.655/0.786, 0.727/0.697/0.758, 0.700/0.650/0.700, and 0.667/0.833/1.000 for <35, 35-37, 38-39, 40-41, and ≥42 years old, respectively, as shown in Figure 3 and Table 5. There were no significant differences in any of the methods with respect to the sensitivities in each age category except age <35 years old. In the age category of <35 years old, the sensitivity of the AI was significantly lower than that of the combination method (*P* = 0.019, by the chi-square test). The sensitivity significantly increased in all methods as a function of age because the *P*-values by the Cochran–Armitage test in the CEE, the AI, and the combination methods were 0.01248, 0.0623, and 0.00453, respectively. The sensitivities of the CEE/AI/combination methods were 0.678 ± 0.059/0.673 ± 0.109/0.779 ± 0.134 (mean ± SD), respectively.

The specificities at the optimal cut-point of the CEE/AI/combination methods were 0.665/0.724/0.592, 0.673/0.685/0.612, 0.725/0.697/0.725, 0.716/0.794/0.816, and 0.922/0.867/0.773 for <35, 35-37, 38-39, 40-41, and ≥42 years old, respectively, as shown in Figure 4 and Table 5. There were no significant differences in any of the methods with respect to the specificities in each age category except age <35, 40-41, and ≥42 years old. In the age of <35 years old, the specificity of AI was significantly higher than that of the combination method (*P* = 0.0014, by the chi-square test). In the age category of 40-41 years old, the specificity of the combination

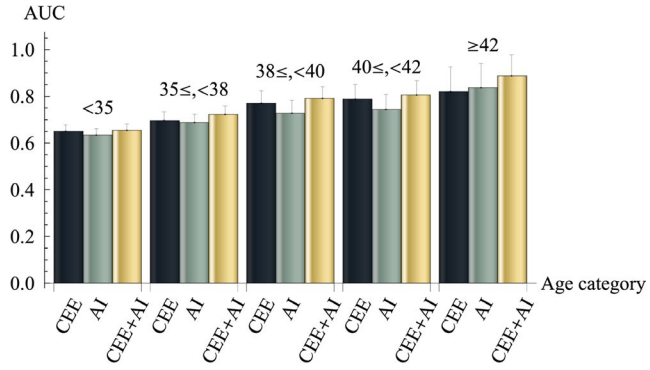

FIGURE 2 The AUC values for predicting live birth achieved by conventional embryo evaluation (CEE), artificial intelligence (AI) applied to blastocyst images from patients categorized by age and a combination method of CEE and AI. No significant differences were observed in each age category. The AUC value significantly increased as a function of age in all methods because the slopes of the linear regression were more than zero in the CEE/AI/combination methods with *P*-values of 0.00425/0.00580/0.00219, respectively, by linear regression analysis. AI, artificial intelligence; CEE, conventional embryo evaluation; CEE + AI, combination method with CEE and AI

method was significantly higher than that of CEE (P = 0.049, by the chi-square test). In the age category of ≥42 years old, the specificity of CEE was significantly higher than that of the combination method (P = 0.001, by the chi-square test). The specificity significantly increased in all methods as a function of age because the *P*-values by the Cochran–Armitage test in the CEE, the AI, and the combination methods were 1.577×10^{-8}, 3.721×10^{-5}, and 2.216×10^{-8}, respectively. The specificities of the CEE/AI/combination methods were 0.740 ± 0.105/0.753 ± 0.076/0.704 ± 0.098 (mean ± SD), respectively.

The sensitivities plus specificities at each optimal cut-point of the CEE/AI/combination methods were 1.245/1.254/1.244, 1.387/1.340/1.398, 1.452/1.394/1.483, 1.416/1.444/1.516, and 1.589/1.700/1.773 for <35, 35-37, 38-39, 40-41, and ≥42 years old, respectively, as shown in Figure 5 and Table 5. The sensitivity plus specificity significantly increased in all methods as a function of age because the slopes of the linear regression were more than zero in the CEE/AI/combination methods with *P*-values of 0.0288/0.0199/0.0091, respectively, by the linear regression analysis. The sensitivities plus specificities of the CEE/AI/combination methods were 1.418 ± 0.124/1.426 ± 0.168/1.483 ± 0.193 (mean ± SD), respectively.

The accuracies at the cut-points for predicting live birth accomplished by the CEE/AI/combination methods were 0.631/0.647/0.616, 0.687/0.675/0.671, 0.725/0.697/0.732, 0.714/0.776/0.801, and 0.910/0.866/0.784 for <35, 35-37, 38-39, 40-41, and ≥42 years old, respectively, as shown in Figure 6 and Table 5. No significant differences in any methods with respect to accuracies in each age category except for age ≥42 years old were observed. In the age category of ≥42 years old, the accuracy of the combination method was significantly lower than that of CEE (P = 0.004 by the chi-square test). The accuracy significantly increased in all methods as a function of age because the *P*-values by the Cochran–Armitage test in the CEE, the AI, and the combination methods were 3.812×10^{-10}, 7.306×10^{-8}, and 1.238×10^{-7}, respectively. The accuracies of the CEE/AI/combination methods were 0.733 ± 0.105/0.733 ± 0.089/0.721 ± 0.077 (mean ± SD), respectively.

4 | DISCUSSION

Here, we developed new multivariate logistic functions in combination with both the CEE and the AI that we had previously published[25] with improvement for predicting live birth in patients categorized by age, as shown in Table 4. The values of independent variables of the function are values calculated by multivariate regression functions of the CEE parameters and values as confidence scores generated by the AI classifiers of deep learning with a convolutional neural network using images of blastocysts that were categorized by age. This multivariate logistic function, defined as the combination method, demonstrated the best results, although there were mostly no significant differences from those of the CEE or the AI.

We believe the values of AUC, sensitivity, and specificity are the most important statistics for evaluating test methods because they are independent of patient distribution. On the other hand, accuracy is dependent on patient distribution. For example, if a test method that always outputs non-live birth for any inputted data would be applied to the images of patients aged ≥42 years, it would demonstrate very high accuracy because most patients aged ≥42 years belong to the non-live birth category.

The AUC values for predicting live birth accomplished by the CEE/AI/combination methods are shown in Figure 2 and Table 5. Although there were no significant differences in each age category, the combination method showed the best predictive results in each category. There seems to be no comparable study for predicting live birth. However, in terms of the AUC from pre-implantation genetic screening, there is a report that a prediction model classifying embryos into low-, medium-, or high-risk categories achieved an AUC of 0.74.[58] The AUC values of the combination method, especially for ages ≥38 years old in this study, were better. The AUC values of the combination method for age ≥40 years in this study were higher than 0.8, indicating good predictive results.

The sensitivities at the optimal cut-point of the CEE/AI/combination methods are shown in Figure 3. There were no significant differences among all methods for sensitivities in each age category except age <35 years, at which the sensitivity of the AI was significantly lower than the combination method. The sensitivities of AI were lower than those of the combination method in other age categories, although these values were not significantly different. The AI classifiers of deep learning with a convolutional neural network with architectures in this study seemed to demonstrate moderate sensitivity. Because the incidences of live birth in each category were lower than those of non-live birth, it might be more difficult to prepare a good training dataset of live births for the AI to create the

TABLE 5 Comparison of CEE, artificial intelligence (AI) and the combination of CEE and AI

Patient age (years)	Actual live birth	Actual non-live birth	Predicted live birth	Predicted non-live birth	Accuracy	Sensitivity	Specificity	PPV[a]	NPV[b]	AUC[c]	95% CI[d] of the AUC	Cut-point
The CEE												
<35	181	272	196	257	0.631	0.580	0.665	0.536	0.704	0.651	0.598-0.703	0.420
35-37	84	165	114	135	0.687	0.714	0.673	0.526	0.822	0.697	0.625-0.768	0.279
38-39	33	109	54	88	0.725	0.727	0.725	0.444	0.898	0.771	0.670-0.872	0.190
40-41	20	141	54	107	0.714	0.700	0.716	0.259	0.944	0.788	0.665-0.911	0.190
≥42	6	128	14	120	0.910	0.667	0.922	0.286	0.983	0.820	0.612-1.03	0.184
The AI with deep learning in the convolutional neural network												
<35	181	272	171	282	0.647	0.530	0.724	0.561	0.699	0.634	0.581-0.687	0.414
35-37	84	165	107	142	0.675	0.655	0.685	0.514	0.796	0.688	0.615-0.760	0.314
38-39	33	109	56	86	0.697	0.697	0.697	0.411	0.884	0.728	0.622-0.834	0.226
40-41	20	141	42	119	0.776	0.650	0.794	0.309	0.941	0.743	0.613-0.873	0.268
≥42	6	128	22	112	0.866	0.833	0.867	0.227	0.991	0.837	0.636-1.039	0.118
Combination of the AI and the CEE												
<35	181	272	229	224	0.616	0.652	0.592	0.515	0.719	0.655	0.600-0.707	0.388
35-37	84	165	130	119	0.671	0.786	0.612	0.508	0.849	0.723	0.653-0.793	0.281
38-39	33	109	55	87	0.732	0.758	0.725	0.455	0.908	0.791	0.693-0.889	0.219
40-41	20	141	40	121	0.801	0.700	0.816	0.350	0.950	0.806	0.687-0.925	0.142
≥42	6	128	35	99	0.784	1.000	0.773	0.171	1.000	0.888	0.713-1.063	0.037

Notes: The value of the CEE was a function of multivariate logistic regression with no multicollinearity of conventional laboratory and clinical factors to predict the probability of live birth. The value of AI was a confidence score that was the AI-generated predicted probability for live birth from an image of the blastocyst in patients categorized by age. The value of the combination was a function of multivariate logistic regression with CEE and AI as described in Table 4. The optimal cut-point of live birth was the value corresponding to the point with the lowest distance to the upper-left corner of the receiver operator characteristic (ROC) curve.[62] The accuracies, sensitivities, and specificities were obtained by using cut-points.

[a]Positive precidtive value.

[b]Negative predictive value.

[c]The area under the curve.

[d]Confidence interval.

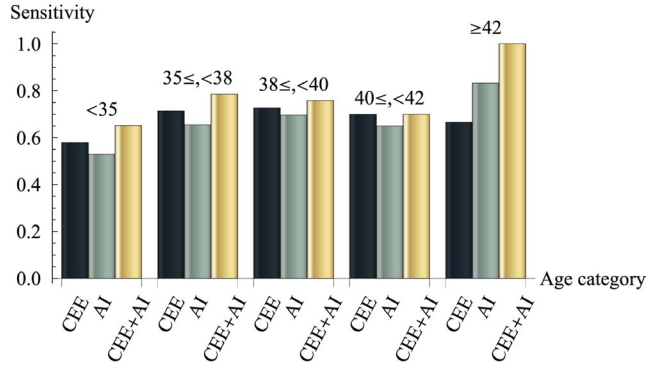

FIGURE 3 The sensitivity for predicting live birth achieved by conventional embryo evaluation (CEE), artificial intelligence (AI) applied to blastocyst images from patients categorized by age and a combination method with CEE and AI. There were no significant differences in all methods for sensitivities in each age category except age <35 years. In the age category of <35 years, the sensitivity of the AI was significantly lower than that of the combination method (P = 0.019, by the chi-square test). The sensitivity significantly increased as a function of age in all methods because the P-values in the CEE, the AI, and the combination method were 0.01248, 0.0623, and 0.00453, respectively, by the Cochran–Armitage test. AI, artificial intelligence; CEE, conventional embryo evaluation; CEE + AI, combination method with CEE and AI. $^{*}P < 0.05$

classifiers. Therefore, more datasets of blastocysts that result in live births might improve the sensitivity.

The specificities at the optimal cut-point of the CEE/AI/combination methods are shown in Figure 4. There were no significant differences among the CEE, AI and combination methods for specificities in each age category except age <35, 40-41, and ≥42 years old. In the age category of <35 years old, the specificity of the AI was significantly higher than the combination method. In the age category of 40-41 years old, the specificity of the combination method was significantly higher than the CEE. In the age category of ≥42 years old, the specificity of the CEE was significantly lower than the combination method. These complicated results suggest that many and various morphological feature types of blastocysts result in non-live births. Therefore, more datasets of blastocysts that result in non-live births might improve the specificity.

The sensitivity plus specificity-1, known as Youden's index,[59] is a statistic value that is useful for the performance of a dichotomous diagnostic test and is often used in ROC analysis. The value of sensitivity plus specificity provides equal weight to false-negative and false-positive values. Therefore, we investigated the sensitivity plus specificity. The sensitivities plus specificities at each optimal cut-point of the CEE/AI/combination methods are shown in Figure 5. The sensitivities plus specificities of the combination method were higher than those of the CEE or the AI in all age categories. The combination methods seemed to be the best from the viewpoint of the performance of a dichotomous diagnostic test.

The accuracies to predict live birth accomplished by the CEE/AI/combination methods are shown in Figure 6. There were no significant differences among all methods for accuracies in each age category except the category of age ≥42 years, in which the accuracy of the CEE was significantly higher than the combination method, probably because the specificity of the CEE, 0.922, was

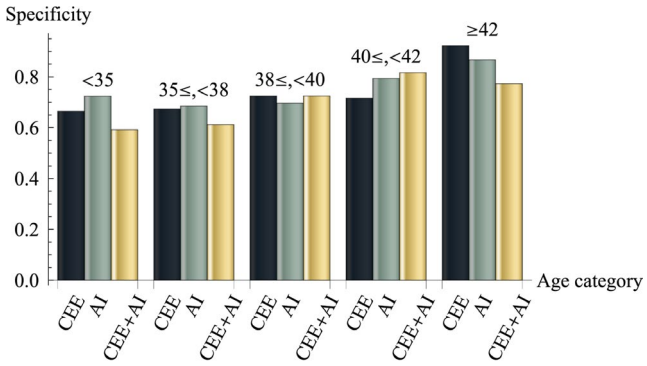

FIGURE 4 The specificity for predicting live birth achieved by conventional embryo evaluation (CEE), artificial intelligence (AI) applied to blastocyst images from patients categorized by age and a combination method with CEE and AI. There were no significant differences in all methods for specificities in each age category except age <35, 40-41 and ≥42 years. In the age category of <35 years, the specificity of the AI was significantly higher than the combination method (P = 0.0014, by the chi-square test). In the age category of 40-41 years old, the specificity of the combination method was significantly higher than the CEE (P = 0.049) by the chi-square test. In the age category of ≥42 years old, the specificity of the CEE was significantly higher than that of the combination method (P = 0.001, by the chi-square test). The specificity significantly increased as a function of age in all methods because the P-values in the CEE, the AI and the combination method were 1.577×10^{-8}, 3.721×10^{-5} and 2.216×10^{-8}, respectively, by the Cochran–Armitage test. AI, artificial intelligence; CEE, conventional embryo evaluation; CEE + AI, combination method with CEE and AI. $^{*}P < 0.05$; $^{**}P < 0.005$

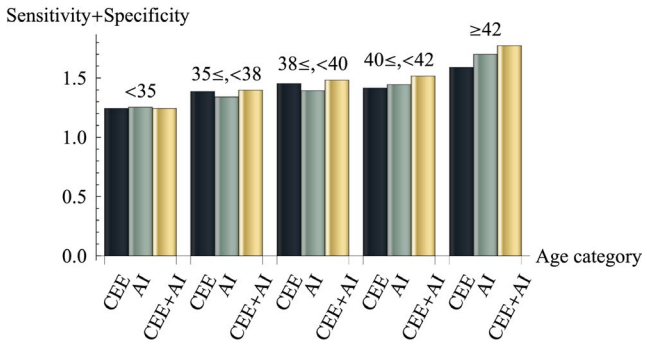

FIGURE 5 The specificity plus specificity for predicting live birth achieved by conventional embryo evaluation (CEE), artificial intelligence (AI) applied to blastocyst images from patients categorized by age and a combination method with CEE and AI. The sensitivity plus specificity significantly increased as a function of age in all methods because the slopes of linear regression were more than zero in the CEE/AI/combination methods with P-values of 0.0288/0.0199/0.0091, respectively, by linear regression analysis. AI, artificial intelligence; CEE, conventional embryo evaluation; CEE + AI, combination method with CEE and AI

superior to that of the other methods and the incidence of non-live birth was 94.6%. If a test with high specificity will be applied to a dataset in which the incidence of a negative result is close to one, the accuracy will be very high. Therefore, we believe the superiority of CEE in terms of the accuracy for age ≥42 years should be acceptable as data bias. The live birth rate for each transfer was reported to be 0.668 based on some clinical factors, such as body mass index.[60] There is another study reported that the TE grade is the single statistically significant independent factor to predict live birth and the live birth probabilities of TE grade A, B, and C are 0.499, 0.339, and 0.080, respectively.[61] The accuracies of the combination method we obtained were superior to the accuracies in these reports. Attention should be paid to the fact that the prediction of the probability of live birth cannot reach 1.00 because there are clinical disincentives to accomplish live birth that the AI classifiers cannot detect, such as uterine factors[60] (eg, uterine myomas,[62] intrauterine adhesions,[63,64] and endometrial polyps[65]); endometriosis[66]; ovarian function[67]; oviduct obstruction[68,69]; immune disorders,[70,71] maternal diseases such as diabetes mellitus,[72] and chronic endometritis[73,74]; and the uterine microbiota[75,76]).

The more advanced the age, the greater was the value of the AUC, sensitivity, specificity, sensitivity plus specificity, and accuracy in the function of the CEE, AI, and combination methods, and this trend was significant. Because the morphological findings in CEE are not modified by age, we believe that the phenomenon could be partially derived from the appropriate selection of the optimal cutoff point of the multivariate regression functions of the CEE parameters and of the confidence score of the AI. In particular, because the incidence of live birth decreases as age advances, lowering the optimal cutoff point could improve the accuracies and the values of the sensitivity plus specificity. In our previous work,[25] in which the cutoff points were set at 0.5 and were derived from the logistic regression model for binary data, the accuracies, sensitivities, and specificities were lower than those in the present study, in which the optimal cutoff points were lower than 0.5 (Table 5). As such, the improvement of the values of AUC, accuracies, sensitivities, and specificities in this study might be achieved not by the progressive skill of the embryologists or physicians but by the improvement of the classifier and selecting the appropriate cutoff point. On the other hand, age is very important for live birth from the viewpoint of biology. The contribution of age was observed in the combination method as well as in the CEE and the AI. Thus, AI seems to recognize some findings related to age from the blastocyst images. There could be some differences in the morphological features of the blastocyst at the same timepoint after insemination in patients of different ages.

It is important to create a good AI classifier to acquire a multivariate logistic function of both CEE and AI. Our AI classifiers have a classification potential almost identical to that of the CEE classifiers. Because the embryologists and physicians who were thought to be specialists provided the CEE in this study, the AI classifiers might be useful in clinical practice. The AI classifiers and the CEE by well-trained embryologists and physicians will improve the combined method. Even for highly skilled embryologists

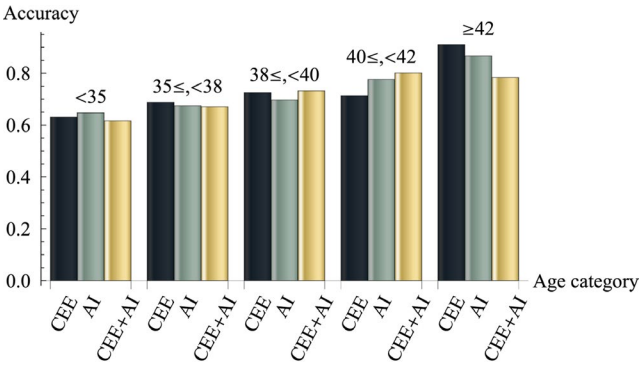

FIGURE 6 The accuracy for predicting live birth achieved by conventional embryo evaluation (CEE), artificial intelligence (AI) applied to blastocyst images from patients categorized by age, and the combination method with CEE and AI. There were no significant differences in all methods for accuracies in each age category except for age ≥42 years. In the age category of ≥42 years old, the accuracy of the combination method was significantly lower than that of the CEE ($P = 0.004$, by the chi-square test). The accuracy significantly increased as a function of age in all methods because the *P*-values in the CEE, the AI, and the combination methods were 3.812×10^{-10}, 7.306×10^{-8}, and 1.238×10^{-7}, respectively, by the Cochran–Armitage test. AI, artificial intelligence; CEE, conventional embryo evaluation; CEE + AI, combination method with CEE and AI. $^{*}P < 0.005$

and physicians, it would be useful to refer to the prediction values calculated by the combination method. When the morphological grades of the blastocysts[77-79] are identical, the several-digit real number of the calculated prediction values would be helpful to determine the order of blastocyst selection. We think it is desirable for embryologists and physicians to refer the combination method that shows the predicted probability of live birth. The blastocysts can be ordered by the prediction values for embryo transfer. The combination method might psychologically confirm the historical experience of embryologists and physicians using AI in practice for the first time.

It is emphasized that blastocysts evaluated by CEE, AI and combination methods are intact and can be transferred to the uterus with no ethical problems. It took 0.15 seconds/image to calculate the prediction probability by the AI. Because CEE information is usually obtained as a routine conventional procedure, the prediction probability of the combined method will be generated in a moment. It is also emphasized that the images of the blastocyst can be transferred, and the probabilities can be reported either by e-mail or by saving the information to cloud spaces via the Internet so that expensive facilities and equipment might not be necessary for medical institutes worldwide.

The AI classifier seems to be almost as good and might be superior to novice practitioners for classifying blastocysts. In clinical practice, AI used by embryologists who are trained insufficiently or are unskilled would be able to facilitate the prediction of live births from blastocysts to a level similar to that of specialists. Time and financial costs of training could be saved. This efficiency will afford

embryologists and physicians time to work on other tasks, such as counseling patients.

We would like to improve the AI, which would result in an improved combination method. Improvements in the architecture of the network and the hyperparameters used for training would be able to create the classifiers better, despite some of the clinical disincentives to accomplish live birth. The following neural networks have made progress: LeNet[80] in 1998, AlexNet[81] in 2012, GoogLeNet[82] in 2014, ResNet[83] in 2015, and Squeeze-and-Excitation networks[84] in 2017. We previously tested ResNet with modification for our dataset, but the result was inferior to the neural network architectures we created in this study (data not shown). The AI for image recognition is still being developed. We believe that we may need more varied patterns of images for datasets. Usually, 500-1000 images may be required for each class with deep learning.[85] Such a large number of datasets for each age category will improve the value of AUC, sensitivity, specificity, and accuracy of the classifier with deep learning. It is also considered to investigate the image size,[85,86] the appropriate number of training datasets, the appropriate timing after insemination to capture images, and the regularization values for further study to improve the accuracies and to avoid overfitting[87-92] that is an error that occurs when a classifier is too fit to a limited set of data. It would be better to study other parameters, such as information of time lapses regarding their potential to predict live birth. The time-lapse information might be included in the future. When the AI progresses, better results will be delivered.

Deep learning with a convolutional neural network was applied to develop classifiers to predict the probability of a live birth from a blastocyst image categorized by age. The combination method with both CEE and AI demonstrated good AUC values in the range of 0.655-0.888. Less than 0.15 second was necessary to complete the analysis of an image. It should be emphasized that this method causes no harm to the embryo. The embryo can be transferred after the prediction of the probability is acquired. It could provide financial savings for clinical institutes as well as patients. It could supply a rapid and useful diagnosis of the classification and allows tests over distances. It is thought that this AI will be useful in reproductive medicine. Although further or advanced study may be required for validation, this system indicates that this combination method would be feasible and may offer profits to both medical workers as well as patients.

The contents in the article have been recognized as patentable in Japan (patent no. 6468576).

CONFLICT OF INTEREST

Yasunari Miyagi, Toshihiro Habara, Rei Hirata, and Nobuyoshi Hayashi declare they have no conflicts of interest.

ETHICAL APPROVAL

In this study, human rights statements and informed consent: All procedures followed were performed in accordance with the ethical standards of the responsible committee on human experimentation and with the Helsinki Declaration of 1964 and its later amendments. Informed consent was obtained from all patients for inclusion. We obtained additional informed consent from all patients for whom identifying information is included. A Web site with additional information including an "opt-out" option was set up for this study.

Animal studies: No animals were used in this study.

The protocol for this research project including the use of human patients was approved by the IRB at Okayama Couples' Clinic (IRB no. 18000128-05).

Clinical trial registry: This study is not a clinical trial.

ORCID

Yasunari Miyagi https://orcid.org/0000-0003-0962-033X

REFERENCES

1. Rienzi L, Vajta G, Ubaldi F. Predictive value of oocyte morphology in human IVF: a systematic review of the literature. *Hum Reprod*. 2011;17:34-45.
2. Kirkegaard K, Ahlström A, Ingerslev HJ, Hardarson T. Choosing the best embryo by time lapse versus standard morphology. *Fertil Steril*. 2015;103:323-332.
3. Finn A, Scott L, O'Leary T, Davies D, Hill J. Sequential embryo scoring as a predictor of aneuploidy in poor-prognosis patients. *Reprod Biomed Online*. 2010;21:381-390.
4. Eaton JL, Hacker MR, Barrett CB, Thornton KL, Penzias AS. Influence of patient age on the association between euploidy and day-3 embryo morphology. *Fertil Steril*. 2010;94:365-367.
5. Eaton JL, Hacker MR, Harris D, Thornton KL, Penzias AS. Assessment of day-3 morphology and euploidy for individual chromosomes in embryos that develop to the blastocyst stage. *Fertil Steril*. 2009;91:2432-2436.
6. Wells D. Embryo aneuploidy and the role of morphological and genetic screening. *Reprod Biomed Online*. 2010;21:274-277.
7. Alfarawati S, Fragouli E, Colls P, et al. The relationship between blastocyst morphology, chromosomal abnormality, and embryo gender. *Fertil Steril*. 2011;95:520-524.
8. Ziebe S, Lundin K, Loft A, et al. FISH analysis for chromosomes 13, 16, 18, 21, 22, X and Y in all blastomeres of IVF pre-embryos from 144 randomly selected donated human oocytes and impact on pre-embryo morphology. *Hum Reprod*. 2003;18:2575-2581.
9. Dahdouh EM, Balayla J, Audibert F, et al. Technical update: preimplantation genetic diagnosis and screening. *J Obstet Gynaecol Can*. 2015;37:451-463.
10. Brezina PR, Kutteh WH. Clinical applications of preimplantation genetic testing. *BMJ*. 2015;19:350.
11. Gleicher N, Metzger J, Croft G, Kushnir VA, Albertini DF, Barad DH. A single trophectoderm biopsy at blastocyst stage is mathematically unable to determine embryo ploidy accurately enough for clinical use. *Reprod Biol Endocrinol*. 2017;15:33.
12. 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.
13. Weiss RV, Clapauch R. Female infertility of endocrine origin. *Arq Bras Endocrinol Metabol*. 2014;58:144-152.
14. Shirasuna K, Iwata H. Effect of aging on the female reproductive function. *Contracept Reprod Med*. 2017;2;23. eCollection.
15. Crawford NM, Steiner AZ. Age-related Infertility. *Obstet Gynecol Clin*. 2015;42:15-25.
16. American College of Obstetricians and Gynecologists Committee on Gynecologic Practice and Practice Committee. Female

age-related fertility decline. Committee Opinion No. 589. *Fertil Steril*. 2014;101:633-634.
17. Igarash H, Takahashi T, Nagase S. Oocyte aging underlies female reproductive aging: biological mechanisms and therapeutic strategies. *Reprod Med Biol*. 2015;14:159-169.
18. Sauer MV. Reproduction at an advanced maternal age and maternal health. *Fertil Steril*. 2015;103:1136-1143.
19. Kato K, Ueno S, Yabuuchi A, et al. Women's age and embryo developmental speed accurately predict clinical pregnancy after single vitrified-warmed blastocyst transfer. *Reprod Biomed Online*. 2014;29:411-416.
20. Saito H, Jwa SC, Kuwahara A, et al. Assisted reproductive technology in Japan: a summary report for 2015 by The Ethics Committee of The Japan Society of Obstetrics and Gynecology. *Reprod Med Biol*. 2018;17(1):20-28. Published online 2017 Nov 29.
21. Fukushima K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol Cybern*. 1980;36:193-202.
22. Hubel D, Wiesel T. Receptive fields and functional architecture of monkey striate cortex. *J Physiol*. 1968;195:215-243.
23. Hubel DH, Wiesel TN. Receptive fields of single neurones in the cat's striate cortex. *J Physiol*. 1959;148:574-591.
24. Schmidhuber J. Deep learning in neural networks: an overview. *Neural Netw*. 2015;61:85-117.
25. Miyagi Y, Habara T, Hirata R, Hayashi N. Feasibility of deep learning for predicting live birth from a blastocyst image in patients classified by age. *Reprod Med Biol*. 2019;18:190-203.
26. Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. *J Biomed Inform*. 2002;35:352-359.
27. Ben-Bassat M, Klove KL, Weil MH. Sensitivity analysis in Bayesian classification models: multiplicative deviations. *IEEE Trans Pattern Anal Mach Intell*. 1980;3:261-266.
28. Friedman J, Baskett F, Shustek L. An algorithm for finding nearest neighbors. *IEEE Trans Comput*. 1975;100:1000–1006.
29. Rumelhart D, Hinton G, Williams R. Learning representations by back-propagating errors. *Nature*. 1986;323:533-536.
30. Breiman L. Random forests. *Mach Learn*. 2001;45:5-32.
31. Alpha Scientists in Reproductive Medicine and ESHRE Special Interest Group of Embryology. The Istanbul consensus workshop on embryo assessment: proceedings of an expert meeting. *Human Reprod*. 2011;26:1270-1283.
32. Bengio Y, Courville A, Vincent P. Representation learning: a review and new perspectives. *IEEE Trans Pattern Anal Mach Intell*. 2013;35:1798-1828.
33. LeCun Y, Bottou L, Orr GB, Müller KR. Efficient BackProp. In: G Montavon, GB Orr, KR Müller, eds.*Neural networks: tricks of the trade.Lecture notes in computer science*. Berlin, Heidelberg: Springer; 2012; vol 7700:9-48.
34. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc IEEE*. 1998;86:2278-2324.
35. LeCun Y, Boser B, Denker JS, et al. Backpropagation applied to handwritten zip code recognition. *Neural Comput*. 1989;1:541-551.
36. Serre T, Wolf L, Bileschi S, Riesenhuber M. Robust object recognition with cortex-like mechanisms. *IEEE Trans Pattern Anal Mach Intell*. 2007;29:411-426.
37. Wiatowski T, Bölcskei H. A mathematical theory of deep convolutional neural networks for feature extraction. *IEEE Trans Inform Theory*. 2018;64:1845-1866.
38. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Lean Res*. 2014;15:1929-1958.
39. Nowlan SJ, Hinton GE. Simplifying neural networks by soft weight-sharing. *Neural Comput*. 1992;4:473-493.
40. Bengio Y. Learning deep architectures for AI. *Found Trends® Mach Learn*. 2009;2:1-127.
41. Mutch J, Lowe DG. Object class recognition and localization using sparse features with limited receptive fields. *Int J Comput Vision*. 2008;80:45-57.
42. Neal RM. Connectionist learning of belief networks. *Artif Intell*. 1992;56:71-113.
43. Ciresan D, Meier U, Masci J, Maria Gambardella L, Schmidhuber J. *Flexible, high performance convolutional neural networks for image classification*. IJCAI Proceedings-international joint conference on artificial intelligence. 2011;22:1237-1242.
44. Scherer D, Müller A, Behnke S. Evaluation of pooling operations in convolutional architectures for object recognition. In: Diamantaras K, Duch W, Iliadis LS, eds. *Artificial neural networks–ICANN 2010. Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer; 2010:92–101.
45. Huang FJ, LeCun Y. *Large-scale learning with SVM and convolutional for generic object categorization. Computer vision and pattern recognition*. 2006 IEEE Computer Society Conference. 2006;1:284–291.
46. Jarrett K, Kavukcuoglu K, Ranzato M, LeCun Y. *What is the best multi-stage architecture for object recognition? Computer vision*. 2009 IEEE International Conference. 2009:2146–2153.
47. Zheng Y, Liu Q, Chen E, Ge Y, Zhao JL. *Time series classification using multi-channels deep convolutional neural networks*. International conference on web-age information management. Springer: Cham; 2014:298–310.
48. Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning. *Nature*. 2015;518:529.
49. Szegedy C, Liu W, Jia Y, et al. *Going deeper with convolutions*. Proceedings of the IEEE conference on computer vision and pattern recognition. 2015:1-9.
50. Glorot X, Bordes A, Bengio Y. *Deep sparse rectifier neural networks*. Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics. 2011:315–323.
51. Nair V, Hinton G. *Rectified linear units improve restricted Boltzmann machines*. Proceedings of International Conference on Machine Learning. 2010:807–814.
52. Krizhevsky A, Sutskever I, Hinton GE. *Imagenet classification with deep convolutional neural networks*. Advances in neural information processing systems. 2012;1097–1105.
53. Bridle JS. Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. In: Soulié FF, Hérault J, eds. *Neurocomputing*, pp. 227-236.Berlin, Heidelberg: Springer; 1990.
54. Kohavi R. *A study of cross-validation and bootstrap for accuracy estimation and model selection*. Proceedings of the 14th International Joint Conference on Artificial Intelligence. 1995; 2: 1137–1143.
55. Schaffer C. Selecting a classification method by cross-validation. *Mach Learn*. 1993;13:135–143.
56. Refaeilzadeh P, Tang L, Liu H. Cross-Validation. In Liu L, Özsu MT, eds. *Encyclopedia of Database Systems*, pp. 532-538. New York, US: Springer US; 2009.
57. Unal I. Defining an optimal cut-point value in ROC analysis: an alternative approach. *Comput Math Methods Med*. 2017;2017:3762651.
58. Campbell A, Fishel S, Bowman N, Duffy S, Sedler M, Thornton S. Retrospective analysis of outcomes after IVF using an aneuploidy risk model derived from time-lapse imaging without PGS. *Reprod Biomed Online*. 2013;27:140–146.
59. Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950;3:32–35.
60. Sanders B. Uterine factors and infertility. *J Reprod Med*. 2006;51:169–176.
61. Ahlström A, Westin C, Reismer E, Wikland M, Hardarson T. Trophectoderm morphology: an important parameter for predicting live birth after single blastocyst transfer. *Hum Reprod*. 2011;26:3289–3296.

62. Ikhena DE, Bulun SE. Literature review on the role of uterine fibroids in endometrial function. *Reprod Sci*. 2018;25:635–643.
63. Pabuccu R, Atay V, Orhon E, Urman B, Ergün A. Hysteroscopic treatment of intrauterine adhesions is safe and effective in the restoration of normal menstruation and fertility. *Fertil Steril*. 1997;68:1141–1143.
64. Liu L, Huang X, Xia E, Zhang X, Li TC, Liu Y. A cohort study comparing 4 mg and 10 mg daily doses of postoperative oestradiol therapy to prevent adhesion reformation after hysteroscopic adhesiolysis. *Human Fertil*. 2018;5:1–7.
65. Taylor E, Gomel V. The uterus and fertility. *Fertil Steril*. 2008;89:1–16.
66. Tomassett C, D'Hooghe T. Endometriosis and infertility: insights into the causal link and management strategies. *Best Pract Res Clin Obstet Gynaecol*. 2018;51:25–33.
67. Christ JP, Gunning MN, Palla G, et al. Estrogen deprivation and cardiovascular disease risk in primary ovarian insufficiency. *Fertil Steril*. 2018;109:594–600.
68. Arronet GH, Eduljee SY, O'Brien JR. A nine-year surgery of fallopian tube dysfunction in human infertility: diagnosis and therapy. *Fertil Steril*. 1969;20:903–918.
69. Segars JH, Hill GA, Herbert CM, Colston A, Moore DE, Winfield AC. Selective fallopian tube cannulation: initial experience in an infertile population. *Fertil Steril*. 1990;53:357–359.
70. Practice Committee of the American Society for Reproductive Medicine. The role of immunotherapy in in vitro fertilization: a guideline. *Fertil Steril*. 2018;110:387-400.
71. Hong YH, Kim SJ, Moon KY, et al. Impact of presence of antiphospholipid antibodies on in vitro fertilization outcome. *Obstet Gynecol Sci*. 2018;61:359–366.
72. Ota K, Ohta H, Yamagishi S. *Diabetes and Female Sterility/Infertility: Diabetes and Aging-related Complications*. Singapore: Springer; 2018.
73. Cicinelli E, De Ziegler D, Nicoletti R, et al. Chronic endometritis: correlation among hysteroscopic, histologic, and bacteriologic findings in a prospective trial with 2190 consecutive office hysteroscopies. *Fertil Steril*. 2008;89:677–684.
74. Johnston-MacAnanny EB, Hartnett J, Engmann LL, Nulsen JC, Sanders MM, Benadiva CA. Chronic endometritis is a frequent finding in women with recurrent implantation failure after in vitro fertilization. *Fertil Steril*. 2010;93:437–441.
75. Moreno I, Simon C. Relevance of assessing the uterine microbiota in infertility. *Fertil Steril*. 2018;110:337–343.
76. Kroon SJ, Ravel J, Huston WM. Cervicovaginal microbiota, women's health, and reproductive outcomes. *Fertil Steril*. 2018;110:327–336.
77. Gardner DK, Lane M, Stevens J, Schlenker T, Schoolcraft WB. Blastocyst score affects implantation and pregnancy outcome: towards a single blastocyst transfer. *Fertil Steril*. 2000;73:1155–1158.
78. Gardner DK, Schoolcraft WB, Wagley L, Schlenker T, Stevens J, Hesla J. A prospective randomized trial of blastocyst culture and transfer in in-vitro fertilization. *Human Reprod*. 1998;13:3434–3440.
79. Gardner DK, Surrey E, Minjarez D, Leitz A, Stevens J, Schoolcraft WB. Single blastocyst transfer: a prospective randomized trial. *Fertil Steril*. 2004;81:551–555.
80. LeCun Y, Haffner P, Bottou L, Bengio Y. *Object recognition with gradient-based learning. In Shape, contour and grouping in computer vision*. Berlin, Heidelberg: Springer; 1999.
81. Krizhevsky A, Sutskever I, Hinton GE. *Imagenet classification with deep convolutional neural networks*. In Proc. of neural information processing systems. 2012:1097-1105.
82. Szegedy C, Liu W, Jia Y, et al. *Going deeper with convolutions*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015:1-9.
83. He K, Zhang X, Ren S, Sun J. *Deep residual learning for image recognition*. In Proceedings of the IEEE conference on computer vision and pattern recognition. 2016:770-778.
84. Hu J, Shen L, Sun G. *Squeeze-and-excitation networks*. arXiv:1709.01507, 2017.
85. Sato M, Horie K, Hara A, et al. Application of deep learning to the classification of images from colposcopy. *Oncol Lett*. 2018;15:3518–3523.
86. Kudva V, Prasad K, Guruvare S. Automation of detection of cervical cancer using convolutional neural networks. Critical Reviews™ in Biomed. *Engineering*. 2018;46(2):135–145.
87. Yu L, Chen H, Dou Q, Qin J, Heng PA. Automated melanoma recognition in dermoscopy images via very deep residual networks. *IEEE Trans Med Imaging*. 2017;36:994–1004.
88. Caruana R, Lawrence S, Giles CL. *Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping*. Advances in Neural Information Processing Systems. 2001:402-408.
89. Baum EB, Haussler D. What size net gives valid generalization? *Neural Comput*. 1989;1:151–160.
90. Geman S, Bienenstock E. Neural networks and the bias/variance dilemma. *Neural Comput*. 1992;1992(4):1–58.
91. Krogh A, Hertz JA. A simple weight decay can improve generalization. *Adv Neural Inform Proc Syst*. 1992;4:950–957.
92. Moody JE. The effective number of parameters: An analysis of generalization and regularization in nonlinear learning systems. *Adv Neural Infor Proc Syst*. 1992;4:847–854.

How to cite this article: Miyagi Y, Habara T, Hirata R, Hayashi N. Feasibility of predicting live birth by combining conventional embryo evaluation with artificial intelligence applied to a blastocyst image in patients classified by age. *Reprod Med Biol*. 2019;18:344–356. https://doi.org/10.1002/rmb2.12284