Received: 8 March 2022 | Accepted: 8 September 2022

DOI: 10.1002/rmb2.12486

Reproductive Medicine and Biology

WILEY

ORIGINAL ARTICLE

Shallow artificial networks with morphokinetic time-lapse parameters coupled to ART data allow to predict live birth

Mehdi Benchaib[1,2,3] | Elsa Labrune[1,3,4] | Sandrine Giscard d'Estaing[1,5,4] | Bruno Salle[1,5,4] | Jacqueline Lornage[1,5,4]

[1]Hospices Civil de Lyon, HFME, Médecine de la Reproduction & Préservation de la Fertilité Féminine, Bron cedex, France

[2]UMR CNRS 5558, LBBE, Villeurbanne Cedex, France

[3]Université Lyon I, Faculté de Médecine Lyon Est, Lyon, France

[4]Inserm U1208, Bron cedex, France

[5]Université Lyon I, Faculté de Médecine Lyon Sud, Oullins cedex, France

Correspondence
Mehdi Benchaib, Hôpital Femme Mère Enfant, Service de Médecine de la Reproduction, 59 Boulevard Pinel, 69675 Bron cedex, France.
Email: mehdi.benchaib@chu-lyon.fr

Funding information
Hospices Civils de Lyon

Abstract

Purpose: The purpose of this work was to construct shallow neural networks (SNN) using time-lapse technology (TLT) from morphokinetic parameters coupled to assisted reproductive technology (ART) parameters in order to assist the choice of embryo(s) to be transferred with the highest probability of achieving a live birth (LB).

Methods: A retrospective observational single-center study was performed, 654 cycles were included. Three SNN: multilayers perceptron (MLP), simple recurrent neuronal network (simple RNN) and long short term memory RNN (LSTM-RNN) were trained with K-fold cross-validation to avoid sampling bias. The predictive power of SNNs was measured using performance scores as AUC (area under curve), accuracy, precision, Recall and F1 score.

Results: In the training data group, MLP and simple RNN provide the best performance scores; however, all AUCs were above 0.8. In the validating data group, all networks were equivalent with no performance scores difference and all AUC values were above 0.8.

Conclusion: Coupling morphokinetic parameters with ART parameters allows to SNNs to predict the probability of LB, and all SNNs seems to be efficient according to the performance scores. An automatic time recognition system coupled to one of these SNNs could allow a complete automation to choose the blastocyst(s) to be transferred.

KEYWORDS
artificial intelligence, blastocyst, embryo selection, time lapse

1 | INTRODUCTION

Since the first child issued from in vitro fertilization (IVF) was born in 1978[1]; physicians have never stopped trying to improve people's chances of getting a child. All aspects of the artificial reproductive technology (ART) process have benefited from this effort: ovarian hyper-stimulation protocols,[2] embryo culture conditions,[3,4] embryo transfer procedures,[5,6] and embryo freezing techniques[7,8] have been improved. The choice of embryo to be transferred is based on either morphological criteria at D2 (day 2 post fertilization),[9] at D3,[10] or at the blastocyst stage,[11] or on preimplantation genetic diagnosis (PGD) techniques.[12,13]

In order to improve the embryos quality, incubators can be equipped with a time-lapse technology (TLT) that enables the culture conditions improvement and the embryo growth longitudinal follow-up. Thus, new information is available to embryologists regarding the choice of the embryo(s) to be transferred, that is morphokinetic data and embryos images at different stages of their growth. Importantly, in the context where the transfer of a single embryo can be preferred to reduce the risks inherent to multiple pregnancy; the choice of the embryo to be transferred is all the more essential.[14,15]

Several algorithms for embryo selection based on morphokinetic data have been proposed.[16] The first algorithms were derived from decision trees based solely on these morphokinetic data.[17–22] More recent algorithms have included deep learning technologies for data analysis, and preferentially convolutional neural network (CNN) for images classification. However, deep learning requires a significant amount of data and resources to be efficient in its development. Hence, the embryos classification could be performed by: machine learning algorithms based on morphokinetic data (Blank et al., 2019; 2019), deep learning analyzing time-lapse images,[23,24] or standard optical light microscope images from mounted camera[25,26]. Some of the methods mentioned above involve very sophisticated algorithms (genetic algorithm, Google's inception technology), which allow accurate results in embryo classification.[23–26] Shallow neural network (SNN) is more often used for data analysis for classification or regression, but in some cases it can also be used for image analysis.[27,28] However, in the majority of the neural networks used to predict live birth from time-lapse data (images or morphokinetic parameters), no bio-clinical data was used.

The chronology of the ART process is somewhat stereotyped: gonadotropin is administered to a woman, oocytes are collected, fertilized, and some embryos are obtained. Some of the embryos reach the blastocyst stage following specific kinetics, and finally the transfer of one or more blastocysts can be performed. The recurrent neural networks (RNN) are simple and able to consider this chronology.

The aim of the present study was to build three SNN: a multilayers perceptron (MLP) not able to consider the events chronology, and two RNNs (simple RNN and long short term memory RNN, (LSTM-RNN)) able to consider the events chronology. The capacity these SNN to predict the best embryos to be transferred according to their likelihood of achieving a live birth; and if the events chronology allows to RNNs to provide better prediction compared to MLP neural network were studied.

2 | MATERIALS AND METHODS

2.1 | Study population

This retrospective observational study took place between January 2013 and December 2018 at the *Hospices Civils de Lyon*, France (*Hôpital Femme Mère Enfant*). This study was approved by the Institutional Review Board of the *Hospices Civils de Lyon*. Cycles from couples with embryos cultured in time-lapse systems until the blastocyst stage were included in the study; no PGD on embryos was performed. Cycles resulting from oocyte donation were not included. Participants age was not an exclusion criterion.

2.2 | Assisted reproduction process

Gonadotrophin releasing hormone (GnRH) agonist protocol was used for the controlled ovarian stimulation.[29,30] The follicle stimulating hormone starting dose was adjusted according to the female age, ovarian reserve, and previous ovarian stimulation outcomes when available. Ovulation was triggered using recombinant human chorionic gonadotropin (Ovitrelle (MerckSerono, Darmstadt, Germany) when at least 3 follicles reached 18 mm. Only intra-cytoplasmic sperm injection (ICSI) cycles were included, allowing to control the time of insemination and to report oocyte- and fertilization-related measures. Sequential media were used for embryo culture, and half-medium change was performed in the afternoon of D2. All embryos were cultured in the Embryoscope or Embryoscope Plus (Vitrolife, Copenhagen, Denmark) TLT incubators. For the present study, only the outcome regarding the transfer of fresh embryos was used. One or two blastocysts per cycle were transferred depending on medical history. In case of live birth, ART procedures for which the number of babies did not match the number of transferred blastocysts were excluded. In case of failure, cycles were included in the study.

2.3 | Statistic

2.3.1 | Shallow neural networks

The data analysis was performed using R software (4.05).[31] Each SNN was optimized for the hyperparameters. Keras[32] with "Tensorflow"[33] as backend was used to construct the SNN: MLP, simple RNN, and LSTM-RNN.[34] In ART process, decisions are made in sequence, each decision being based on previous information, so there is a chronological sequence. The age of the patient will determine the type of treatment, hence the number of oocytes punctured, hence the number of embryos obtained, and the rate of cell division, which corresponds to embryonic development. The hypothesis as to the use of neural network including chronological sequence, as simple RNN or LSTM-RNN, would allow obtaining a neural network providing a greater predictive power. Data were normalized: the mean of the feature values was subtracted to the input value and the result was divided by the standard deviation of the feature values. The number of hidden layers and neurons were chosen empirically, and improved in order to increase the accuracy of the SNN.[35] The different SNN architectures were described in Table 1. For simple RNNs, up to 5 hidden layers can be added in order to highlight more or less complex relationships, moreover, given the small number of layers and the few inputs, the simple RNN is not very exposed to the risk of

TABLE 1 Simplified R scripts (Keras package) of neural networks architecture

	MLP	Simple RNN	LSTM-RNN
Inputs ordered ($n = 11$)	Female age, Cumulative FSH dose, Oocytes retrieved, Fertilization rate, Embryos obtained, Blastocysts obtained, Blastulation rate, Transferred blastocysts, tM, tSB, tB.		
Layers	① ***layer_dense*** (units = 6, activation = "relu") layer_dropout(0.2) ② ***layer_dense*** (units = 6, activation = "relu") layer_dropout(0.2)	① ***layer_simple_rnn*** (units = 26, return_sequences = TRUE, dropout = 0.2, recurrent_dropout = 0.2) ② ***layer_simple_rnn*** (units = 26, return_sequences = TRUE, dropout = 0.2, recurrent_dropout = 0.2) ③ ***layer_simple_rnn*** (units = 26, return_sequences = FALSE, dropout = 0.2, recurrent_dropout = 0.2)	① ***layer_lstm*** (units = 26, recurrent_activation = "sigmoid", return_sequences = TRUE, dropout = 0.2, recurrent_dropout = 0.2) ② ***layer_lstm*** (units = 26, recurrent_activation = "sigmoid", return_sequences = TRUE, dropout = 0.2, recurrent_dropout = 0.2) ③ ***layer_lstm*** (units = 26, recurrent_activation = "sigmoid", return_sequences = FALSE, dropout = 0.2, recurrent_dropout = 0.2)
Output	***layer_dense*** (units = 1, activation = "sigmoid")		
Compilation	optimizer = "adam", learning rate = 0.01 loss = "binary_crossentropy", metrics = c("accuracy")		

Abbreviations: ①, layer number; activation, activation function; and sigmoid function for input and forget gates (recurrent_activation). FSH, follicle stimulating hormone; Layer_lstm, the activation function for output gate is a hyperbolic tangent; LSTM-RNN, long short term memory recurrent neural network; MLP, multi layers perceptron; simple RNN, simple recurrent neural network; tB, blastocyst formation time; tM, morula formation time; tSB, sub-blastocyst formation time; units, number of neurons in the layer.

vanishing gradient. In order to enhance the memory process and to decrease the risk of vanishing gradient, the recurrent LSTM network was tested. The argument of "layer_lstm" had been let as Keras defaults parameters[34]; however, for all cells (name of neuron in recurrent neural network) in the hidden layers, the activation function for output gate is an hyperbolic tangent, and sigmoid function for input and forget gates. The SNN improvements were performed by back-propagation in order to reduce as much as possible the loss value (Table 1). Belonging to a category (birth failure or birth success) was determined from the embryo birth probability calculated according to the used SNN. The cut-off value of each probability provided by the neural network allowed transforming a numerical variable to binary variable, was determined by an automatic thresholding from the ROC curve (pROC package, (Robin et al., 2011). The SNN architectures were saved to h5 format in order to be easily used in an R or Python script. To measure the SNN efficiency the following performance scores were used: area under the curve (AUC) (pROC package), accuracy, precision, Recall, and F1 score (caret package[36]). The K-folds cross-validation was used to avoid sampling bias and to estimate the performance scores (Figure 1). ANOVA test and Bonferroni's test for pairwise were used for the performance scores comparisons obtained from K-fold validation. A test was considered significant when the p value was lower than 0.05.

2.3.2 | Sampling process

Total data available was separated into two batches by random sampling, which allowed the creation of a training data group (70% of the data) and a validating data group (30% of the data). The SNNs were trained on the training group with K-fold cross validation. With the K-fold validation, the different performance scores are calculated K times, averages are then calculated and it is then easy to compare the scores using statistical tests and to rank the different neural networks according to the performance scores. The final SNNs were trained on the totality of the training data group and the validating data group was used to confirm the performance of the different final SNNs (Figure 1).[34] These results were then confirmed with the performance scores obtained with the final SNNs.

2.3.3 | Morphokinetic parameters

Since the aim was to obtain SNNs developed on real data obtained from routine practice, no time-lapse annotation was retrospectively performed on embryos. To determine the timing of cell division from the TLT system, each embryo images were acquired every 20 min at seven different focal planes, which enabled to determine the developmental events from syngamy to blastocyst stage. From these events, the morphokinetic parameters were obtained. The following morphokinetic parameters were annotated: tPNf as the time of fading of pronuclei, t2, t3, t4, t5, t6, t7 t8, t9 time for correspond cells number, tM as morula time, tSB as sub-blastocyst formation time, and tB as blastocyst formation time. The unit was hour, and the origin was the time of the sperm injection into oocyte. Each embryo has its own temporal origin, so the timing of events is comparable.

2.3.4 | Parameters used

As blastocysts were transferred, from the list of morphokinetics parameters (tPNF to tB), the late morphokinetic parameters were used: tM (time to morula stage), tSB (time to sub-blastocyst stage), and tB (time to blastocyst stage). These morphokinetic parameters allowed obtaining a blastocyst development chronology, eight ART parameters were added to the morphokinetic parameters: woman

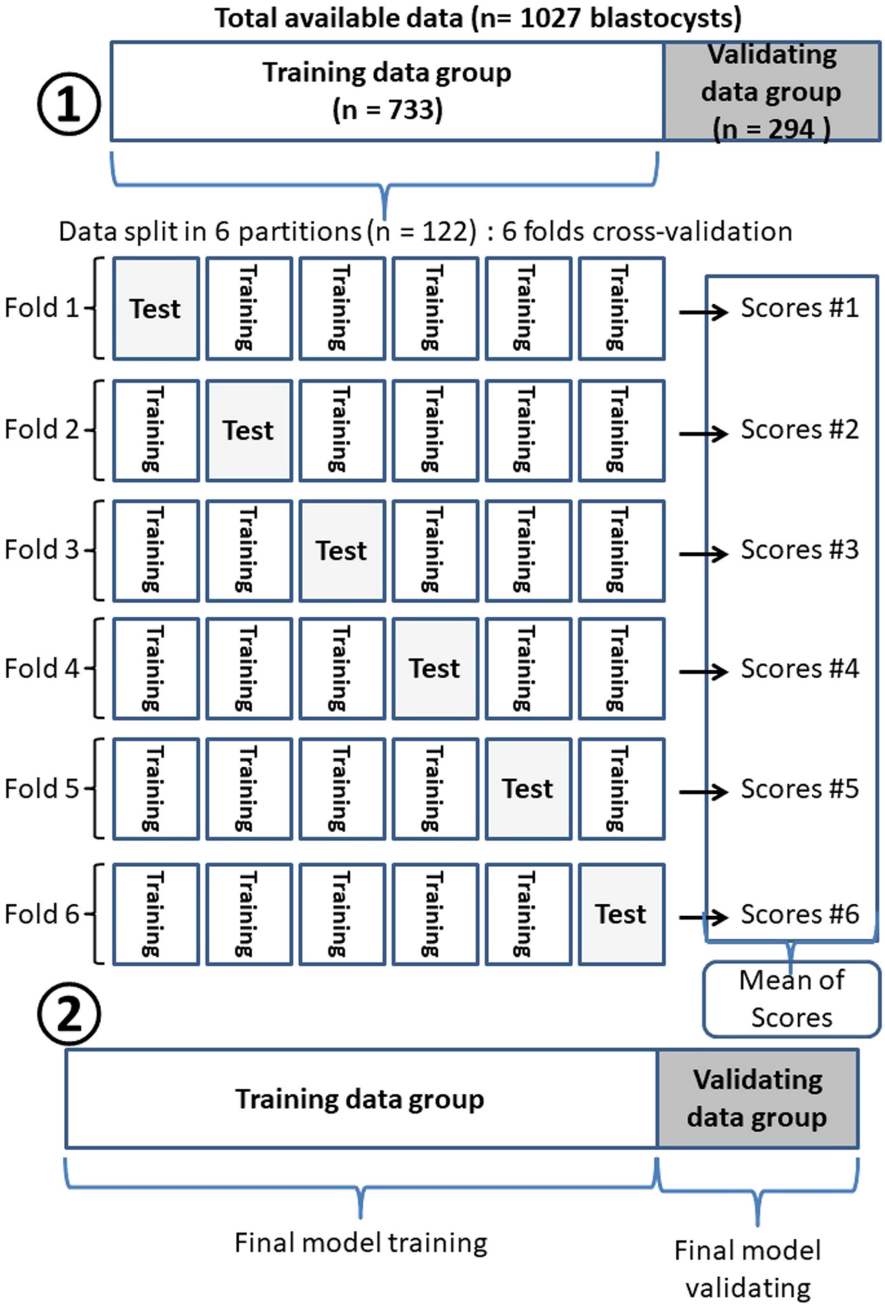

FIGURE 1 ① K-fold cross-validation: The total available data is randomly splitted in "Training data group" (n = 733 blastocysts) and "Validating data group" (n = 294 blastocysts). The "training data group" is splitted in K partitions (K = 6 in our case) of equal size (n = 122 blastocysts). For each partition k, the neural network is trained on the K-1 (training) partitions and tested on partition k (test). With the K-fold validation, the different scores (AUC, accuracy, precision, Recall, F1 score) are calculated K times, means are then calculated and the scores were compared using statistical tests and allow to rank the different neural networks according to the scores. ② the final SNNs were trained on the totality of the" training data group" (n = 733 blastocysts) and the "validating data group" (n = 294 blastocysts) was used to confirm the performance of the different final SNNs

age,[37,38] gonadotropin injected quantity,[37] retrieved oocyte number,[37] obtained embryos number,[37] fertilization rate,[37] obtained blastocyst number, blastulation rate[39] and transferred blastocyst number. The blastulation rate corresponds to blastocysts obtained number divided by embryos number obtained at D2, so eleven parameters were used to feed the SNNs. Among all embryos that have been cultured in the TLM and reached the blastocyst stage, only those whose result of implantation was known have been selected (KID status). For the classification of blastocyst according to KID status, ART procedures with a delivery in which the number of babies did not match the number of transferred blastocysts were excluded. Furthermore, if several blastocysts were transferred but no delivery was obtained, these cases were included in the study. These eleven parameters fed the different neural networks (MLP, simple RNN and LSTM-RNN) (Table 1).

3 | RESULTS

During the study period, 876 cycles have had embryos cultured in TLT incubators, among them, 654 cycles (1027 blastocysts) were included. The training data group was constituted by 458 cycles (733 blastocysts), and the validating data group by 196 cycles (294 blastocysts). The baseline and cycle characteristics for each data group are provided in Table 2. The live birth rate per fresh transfer was 26.0% for the training data group, and 25.0% for the validating data group (Table 2).

TABLE 2 Baseline and cycle characteristics according to delivery outcome (mean ± sd) for training and validating data groups

Training data group				
	Delivery			
	No	**Yes**		**Total**
Number of couples	***n* = 339**	***n* = 119**	***p* value**	***n* = 458**
ART parameters				
Female age (years)	34.5 ± 4.2	32.0 ± 3.5	0.0001	33.8 ± 4.2
Cumulative FSH dose (IU)	2642.2 ± 1139.3	2133.1 ± 912.7	0.0001	2509.9 ± 1106.8
Oocytes retrieved (*n*)	12.6 ± 6.9	12.7 ± 6.0	0.9452	12.7 ± 6.7
Fertilization rate (%)	69.6 ± 21.8	69.3 ± 19.2	0.8802	69.5 ± 21.1
Embryos obtained (*n*)	7.4 ± 5.0	6.9 ± 3.9	0.3123	7.3 ± 4.7
Blastocysts obtained (*n*)	3.5 ± 2.7	3.9 ± 2.6	0.1619	3.6 ± 2.7
Blastulation rate (%)	53.1 ± 26.4	61.3 ± 24.6	0.0022	55.2 ± 26.2
Transferred blastocysts (*n*)	1.6 ± 0.6	1.3 ± 0.5	0.0001	1.6 ± 0.6
Cryopreserved blastocysts (*n*)	1.5 ± 2.2	1.9 ± 1.9	0.0595	1.6 ± 2.1
Delivery rate (%)	–	–	–	26.0
Morphokinetic parameter				
tM (hours)	90.6 ± 10.8	88.3 ± 8.2	0.0422	89.8 ± 10.0
tSB (hours)	102.6 ± 10.0	97.8 ± 7.9	0.0001	101.1 ± 9.7
tB (hours)	110.8 ± 11.6	106.2 ± 10.1	0.0039	109.2 ± 11.3
Validating data group				
Number of couples	**Delivery**			
	No	**Yes**		**Total**
	***n* = 147**	***n* = 49**	***p* value**	***n* = 196**
ART parameters				
Female age (years)	33.6 ± 4.5	30.8 ± 3.7	0.0001	32.9 ± 4.5
Cumulative FSH dose (IU)	2631.8 ± 1157.6	2270.9 ± 1187.8	0.0675	2541.6 ± 1172.7
Oocytes retrieved (*n*)	11.4 ± 6.0	10.9 ± 4.4	0.5325	11.3 ± 5.6
Fertilization rate (%)	68.1 ± 21.8	71.6 ± 20.3	0.3091	69.0 ± 21.4
Embryos obtained (*n*)	6.5 ± 4.1	6.6 ± 3.2	0.9523	6.5 ± 3.9
Blastocysts obtained (*n*)	3.5 ± 2.9	3.9 ± 2.8	0.3762	3.6 ± 2.9
Blastulation rate (%)	55.5 ± 27.4	57.7 ± 25.7	0.6059	56.0 ± 26.9
Transferred blastocysts (*n*)	1.6 ± 0.6	1.2 ± 0.4	0.0001	1.5 ± 0.6
Cryopreserved blastocysts (*n*)	1.4 ± 2.3	2.1 ± 2.2	0.0350	1.6 ± 2.3
Delivery rate (%)	–	–	–	25.0
Morphokinetic parameter				
tM (hours)	92.3 ± 11.6	87.2 ± 6.4	0.0022	90.6 ± 10.4
tSB (hours)	102.3 ± 10.4	98.1 ± 6.1	0.0063	100.9 ± 9.3
tB (hours)	111.3 ± 11.1	104.8 ± 6.0	0.0007	108.7 ± 9.9

Abbreviations: FSH, follicle stimulating hormone; IU, international unit; sd, standard deviation; tB, blastocyst formation time; tM, morula formation time; tSB, sub-blastocyst formation time.

3.1 | Comparison of shallow networks with only morphokinetics parameters with: tM, tSB And tB

When only morphokinetics parameters (tM, tSB and tB) were used to feed the three shallow networks, no difference was observed for the AUCs values and the different performance scores. Theses AUCs means were bellows 0.700 with the values obtained with the K-folds cross validation (Table 3) and below 0.800 for the validating data group (Table 3, Figure 2).

TABLE 3 Neural networks performance scores in training and testing data groups calculating with K-fold cross validation, and neural networks performance score in validating data group, for predicting live birth using included parameters: tM, tSB and tB

K-folds cross validation	Neural networks	AUC (mean ± sd)	Accuracy (mean ± sd)	Precision (mean ± sd)	Recall (mean ± sd)	F1 score (mean ± sd)
Training data group	MLP	0.697 ± 0.013	0.700 ± 0.038	0.602 ± 0.065	0.579 ± 0.120	0.579 ± 0.058
	Simple RNN	0.693 ± 0.016	0.652 ± 0.036	0.525 ± 0.055	0.675 ± 0.107	0.583 ± 0.020
	LSTM-RNN	0.680 ± 0.013	0.677 ± 0.023	0.539 ± 0.042	0.627 ± 0.056	0.577 ± 0.016
Testing data group	MLP	0.643 ± 0.119	0.552 ± 0.072	0.452 ± 0.232	0.439 ± 0.231	0.390 ± 0.117
	Simple RNN	0.612 ± 0.119	0.523 ± 0.067	0.397 ± 0.126	0.522 ± 0.239	0.414 ± 0.104
	LSTM-RNN	0.603 ± 0.094	0.582 ± 0.056	0.468 ± 0.201	0.556 ± 0.184	0.474 ± 0.128
	Neural networks	**AUC [95% CI]**	**Accuracy [95% CI]**	**Precision**	**Recall**	**F1 score**
Validating data group	MLP	0.781 [0.673; 0.785]	0.700 [0.588; 0.804]	0.605	0.793	0.687
	Simple RNN	0.791 [0.680; 0.794]	0.714 [0.594; 0.816]	0.605	0.897	0.722
	LSTM-RNN	0.733 [0.620; 0.734]	0.657 [0.534; 0.767]	0.561	0.793	0.657

Abbreviations: 95% CI, 95% confidence interval; AUC, area under the curve, Included parameters in the neural networks; K-folds cross validation, mean ± standard deviation (sd) of model indicators provided with k-folds cross validation; LSTM-RNN, long short term memory recurrent neural network; MLP, multi layers perceptron; simple RNN, simple recurrent neural network; tB, blastocyst formation time in hours; tM, morula formation time in hours; tSB, sub-blastocyst formation time in hours.

3.2 | Comparison of shallow neural networks

When the eleven parameters were used to feed the three shallow neural networks, with the K-folds cross validation, in the training data group for all SNNs the means of AUC value was above 0.800, above the AUC values obtained with only morphokinetics parameters (Table 3, Table 4 and Figure 2). However, the lowest AUC value was obtained with the LSTM-RNN (AUC = 0.853 ± 0.014, $p < 0.05$). A statistically significant difference was observed for the accuracy values; the lowest accuracy value was obtained with LSTM-RNN (accuracy = 0.757 ± 0.017, $p < 0.05$). A statistically significant difference was observed for the precision values; the lowest precision value was obtained with LSTM-RNN (precision = 0.60 ± 0.041, $p < 0.05$). The Recall value was similar for all tested SNNs. A statistically significant difference was observed for the F1 score values; the lowest F1 score value was obtained with LSTM-RNN (F1 score = 0.721 ± 0.024, $p < 0.05$, Table 4).

With the K-folds cross validation, in the testing data group for all SNNs the AUC value was above 0.700, no statistical difference was observed for the AUC values. A statistically significant difference was observed for the accuracy values; the lowest accuracy value was obtained with LSTM-RNN (accuracy = 0.600 ± 0.078, $p < 0.05$). No statistical difference was observed among the SNNs for precision and Recall values. A statistically significant difference was observed for the F1 score values; the lowest F1 score value was obtained with LSTM-RNN (F1 score = 0.487 ± 0.172, $p < 0.05$, Table 4).

In the validating data group, the AUC values were above 0.800 for all SNNs and no statistical difference was observed for the AUCs; however, the highest AUC value was obtained with the MLP (AUC = 0.866). No difference was observed for the accuracy values; however, the highest accuracy value was obtained with MLP (accuracy = 0.798). The highest precision value was obtained with MLP (precision = 0.700), the highest Recall value was obtained with the LSTM-RNN (Recall = 0.969). The highest F1 score value was obtained with LSTM-RNN (F1 score = 0.785, Table 4).

4 | DISCUSSION

With our data, we have shown herein that it is possible to construct relevant SNNs for predicting live birth using clinical and laboratory data issued from the ART process coupled to morphokinetic data. All presented SNNs provided interesting results with included variables. Among the three SNNs, MLP and simple RNN networks provide the best results in training, testing data groups; however, all tested SNNs were similar in validating data group. The used SNNs provided AUCs close but lower than based on the blastocyst images analysis, whether classified with random forest algorithm,[40] or deep learning algorithm.[23,24,26]

Simple tree decision algorithms based only on morphokinetic data have been reported to predict the quality of blastocysts with an AUC reaching a maximum of 0.65,[41] 0.748 (Storr et al., 2015), and 0.762[22]; however, these algorithms were unable to predict live birth. Random forest could be considered as an extension of algorithms based on decision tree.[17–22] Random forest provide good results as it was previously shown,[42–44] and seem predictive of embryo implantation.[42–46] Others algorithms have been used for pregnancy prediction from blastocyst images. The logistic regression algorithm has been used and resulted in an AUC equal to 0.659.[47] The naive Bayes algorithm was used from blastocyst images with an accuracy of 58% for pregnancy prediction,[47] or from images of D2 or D3 transferred embryos with an accuracy of 80.4%[48] even 85.49%.[49] Blank[42] had

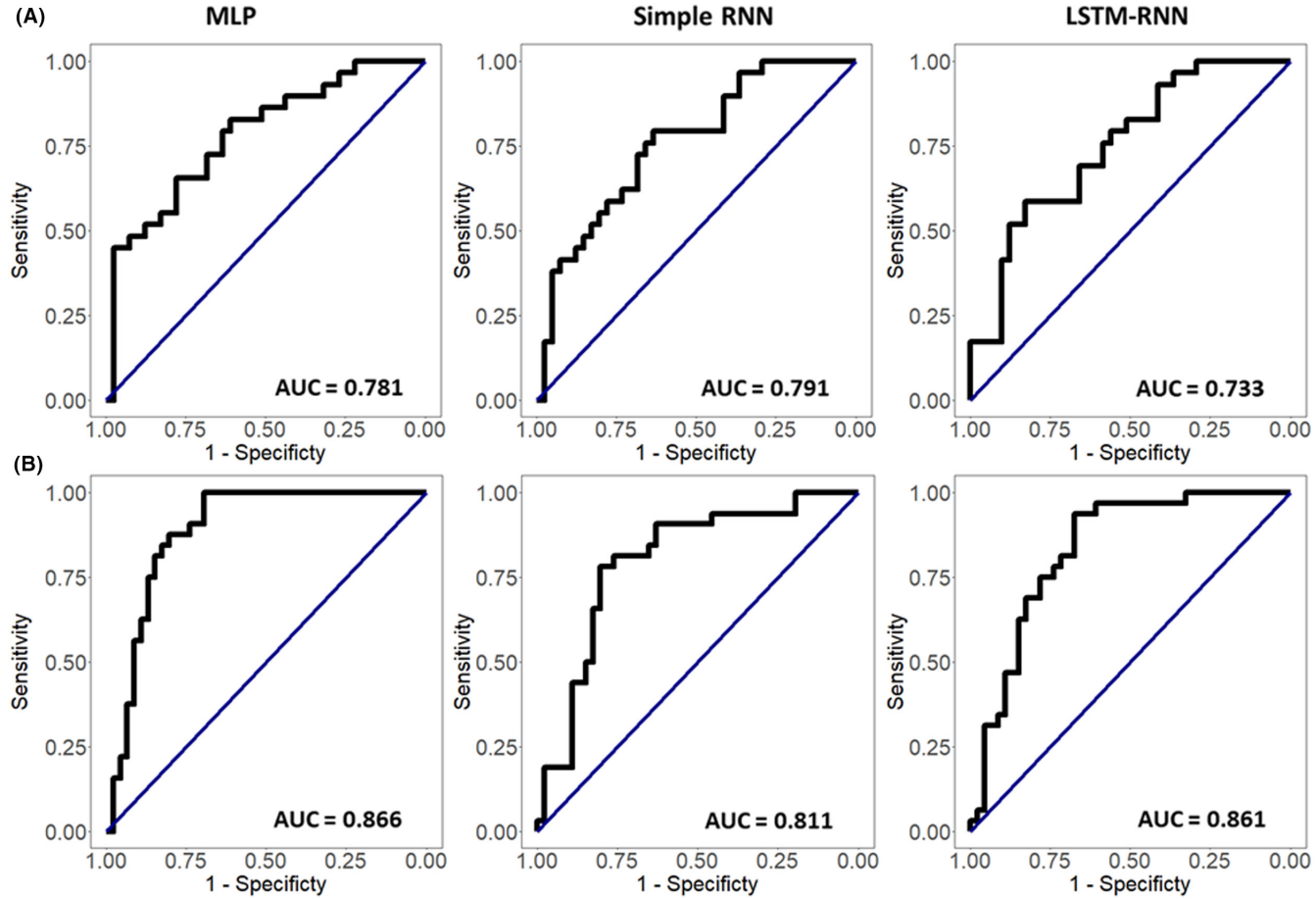

FIGURE 2 ROC curves for the different final neural network. These ROC curves reflect AUC with morphokinetic factors alone (tM, tSB and tB): (A), and with the addition of bio-clinical factors to the morphokinetic factors (woman age, gonadotropin injected quantity, retrieved oocyte number, obtained embryos number, fertilization rate, obtained blastocyst number, blastulation rate, transferred blastocyst number, tM, tSB and tB): (B); according to the three shallow neuronal networks: MLP: multi layers perceptron, RNN_S: simple recurrent neural network, LSTM-RNN: long short term memory recurrent neural network

obtained an AUC of 0.74 when random forest was used and 0.66 for multivariate logistic regression for the prediction of an evolving pregnancy. In Blank's study,[42] even if a greater number of morphokinetic and clinical variables were included, the AUCs were always lower than those obtained with SNNs. Bori et al.,[27] who have linked the morphokinetic data with information from image analysis, had constructed a pregnancy prediction algorithm with an AUC equal to 0.77.

The particularity of the SNNs presented herein, lies in the fact that their construction was based on the coupling of morphokinetic variables with variables issued from the ART procedure. The used of chronology of events with RNN (simple or LSTM) not allow obtaining better results than the MLP network, this means that in our study, the events chronology did not play a major role. However, the ART parameters used in the SNNs represent a key stage of ART process. By example age represents the initial state of the patient and it is known to be involved in the embryo late development[37,38] and to be linked to occurrence of a live birth. The gonadotropin amount is a consequence of the ovarian stimulation; and it can be seen as a reflection of the ovarian stimulation quality.[37] The blastulation rate represents the synthesis of both the clinician's and embryologist's involvement and the blastulation rate relates to the ART chances of success.[39] Among the selected morphokinetic parameters, the morula time seemed predictive of embryo implantation and this is consistent with the fact that this parameter has been shown to be predictive of live birth.[50] Similarly, late blastulation has been shown to be correlated with a drop in the chances of implantation,[51] this is why this morphokinetic parameter (tB), should be included. Thus, the addition of clinical and laboratory variables to embryo growth kinetics allows to accurately predict the embryos that can be implanted and maximize the chances of live birth. The necessity to add clinical and laboratory data to morphokinetic data to increase the AUC of the algorithms further underlines that embryo morphology and its growth dynamic would not be sufficient to predict the occurrence of a birth. This has been shown by Khosravi et al.[23] where the addition of maternal age to embryonic quality allows to predict the attainment of pregnancy through the use of a neural network.

Our study has many limitations, the first being its single-centre retrospective design. It would be necessary to test these SNNs using data from other centers in order to validate them. However,

TABLE 4 Neural networks performance scores in training and testing data groups calculating with K-fold cross validation, and neural networks performance scores in validating data group, for predicting live birth using: Female age, cumulative FSH dose, oocytes retrieved, fertilization rate, embryos obtained, blastocysts obtained, Blastulation rate, transferred blastocysts, tM, tSB, and tB

K-folds cross validation	Neural networks	AUC (mean ± sd)	Accuracy (mean ± sd)	Precision (mean ± sd)	Recall (mean ± sd)	F1 score (mean ± sd)
Training data group	MLP	0.904 ± 0.021	0.855 ± 0.024	0.760 ± 0.087	0.879 ± 0.095	0.808 ± 0.028
	Simple RNN	0.915 ± 0.017	0.849 ± 0.017	0.739 ± 0.061	0.889 ± 0.065	0.804 ± 0.022
	LSTM-RNN	0.853 ± 0.014*	0.757 ± 0.017*	0.603 ± 0.041*	0.902 ± 0.037	0.721 ± 0.024*
Testing data group	MLP	0.788 ± 0.067	0.764 ± 0.079	0.614 ± 0.311	0.599 ± 0.315	0.718 ± 0.064
	Simple RNN	0.766 ± 0.173	0.784 ± 0.108	0.664 ± 0.244	0.784 ± 0.185	0.688 ± 0.167
	LSTM-RNN	0.726 ± 0.145	0.600 ± 0.078*	0.444 ± 0.189	0.660 ± 0.284	0.487 ± 0.172*
	Neural networks	**AUC [95% CI]**	**Accuracy [95% CI]**	**Precision**	**Recall**	**F1 score**
Validating data group	MLP	0.866 [0.781; 0.868]	0.798 [0.688; 0.878]	0.700	0.875	0.778
	Simple RNN	0.811 [0.708; 0.812]	0.756 [0.646; 0.847]	0.667	0.813	0.732
	LSTM-RNN	0.861 [0.768; 0.865]	0.782 [0.674; 0.868]	0.660	0.969	0.785

Abbreviations: *, statistical difference between LSTM-RNN and MLP, and between LSTM-RNN and simple RNN; 95% CI, 95% confidence interval; AUC, area under the curve; FSH, follicle stimulating hormone; Included parameters in the neural networks: female age, cumulative FSH dose, number of oocytes retrieved, fertilization rate, number of embryos obtained, number of blastocysts obtained, blastulation rate, number of transferred blastocysts; K-folds cross validation, mean ± standard deviation (sd) of model indicators provided with k-folds cross validation; LSTM-RNN, long short term memory recurrent neural network; MLP, multi layers perceptron; simple RNN, simple recurrent neural network; tB, blastocyst formation time in hours; tM, morula formation time in hours; tSB, sub-blastocyst formation time in hours.

the use of K-fold cross validation improves the validity of the performance scores. Another limitation was that morphokinetic parameters were annotated by different persons, increasing the possibility of significant variance between morphokinetic measurements. In addition, there were missing data causing a loss of power in the statistical analysis, this bias was partially compensated by the number of included embryos. Data entry was dependent on humans and, therefore, required the involvement of the technical team. In the real world, data entry is time-consuming and adds to the ever-increasing workload, resulting in a decrease in the time that can be devoted to this task, hence the increased risk of missing data. One solution to avoid the problems of non-homogenization and missing data would be to obtain the morphokinetic parameters using automatic time recognition systems.[52] This would allow a homogenization of measurements of morphokinetic parameters and a complete automation regarding the choice of blastocyst to be transferred, which would in turn probably increase the chances of live birth as suggested by Fishel et al.[45]

In conclusion, SNN, are able to predict live birth by coupling morphokinetic data to clinical data. The next steps will be to use one of these SNN coupled to an automatic time recognition system as a support for a complete automation system for the choice of embryo(s) to be transferred.

ACKNOWLEDGMENT

To Hélène Boyer (*Hospices Civils de Lyon*) for help in manuscript preparation.

FUNDING INFORMATION

This study did not receive any specific grant from any funding agency in the public, commercial, or not-for-profit sector.

CONFLICT OF INTEREST

Mehdi Benchaib, Elsa Labrune, Sandrine Giscard d'Estaing, Bruno Salle, and Jacqueline Lornage declare that they have no conflict of interest.

ETHICAL APPROVAL

The protocol for the research project including human subjects has been approved by a suitably constituted Ethics Committee.

HUMAN RIGHTS STATEMENTS AND INFORMED CONSENT

All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1964 and its later amendments. Informed consent was obtained from all patients for being included in the study.

ORCID

Mehdi Benchaib https://orcid.org/0000-0001-5250-6036

REFERENCES

1. Steptoe PC, Edwards RG. Birth after the reimplantation of a human embryo. *Lancet Lond Engl*. 1978;2:366.
2. Pacchiarotti A, Selman H, Valeri C, Napoletano S, Sbracia M, Antonini G, et al. Ovarian stimulation protocol in IVF: an up-to-date review of the literature. *Curr Pharm Biotechnol*. 2016;17:303–15.
3. Baltz JM. Connections between preimplantation embryo physiology and culture. *J Assist Reprod Genet*. 2013;30:1001–7.
4. Ieda S, Akai T, Sakaguchi Y, Shimamura S, Sugawara A, Kaneda M, et al. A microwell culture system that allows group culture and is compatible with human single media. *J Assist Reprod Genet*. 2018;35:1869–80.

5. Biggers JD. IVF and embryo transfer: historical origin and development. *Reprod Biomed Online*. 2012;25:118–27.
6. Caanen MR, van der Houwen LE, Schats R, Vergouw CG, de Leeuw B, Lambers MJ, et al. Embryo transfer with controlled injection speed to increase pregnancy rates: a randomized controlled trial. *Gynecol Obstet Invest*. 2016;81:394–404.
7. Korkmaz C, Gül Yıldız Ü, Fidan U, Baykal B, Temel Ceyhan S, Ağaçayak E. Investigation of transfer results of human embryos that were vitrified and thawed at the cleavage, morula and blastocyst stages. *Zygote Camb Engl*. 2020;28:191–5.
8. Nagy ZP, Shapiro D, Chang C-C. Vitrification of the human embryo: a more efficient and safer in vitro fertilization treatment. *Fertil Steril*. 2020;113:241–7.
9. Ebner T, Yaman C, Moser M, Sommergruber M, Pölz W, Tews G. Embryo fragmentation in vitro and its impact on treatment and pregnancy outcome. *Fertil Steril*. 2001;76:281–5.
10. Sacha CR, Dimitriadis I, Christou G, Souter I, Bormann CL. The effect of day 2 versus day 3 embryo transfer on early pregnancy outcomes in women with a low yield of fertilized oocytes. *J Assist Reprod Genet*. 2018;35:879–84.
11. Gardner DK, Schoolcraft WB. Culture and transfer of human blastocysts. *Curr Opin Obstet Gynecol*. 1999;11:307–11.
12. Fesahat F, Montazeri F, Hoseini SM. Preimplantation genetic testing in assisted reproduction technology. *J Gynecol Obstet Hum Reprod*. 2020;49:101723.
13. Sciorio R, Tramontano L, Catt J. Preimplantation genetic diagnosis (PGD) and genetic testing for aneuploidy (PGT-A): status and future challenges. *Gynecol Endocrinol Off J Int Soc Gynecol Endocrinol*. 2020;36:6–11.
14. Mehta VP, Patel JA, Gupta RH, Shah SI, Banker MR. One plus one is better than two: cumulative reproductive outcomes are better after two elective single blastocyst embryo transfers compared to one double blastocyst embryo transfer. *J Hum Reprod Sci*. 2018;11:161–8.
15. Tiitinen A. Single embryo transfer: why and how to identify the embryo with the best developmental potential. *Best Pract Res Clin Endocrinol Metab*. 2019;33:77–88.
16. Curchoe CL, Bormann CL. Artificial intelligence and machine learning for human reproduction and embryology presented at ASRM and ESHRE 2018. *J Assist Reprod Genet*. 2019;36:591–600.
17. Meseguer M, Rubio I, Cruz M, Basile N, Marcos J, Requena A. Embryo incubation and selection in a time-lapse monitoring system improves pregnancy outcome compared with a standard incubator: a retrospective cohort study. *Fertil Steril*. 2012;98:1481–9.
18. Conaghan J, Chen AA, Willman SP, Ivani K, Chenette PE, Boostanfar R, et al. Improving embryo selection using a computer-automated time-lapse image analysis test plus day 3 morphology: results from a prospective multicenter trial. *Fertil Steril*. 2013;100:412–419.e5.
19. Basile N, Vime P, Florensa M, Aparicio Ruiz B, García Velasco JA, Remohí J, et al. The use of morphokinetics as a predictor of implantation: a multicentric study to define and validate an algorithm for embryo selection. *Hum Reprod Oxf Engl*. 2015;30:276–83.
20. Milewski R, Milewska AJ, Kuczyńska A, Stankiewicz B, Kuczyński W. Do morphokinetic data sets inform pregnancy potential? *J Assist Reprod Genet*. 2016;33:357–65.
21. Fishel S, Campbell A, Montgomery S, Smith R, Nice L, Duffy S, et al. Time-lapse imaging algorithms rank human preimplantation embryos according to the probability of live birth. *Reprod Biomed Online*. 2018;37:304–13.
22. Liu Y, Feenan K, Chapple V, Matson P. Assessing efficacy of day 3 embryo time-lapse algorithms retrospectively: impacts of dataset type and confounding factors. *Hum Fertil Camb Engl*. 2019;22:182–90.
23. Khosravi P, Kazemi E, Zhan Q, Malmsten JE, Toschi M, Zisimopoulos P, et al. Deep learning enables robust assessment and selection of human blastocysts after in vitro fertilization. *NPJ Digit Med*. 2019;2:21.
24. Tran D, Cooke S, Illingworth PJ, Gardner DK. Deep learning as a predictive tool for fetal heart pregnancy following time-lapse incubation and blastocyst transfer. *Hum Reprod Oxf Engl*. 2019;34:1011–8.
25. 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:e55301.
26. VerMilyea M, Hall JMM, Diakiw SM, Johnston A, Nguyen T, Perugini D, et al. Development of an artificial intelligence-based assessment model for prediction of embryo viability using static images captured by optical light microscopy during IVF. *Hum Reprod Oxf Engl*. 2020;35:770–84.
27. Bori L, Paya E, Alegre L, Viloria TA, Remohi JA, Naranjo V, Meseguer M Novel and conventional embryo parameters as input data for artificial neural networks: an artificial intelligence model applied for prediction of the implantation potential. *Fertil Steril [Internet]* 2020;114, 1232–1241. [cited 2020 Nov 2]; Available from: http://www.sciencedirect.com/science/article/pii/S0015028220307779
28. Rocha JC, Passalia FJ, Matos FD, Takahashi MB, de Souza Ciniciato D, Maserati MP, et al. A method based on artificial intelligence to fully automatize the evaluation of bovine blastocyst images. *Sci Rep*. 2017;7:7659.
29. 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;27:140–6.
30. Fishel S, Baker D, Elson J, Ragunath M, Atkinson G, Shaker A, et al. Precision medicine in assisted conception: a multicenter observational treatment cohort study of the annexin A5 M2 haplotype as a biomarker for antithrombotic treatment to improve pregnancy outcome. *EBioMedicine*. 2016;10:298–304.
31. R Core Team. *R: a language and environment for statistical computing [internet]*. Vienna, Austria: R Foundation for Statistical Comput Secur; 2021. Available from: https://www.R-project.org/
32. Allaire JJ, Chollet F. keras: R Interface to "Keras" [Internet] 2020. Available from: https://CRAN.R-project.org/package=keras
33. Allaire JJ, Tang Y. tensorflow: R Interface to "TensorFlow" [Internet]. 2020. Available from: https://CRAN.R-project.org/package=tensorflow
34. Chollet F. Deep learning with R/François Chollet; with J.J. Allaire (ed). *Deep learn. R*. Shelter Island, New York: Manning Publications; 2018.
35. Karsoliya S. Approximating number of hidden layer neurons in multiple hidden layer BPNN architecture. *Int J Eng Trends Technol*. 2012;3:715–7.
36. Kuhn M. caret: Classification and Regression Training [Internet]. 2020. Available from: https://CRAN.R-project.org/package=caret
37. Labrune E, Mery L, Lornage J, Aknin I, Guérin JF, Benchaib M. An ART score to note objectively the quality of an ART procedure. *Eur J Obstet Gynecol Reprod Biol*. 2018;221:52–7.
38. Warshaviak M, Kalma Y, Carmon A, Samara N, Dviri M, Azem F, et al. The effect of advanced maternal age on embryo Morphokinetics. *Front Endocrinol*. 2019;10:686.
39. Racowsky C. High rates of embryonic loss, yet high incidence of multiple births in human ART: is this paradoxical? *Theriogenology*. 2002;57:87–96.
40. Chavez-Badiola A, Flores-Saiffe Farias A, Mendizabal-Ruiz G, Garcia-Sanchez R, Drakeley AJ, Garcia-Sandoval JP. Predicting pregnancy test results after embryo transfer by image feature extraction and analysis using machine learning. *Sci Rep*. 2020;10:4394.
41. Barrie A, Homburg R, McDowell G, Brown J, Kingsland C, Troup S. Examining the efficacy of six published time-lapse imaging embryo selection algorithms to predict implantation to demonstrate

the need for the development of specific, in-house morphokinetic selection algorithms. *Fertil Steril*. 2017;107:613–21.
42. Blank C, Wildeboer RR, DeCroo I, Tilleman K, Weyers B, de Sutter P, et al. Prediction of implantation after blastocyst transfer in in vitro fertilization: a machine-learning perspective. *Fertil Steril*. 2019;111:318–26.
43. Fishel S, Campbell A, Montgomery S, Smith R, Nice L, Duffy S, et al. Live births after embryo selection using morphokinetics versus conventional morphology: a retrospective analysis. *Reprod Biomed Online*. 2017;35:407–16.
44. Reignier A, Girard J-M, Lammers J, Chtourou S, Lefebvre T, Barriere P, et al. Performance of day 5 KIDScore™ morphokinetic prediction models of implantation and live birth after single blastocyst transfer. *J Assist Reprod Genet*. 2019;36:2279–85.
45. Fishel S, Campbell A, Foad F, Davies L, Best L, Davis N, et al. Evolution of embryo selection for IVF from subjective morphology assessment to objective time-lapse algorithms improves chance of live birth. *Reprod Biomed Online*. 2020;40:61–70.
46. Adolfsson E, Porath S, Andershed AN. External validation of a time-lapse model; a retrospective study comparing embryo evaluation using a morphokinetic model to standard morphology with live birth as endpoint. *JBRA Assist Reprod*. 2018;22:205–14.
47. 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;18:204–11.
48. Uyar A, Bener A, Ciray HN. Predictive modeling of implantation outcome in an in vitro fertilization setting: an application of machine learning methods. *Med Decis Mak Int J Soc Med Decis Mak*. 2015;35:714–25.
49. Morales DA, Bengoetxea E, Larrañaga P. Selection of human embryos for transfer by Bayesian classifiers. *Comput Biol Med*. 2008;38:1177–86.
50. Rienzi L, Cimadomo D, Delgado A, Minasi MG, Fabozzi G, Gallego RD, et al. Time of morulation and trophectoderm quality are predictors of a live birth after euploid blastocyst transfer: a multicenter study. *Fertil Steril*. 2019;112:1080–93.
51. Kimelman D, Confino R, Okeigwe I, Lambe-Steinmiller J, Confino E, Shulman LP, et al. Assessing the impact of delayed blastulation using time lapse morphokinetics and preimplantation genetic testing in an IVF patient population. *J Assist Reprod Genet*. 2019;36:1561–9.
52. Feyeux M, Reignier A, Mocaer M, Lammers J, Meistermann D, Barrière P, et al. Development of automated annotation software for human embryo morphokinetics. *Hum Reprod Oxf Engl*. 2020;35:557–64.

How to cite this article: Benchaib M, Labrune E, Giscard d'Estaing S, Salle B, Lornage J. Shallow artificial networks with morphokinetic time-lapse parameters coupled to ART data allow to predict live birth. *Reprod Med Biol*. 2022;21:e12486. doi: 10.1002/rmb2.12486