

# Automated Evaluation of Human Embryo Blastulation and Implantation Potential using Deep-Learning

Yoav Kan-Tor, Nir Zabari, Ity Erlich, Adi Szeskin, Tamar Amitai, Dganit Richter, Yuval Or, Zeev Shoham, Arye Hurwitz, Iris Har-Vardi, Matan Gavish, Assaf Ben-Meir,\* and Amnon Buxboim\*

In *in vitro* fertilization (IVF) treatments, early identification of embryos with high implantation potential is required for shortening time to pregnancy while avoiding clinical complications to the newborn and the mother caused by multiple pregnancies. Current classification tools are based on morphological and morphokinetic parameters that are manually annotated using time-lapse video files. However, manual annotation introduces interobserver and intraobserver variability and provides a discrete representation of preimplantation development while ignoring dynamic features that are associated with embryo quality. A fully automated and standardized classifiers are developed by training deep neural networks directly on the raw video files of >6200 blastulation-labeled and >5500 implantation-labeled embryos. Prediction of embryo implantation is more accurate than the current state-of-the-art morphokinetic classifier. Embryo classification improves with video length where the most predictive images show only partial association with morphological features. Deep learning substitute to human evaluation of embryo developmental competence thus contributes to implementing single embryo transfer methodology.

## 1. Introduction


During *in vitro* fertilization (IVF) cycles, multiple mature oocytes are retrieved from the ovary and are fertilized in the lab. The newly generated embryos can be transferred into the uterus on day-3, -4, or -5 of incubation, cryopreserved for subsequent transfers or discarded. Lacking a reliable noninvasive evaluation method of the potential to implant, pregnancy rates can be improved by cotransferring multiple embryos thus introducing health risks that are associated with multiple pregnancies.<sup>[1]</sup> Hence, the evaluation of embryo quality is required for improving live birth rates while minimizing medical complications and shortening time to pregnancy.<sup>[2–6]</sup> Machine learning was used for assessing the potential of embryos to blastulate<sup>[7,8]</sup> and to implant<sup>[9–11]</sup> based on manually annotated morphological and/or morpho-

kinetic features. Deep learning, which offers a powerful toolbox for carrying out automated and standardized classification tasks

Y. Kan-Tor, N. Zabari, I. Erlich, A. Szeskin, T. Amitai, Dr. M. Gavish, Dr. A. Buxboim  
School of Computer Science and Engineering  
The Hebrew University of Jerusalem  
Rothberg Family Buildings, The Edmond J. Safra Campus, Jerusalem 9190416, Israel  
E-mail: amnon.buxboim@mail.huji.ac.il

Y. Kan-Tor, Dr. A. Buxboim  
Center for Interdisciplinary Data Science Research  
The Hebrew University of Jerusalem  
Jerusalem 9190416, Israel

A. Szeskin, Dr. A. Buxboim  
The Alexander Grass Center for Bioengineering  
The Hebrew University of Jerusalem  
Edmond J. Safra Campus–Givat Ram, Jerusalem 9190401, Israel

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/aisy.202000080>.

© 2020 The Authors. Published by WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1002/aisy.202000080

D. Richter, Dr. I. Har-Vardi  
Department of Obstetrics and Gynecology  
Soroka University Medical Center  
Ben-Gurion University of the Negev  
POB 653, Beer Sheva, Israel

Dr. Y. Or, Prof. Z. Shoham  
Fertility and IVF Unit  
Kaplan Medical Center  
The Hebrew University of Jerusalem  
POB 1, Rehovot 76100, Israel

Dr. A. Hurwitz, Dr. A. Ben-Meir  
Fertility and IVF Unit  
Department of Obstetrics and Gynecology  
The Hebrew University of Jerusalem  
Hadassah Medical Center  
Kiryat Hadassah POB 12000, Jerusalem 91120, Israel  
E-mail: assafb@hadassah.org.il, assaf.benmeir@gmail.com

Dr. A. Buxboim  
The Institute of Life Sciences  
The Hebrew University of Jerusalem  
Edmond J. Safra Campus–Givat Ram, Jerusalem 9190401, Israel

involving expansive datasets, is gradually incorporated into the health-care system worldwide<sup>[12–14]</sup> and IVF clinics specifically.<sup>[15]</sup> Deep learning was used for performing automated morphokinetic annotation<sup>[16]</sup> and morphological scoring of blastocysts based on the last frame that was acquired just before transfer.<sup>[17]</sup> Tran et al. also used deep learning for predicting the implantation outcome of day-5 embryos by training on the time-lapse images. However, negative implantation outcome was mislabeled by including discarded embryos that were not transferred to the uterus (80% of the entire dataset).<sup>[18]</sup> A recent publication reports the prediction of implantation of day-5 transferred embryos using single “static” images of blastocyst stage.<sup>[19]</sup> By training a deep neural network (DNN) using whole-video time-lapse images of preimplantation embryo development, we achieved high accuracy and provide an early prediction of embryo blastulation and embryo implantation already on day-3.

## 2. Results

### 2.1. The SHIFRA Database

We conducted a retrospective study in which we trained DNNs directly on the raw video files to generate automated and standardized classification algorithms of embryo blastulation and implantation. DNN training requires large datasets of labeled embryos. Therefore, we assembled and compiled a database that we term “SHIFRA” by collecting video files of >20 000 fresh embryos that were cultured in nine time-lapse incubators (Embryoscope, Vitrolife) during the past 6 years in four medical centers dispersed across Israel (Figure S1a and Table S1, Supporting Information). Our data-providing clinics accept patients with a heterogeneous ethnic and racial backgrounds (Eastern and Western European Jews, North African and Middle Eastern Jews, Arabs, and others)<sup>[20]</sup> and span different maternal age groups (Figure S1, Supporting Information), thus decreasing the effect of confounding variables and increasing embryo classification generality. Seven-frame z-stacks, 15  $\mu$ m apart, were recorded at 18-to-20 min intervals for up to 6 days of incubation, providing a continuous 3D imaging of preimplantation embryo development (Figure 1a). Based on time-lapse imaging, morphokinetic profiles of 16 000 embryos were annotated, specifying the time series of discrete events (Figure S2b-i, Supporting Information): pronuclei appearance and fading (tPNa/f), cleavage of *N* cells (tN; *N* = 2 to 9), morula compaction (tM), and start of blastulation (tSB). Morphokinetic annotations were determined by qualified and trained embryologists according to established protocols (2-to-3 annotators).<sup>[2,21]</sup> Quality assurance (QA) of morphokinetic annotations was conducted in a blindly fashion by an additional expert embryologist (Figure S2a, Supporting Information). The temporal intervals between consecutive morphokinetic events were included as well, showing temporal separation between cleavage cycles (Figure S2b-ii, Supporting Information).

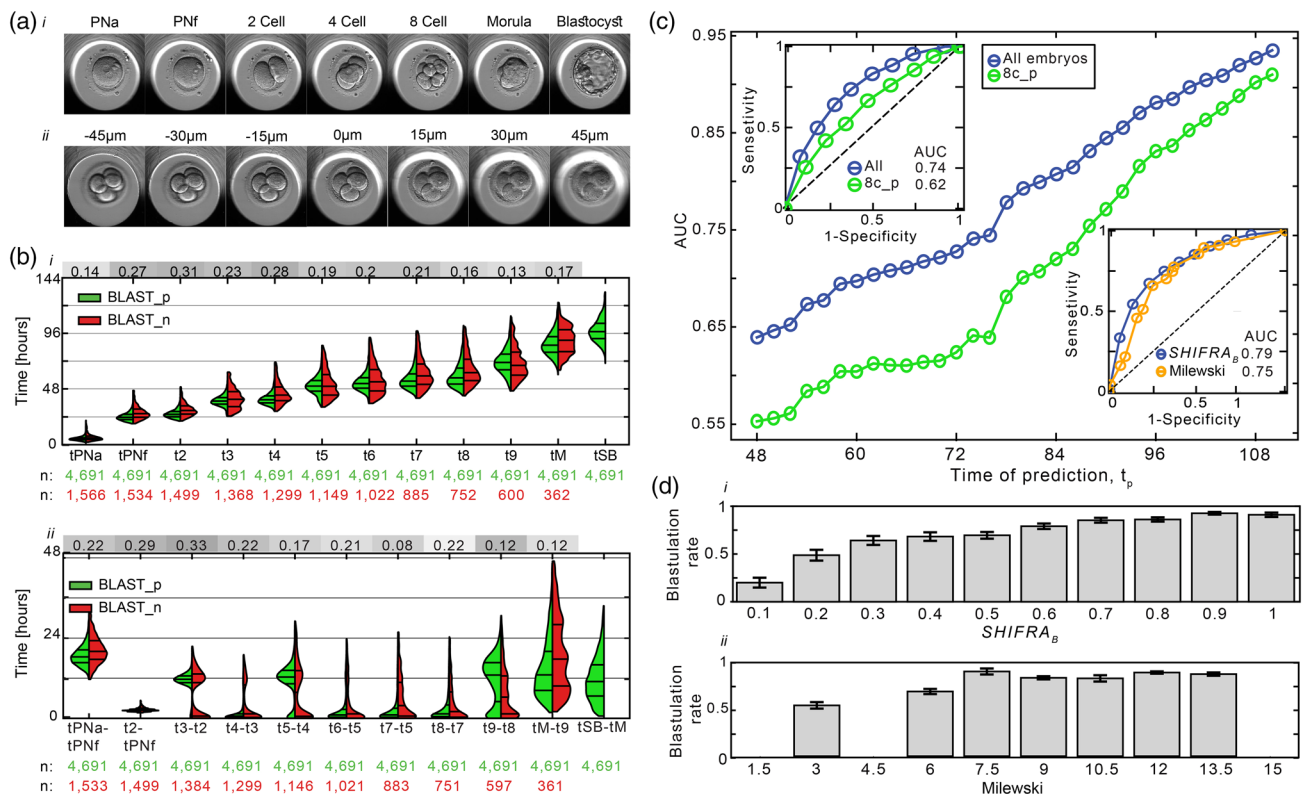
Labeling of embryo blastulation is based on morphokinetic histories. Embryos that reached start of blastulation (tSB) inside the incubator were labeled blastulation-positive (BLAST<sub>p</sub>). BLAST-negative (BLAST<sub>n</sub>) embryos that had been arrested at earlier developmental states and BLAST-unknown (BLAST<sub>u</sub>)

embryos were identified by projecting their morphokinetic profiles onto the time windows that permit transitioning from one embryo state to the next (Figure S3a, Table S2 and S3, Supporting Information; see Experimental Section). The SHIFRA database contains additional metadata, including maternal age, day-of-transfer, and co-transferred embryo statistics (Figure S1a,b, Supporting Information). Known Implantation Data (KID) labeling was determined based on embryo transfer statistics and the number of gestational sacs and fetal heart beats as measured on weeks 5–7 of pregnancy (see Experimental Section). Positive and negative implantation outcomes (KID<sub>p</sub> and KID<sub>n</sub>) refer to embryos that were successfully implanted or failed to implant, respectively, and embryos whose implantation outcome was uncertain are labeled KID unknown (KID<sub>u</sub>; Figure S1c, Supporting Information).

### 2.2. Early Prediction of Blastocyst Formation

The potential of an embryo to undergo blastulation marks its developmental quality and is linked with its potential to implant in the uterus.<sup>[22]</sup> Despite the fact that BLAST<sub>n</sub> embryos were arrested at earlier stages, their temporal overlap with the morphokinetic profiles of BLAST<sub>p</sub> embryos is large (KS < 0.28] Two-sample Kolmogorov Smirnov test quantifies the distance between two distributions.]; Figure 1b-i,ii). We divided a total of >6200 BLAST-labeled embryos into a train-validation set and a strictly uncontaminated test set of randomly selected 28% of the embryos. A fully automated blastulation classifier termed as SHIFRA<sub>B</sub> was designed as consisting of two parts (see Experimental Section). *Packet learning*: A DNN was trained directly on the time-lapse image packets of the embryos outputting a scalar value for each frame (temporal features) from the start of fertilization to time of prediction (*t<sub>p</sub>*). *Embryo learning*: The potential to blastulate was evaluated based on the packet scores (temporal features) of each embryo by training a Random Forest classifier. Random Forest was chosen as it provided the most accurate predictions compared with other classifiers that are suited for this task (XG-Boost and Logistic Regression).

Predictive strength was quantified using the area under the curve (AUC) of the receiver operating characteristic (ROC). Blastulation prediction AUC, which was evaluated for test-set embryos, increased monotonically with time of prediction, *t<sub>p</sub>* (Figure 1c): 0.65 at 48 h, 0.73 at 72 h, 0.88 at 96 h, and 0.94 at 110 h. To confirm that SHIFRA<sub>B</sub> can significantly predict blastulation also of high-quality embryos, we calculated the AUC for a cohort of embryos that reached at least 8-cell cleavage state (8C<sub>p</sub>). As expected, BLAST-prediction AUC of 8C<sub>p</sub> embryos was lower than for total embryo population, yet it reached 0.63 at 72 h, 0.84 at 96 h, and 0.91 at 110 h (Figure 1c inset). Automated blastulation prediction by SHIFRA<sub>B</sub> was compared with the manual morphokinetic classifier developed by Milewski et al. for five-cells positive (5C<sub>p</sub>) embryos.<sup>[8]</sup> AUC values were comparable with small advantage to SHIFRA<sub>B</sub>. Blastulation rate shows a near-perfect monotonic correlation with SHIFRA<sub>B</sub> prediction but not with Milewski prediction (Figure 1d). In the clinic, binary classification is obtained by setting the threshold values of SHIFRA<sub>B</sub> that define negative prediction (below threshold) and positive prediction (above threshold).



**Figure 1.** Automated prediction of embryo blastulation (SHIFRA<sub>B</sub>). a-i) Time-lapse image acquisition of preimplantation embryo development is performed at 18-to-20 min intervals for up to 6 days of culture inside a time-lapse incubator. ii) At each time point, a z-stack of seven focal planes 15 μm apart is recorded. b) High-resolution temporal distributions of the i) morphokinetic events and ii) intervals between consecutive events of positive and negative blastulation-labeled embryos (BLAST\_p and BLAST\_n) are evaluated based on thousands of annotated profiles. The temporal overlap between BLAST\_p and BLAST\_n distributions is quantified by K-S distances (top rows). c) AUC of automated blastulation prediction by SHIFRA<sub>B</sub> of test-set embryos (blue  $n = 1621$ ) and high-quality embryos that reached 8-cells (8C\_p; green  $n = 1456$ ) with video length  $\geq 110$  h increase monotonically with time of prediction,  $t_p$ . Left inset: ROC curves and AUC values calculated at  $t_p = 72$  h of test-set embryos and 8C\_p embryos with sufficiently long videos. Right inset: A comparison between SHIFRA<sub>B</sub> and Milewski BLAST classifier at  $t_p = 86$  h.<sup>[8]</sup> d-i) Blastulation rate of test-set embryos is plotted versus i) SHIFRA<sub>B</sub> prediction at 72 h and ii) Milewski prediction. The monotonic increase in blastulation rate is scored by Spearman's rank coefficient of correlation. SHIFRA<sub>B</sub>: 0.99 ( $p$ -value  $< 0.001$ ); Milewski: 0.80 ( $p$ -value = 0.01). K-S: Kolmogorov Smirnov. ROC: Receiver operating characteristic. AUC: Area under the ROC curve.

Herein, we present retrospective blastocyst prediction statistics by setting two threshold values that support positive predictive value 0.91 (Figure S8a-i, Supporting Information) and sensitivity 0.98 (Figure S8a-ii, Supporting Information). To test generality, we conducted a fivefold stratified cross-validation yielding an average AUC =  $0.83 \pm 0.02$ STD (Figure S6a, Supporting Information). In addition, we compared the blastulation prediction of test-set embryos from each clinic separately (AUC =  $0.78 \pm 0.07$ STD; Figure S6b, Supporting Information) and performed leave-one-clinic-out cross-validation where classifiers were trained on embryos from three clinics and tested on the fourth (Figure S6c, Supporting Information). As H3 clinic is the largest data provider of BLAST-labeled embryos, H3 AUC was smaller than the other clinics. The above-average AUC of H4 is likely skewed by the small number of and the uneven ratio between BLAST\_p and BLAST\_n test-set embryos. Temporal comparison between the morphokinetic events and intervals of embryos that were obtained from young (age  $< 32$ ) and older (age  $> 38$ ) women showed negligible differences (KS  $< 0.08$ )

except for time of start blastulation (tSB), which occurred on average 3 h faster, and tM–tSB interval, which was 1.5 h shorter in embryos derived from young women (Figure S9, Supporting Information). To verify that blastulation prediction was not confounded by maternal age, we divided all test-set embryos to four age groups and calculated AUC separately for each (Figure S6d, Supporting Information). Indeed, blastulation prediction AUC was comparable between maternal age groups: AUC =  $0.75 \pm 0.03$  STD. Collectively, we establish accuracy, robustness, and generality of a fully automated day-3 prediction of embryo blastulation.

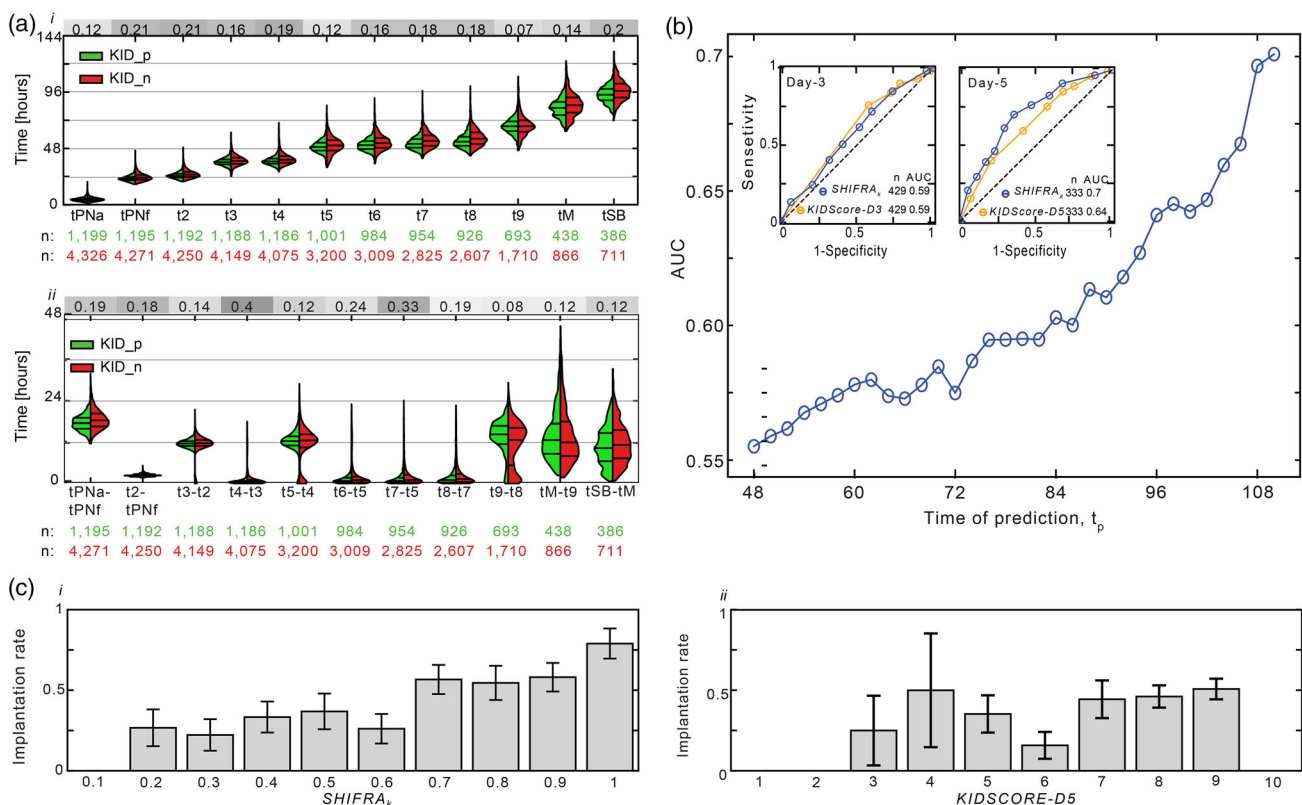
### 2.3. Prediction of Embryo Implantation

Relative to blastulation prediction, KID prediction is required to overcome two major obstacles. 1) Unlike blastulation, which depends on the capacity of the embryo to develop in the incubator under controlled conditions, implantation also depends on endometrial receptivity—a parameter that is not accounted for during

training. 2) Training is limited to embryos that had been preselected for transfer according to existing morphological and/or morphokinetic protocols. As a result, training is restricted to a dataset of morphokinetically homogenous KID-labeled embryos. Accordingly, the temporal distributions of morphokinetic events and intervals of KID\_p and KID\_n embryos are almost fully overlapping (Figure 2a-i,ii). We divided the dataset of >5500 KID-labeled embryos into a train-validation set and an uncontaminated test set that consisted of randomly selected 21% of the embryos (Figure S1c, Supporting Information). We designed an automated two-stage KID classifier that consisted of packet-learning followed by embryo learning. Similar to SHIFRA<sub>B</sub>, we term the KID classifier as SHIFRA<sub>K</sub>. To improve predictive strength and robustness, packet learning combined three DNNs, which were trained separately on KID-labeled embryos, and one DNN that was trained on BLAST-labeled embryos. In all cases, we used the same preprocessing steps and all networks had the same architecture (Figure S5, Supporting Information). The temporal features were obtained by summing together the four scalar values that were generated by each network for each frame. Embryo learning was performed using a logistic regression classifier that was trained on all the

temporal features that belong to each embryo (see Experimental Section).

KID predictive strength increased with time of prediction  $t_p$ , as evaluated for the same cohort of day-5 transferred test-set embryos (Figure 2b). AUC increases slowly from 48 to 84 h and more rapidly from 84 h onward. Compared with the manual-morphokinetic KIDScore decision support tools (Vitrolife), embryo implantation prediction by SHIFRA<sub>K</sub> is as accurate as KIDScore-D3 on day-3 and more accurate than KIDScore-D5 on day-5 as evaluated for the same test-set embryos (Figure 2b-inset). Day-5 predictive strength of SHIFRA<sub>K</sub> remains high despite the fact that 98% of the transferred embryos were blastocysts (very high-quality embryos) and endometrial receptivity was likely an important factor. Prediction of implantation of day-5 transferred embryos reaches 0.71 (calculated on test-set embryos). Satisfyingly, implantation rate positively correlates with SHIFRA<sub>K</sub>, thus demonstrating its clinical utility (Figure 2c-i). In comparison, monotonic correlation with implantation rate is not demonstrated by the manual classifier KIDScore-D5 (Figure 2c-ii). We demonstrate retrospective implantation prediction statistics by setting SHIFRA<sub>K</sub> threshold values that support positive predictive value 0.59 (Figure S8b-i, Supporting Information) and sensitivity 0.89



**Figure 2.** Automated prediction of embryo implantation (SHIFRA<sub>K</sub>). a) High-resolution temporal distributions of the i) morphokinetic events and ii) intervals between consecutive events of positive- and negative-known implantation data-labeled (KID\_p and KID\_n) embryos are evaluated based on thousands of annotated profiles. KID\_p and KID\_n distributions are almost indistinguishable as quantified by K-S distances (top rows). b) AUC of automated KID prediction by SHIFRA<sub>K</sub> of day-5 transferred embryos ( $n = 359$ ; 355 transferred blastocysts) increases with prediction time,  $t_p$ . Inset: ROC curves of automated KID prediction by SHIFRA<sub>K</sub> at 68 h (left) and at 110 h (right) are compared with implantation prediction by KIDScore-D3 and KIDScore-D5 manual-morphokinetic algorithms, respectively. c) Implantation rate of test-set embryos increases monotonically with i) day-5 SHIFRA<sub>K</sub> prediction (Spearman's rank correlation coefficient: 0.89 [ $p$ -value = 0.01]) but not with ii) KIDScore-D5 (Spearman's rank correlation coefficient: 0.66 [ $p$ -value = 0.07]). K-S: Kolmogorov Smirnov. ROC: Receiver operating characteristic. AUC: Area under the ROC curve. PPV: positive predictive value.



(Figure S8b-ii, Supporting Information). Generality of SHIFRA<sub>K</sub> is verified first via fivefold cross-validation of day-5 transferred embryos, showing negligible variation of KID prediction: AUC =  $0.69 \pm 0.02$  STD (Figure S7a, Supporting Information). Next we compared KID prediction of test-set embryos obtained from different clinics, giving AUC =  $0.64 \pm 0.04$  STD (Figure S7b, Supporting Information). AUC values by H1 (above average) and H2 (below average) clinics were likely skewed due to a highly uneven ratio between KID<sub>p</sub> and KID<sub>n</sub> train-set embryos (Figure S1c, Supporting Information). We also tested generality via leave-one-clinic-out cross-validation: AUC =  $0.64 \pm 0.04$  STD (Figure S7c, Supporting Information). The below-average AUC by H3 is consistent with the smallest available train-set once H3 embryos were removed. The fraction of implanted embryos out of all transferred embryos (considering also KID<sub>u</sub> transfers) was 2.7-fold higher for young women (age < 32; 39%) than older women (age > 38, 14%). To verify that SHIFRA<sub>K</sub> is not biased by maternal age, we divided the embryos into four age groups and tested KID-prediction on each. Satisfyingly, variation in day-5 KID prediction across age groups was small: average AUC =  $0.75 \pm 0.03$  STD (Figure S7d, Supporting Information). We thus confirm automated prediction of embryo implantation with superior predictive strength of day-5 transfers and verify generality across clinics and age groups.

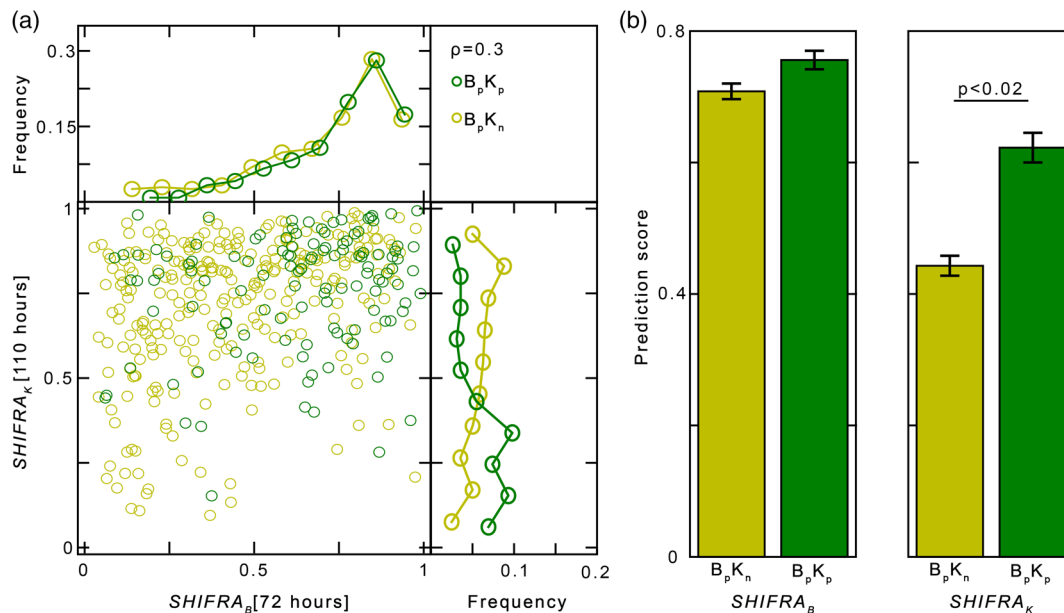
#### 2.4. SHIFRA<sub>K</sub> Differentiates between Implanted and Nonimplanted Blastocysts

Aneuploidy might impair embryo implantation but permit blastocyst formation.<sup>[23]</sup> To study the relationship between blastulation and implantation potentials, we analyze the classification

of blastulation and implantation colabeled embryos: 121 BLAST<sub>p</sub>-KID<sub>n</sub> (B<sub>p</sub>K<sub>n</sub>) embryos and 275 BLAST<sub>p</sub>-KID<sub>p</sub> (B<sub>p</sub>K<sub>p</sub>) embryos (Figure 3a). Blastulation and implantation prediction scores are weakly positively correlated (Pearson correlation score 0.3), which is indicative of common visual elements. Consistent with their BLAST<sub>p</sub> labels, blastulation prediction score distributions of B<sub>p</sub>K<sub>n</sub> and B<sub>p</sub>K<sub>p</sub> embryos overlap. In contrast, average implantation scores of the latter embryos were 40% higher than B<sub>p</sub>K<sub>n</sub> embryos (Figure 3b), indicating that SHIFRA<sub>K</sub> differentiates between blastocysts that have the capacity to implant and blastocysts that do not.

#### 2.5. The last <10 h Are Sufficient for Prediction of Embryo Quality

To gain additional insight into the underlying mechanisms of embryo classification, we analyze the visual information that is embedded within the time-lapse images that SHIFRA<sub>B</sub> and SHIFRA<sub>K</sub> are sensitive to. To this end, we use the SHapley Additive exPlanations (SHAP) methodology for quantifying the impact of the temporal features on embryo prediction.<sup>[24]</sup> We identify the frames that contribute the most to accurate embryo classification by setting the sign of the SHAP values of each feature according to the BLAST or KID label of the embryos (negative: -1; positive: +1) and averaged across embryos (mean adjusted SHAP). In this manner, positive (negative) SHAP values of temporal features of positively (negatively) labeled embryos contribute oppositely to the mean-adjusted SHAP compared with positive (negative) SHAP values of temporal features of negatively (positively) labeled embryos. We find that the contribution of the temporal features to blastulation prediction

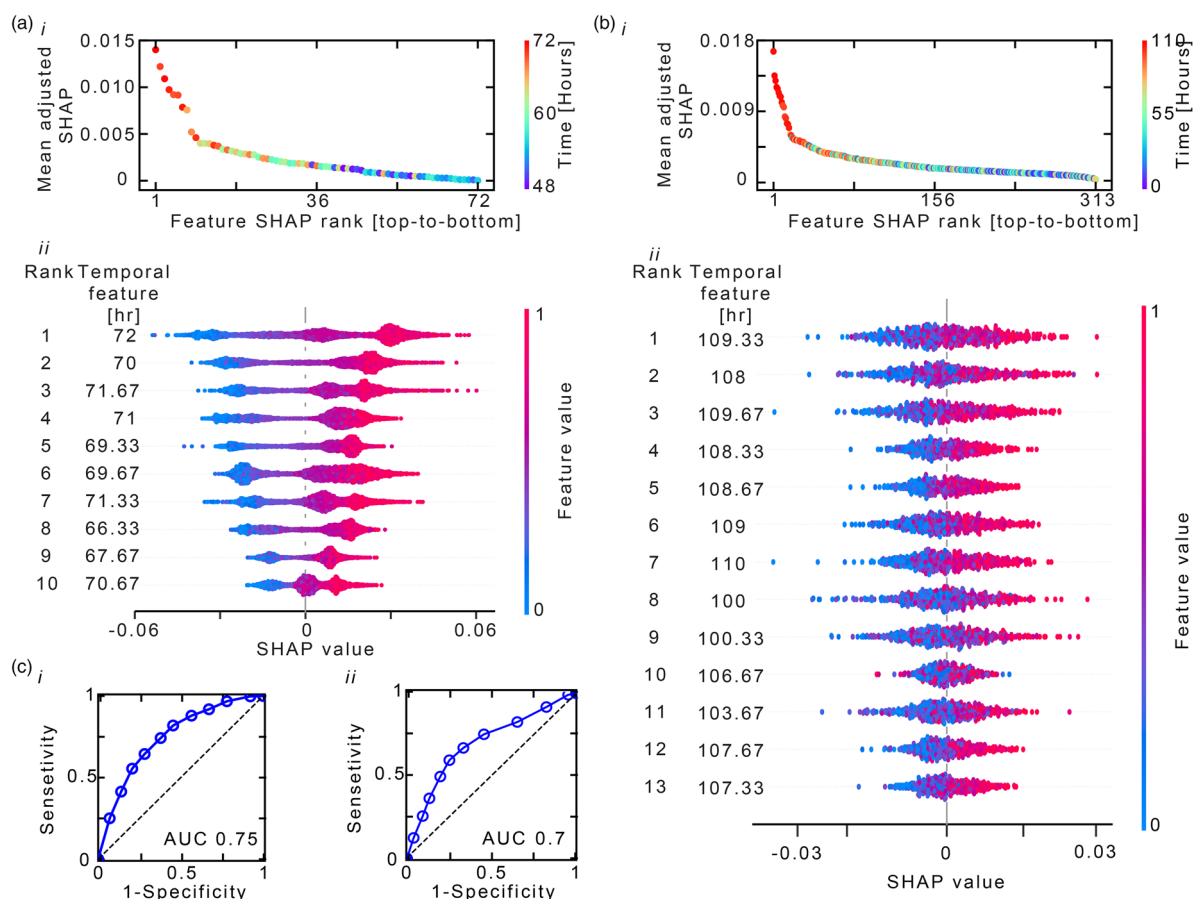


**Figure 3.** SHIFRA<sub>B</sub> and SHIFRA<sub>K</sub> are optimized for prediction of embryo blastulation and implantation. a) Scores of KID prediction (110 h) and BLAST prediction (72 h) of BLAST<sub>p</sub>-KID<sub>n</sub> (B<sub>p</sub>K<sub>n</sub>) and BLAST<sub>p</sub>-KID<sub>p</sub> (B<sub>p</sub>K<sub>p</sub>) colabeled embryos ( $n = 396$ ) are weakly positively correlated (Spearman's correlation 0.3). Consistent with their blastulation labels, BLAST prediction histograms of B<sub>p</sub>K<sub>n</sub> and B<sub>p</sub>K<sub>p</sub> embryos overlap (top panel). b) The average blastulation scores of B<sub>p</sub>K<sub>n</sub> and B<sub>p</sub>K<sub>p</sub> embryos are comparable (left). However, B<sub>p</sub>K<sub>p</sub> embryos are scored 40% higher than B<sub>p</sub>K<sub>n</sub> embryos only by SHIFRA<sub>K</sub> consistent with their implantation labels (right).

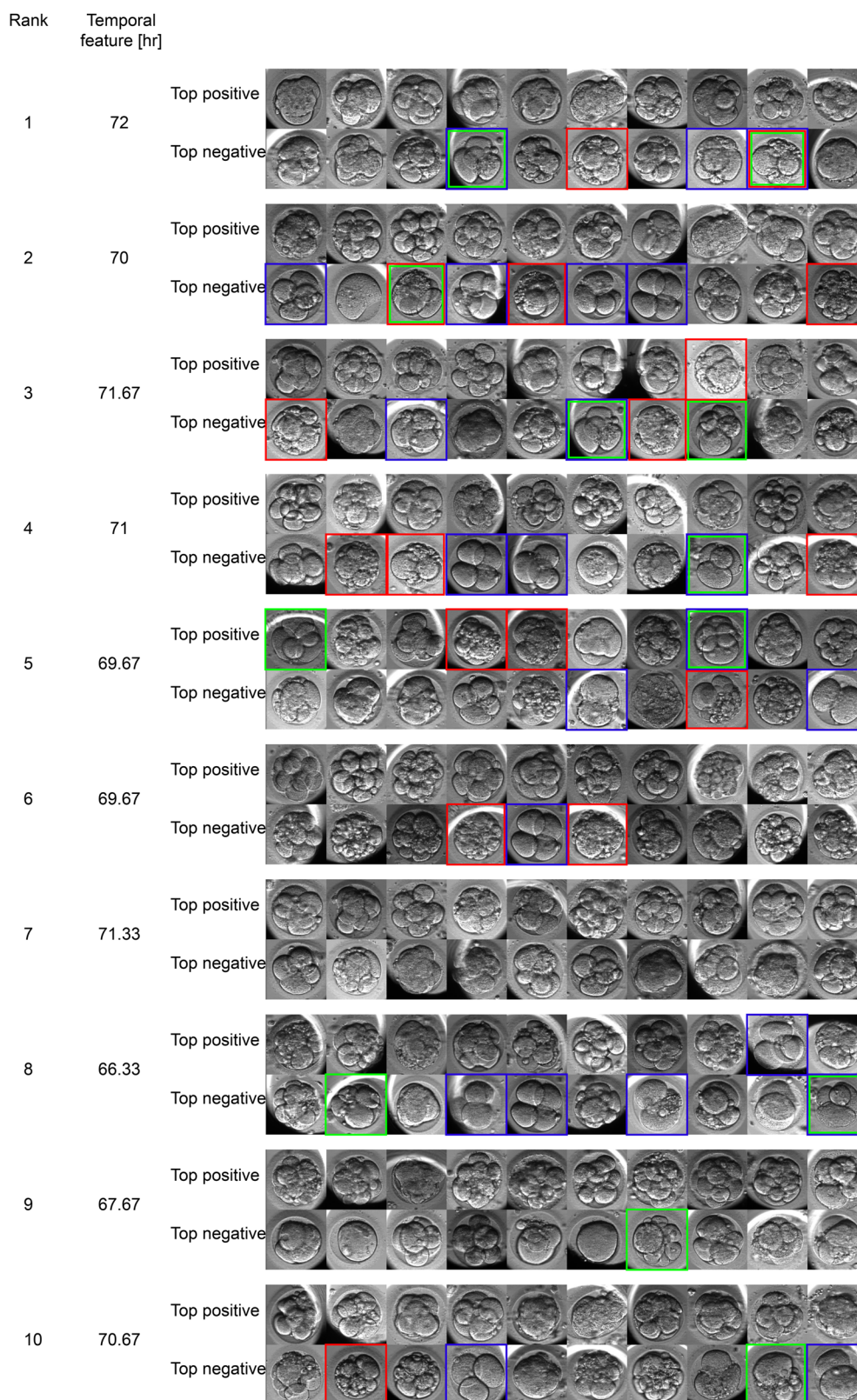
(Figure 4a-i) and to implantation prediction (Figure 4b-i) changes greatly between different frames. Temporal features of low mean adjusted SHAP scores were associated with early time points. The top ten SHIFRA<sub>B</sub> features (adjusted SHAP > 0.004) and top 13 SHIFRA<sub>K</sub> features (adjusted SHAP > 0.007) were associated with latest frames. The SHAP values and the feature values of the top-ranked SHIFRA<sub>B</sub> and SHIFRA<sub>K</sub> temporal features were correlated across individual embryos (Figure 4a-ii,b-ii). To study whether the top-ranked temporal features can direct blastulation and implantation prediction without including the rest of the temporal features, we trained again the embryo-learning blastulation and implantation classifiers using only the top-ranked features. Indeed, we found that the top-ranked features were sufficient for reaching high predictive power with comparable AUC values (BLAST: AUC = 0.75; KID: 0.7 Figure 4c-i,ii) as SHIFRA<sub>B</sub> (AUC = 0.74, Figure 1c-inset) and SHIFRA<sub>K</sub> (AUC = 0.71, Figure 2b-inset). This indicates that the top-ranked temporal features as defined here mark the developmental potential of the embryos to blastulate and to implant.

## 2.6. SHIFRA<sub>B</sub> & SHIFRA<sub>K</sub> Are Sensitive to Dynamic Features Beyond Morphological/Morphokinetic States

The ambiguity in identifying the actual visual elements that direct neural network prediction is one of the major drawbacks in deep learning. To obtain a mechanistic insight into how embryo prediction by SHIFRA<sub>B</sub> and SHIFRA<sub>K</sub> work, we present the top-ten positive SHAP images and top-ten negative SHAP images for each of the top-ranked SHIFRA<sub>B</sub> and SHIFRA<sub>K</sub> temporal features (Figure 5 and 6). Herein, we mark the cleavage-stage embryos with four cells or less (blue frames), embryos with asymmetric blastomeres (green frames), and highly fragmented embryos (red frames) in the images that contributed the most to positive and negative blastulation prediction by SHIFRA<sub>B</sub> (Figure 5). Temporal features 1 (72 h), 2 (70 h) and 4 (71 h) were most sensitive to four-cell cleavage stage embryos and were identified as SHAP-negative images. Temporal features 1, 3 (71.7 h) and 8 (66.3 h) were most sensitive to uneven blastomere size and temporal features 2 and 4 were most sensitive to embryo fragmentation, which also obtained negative-SHAP scores

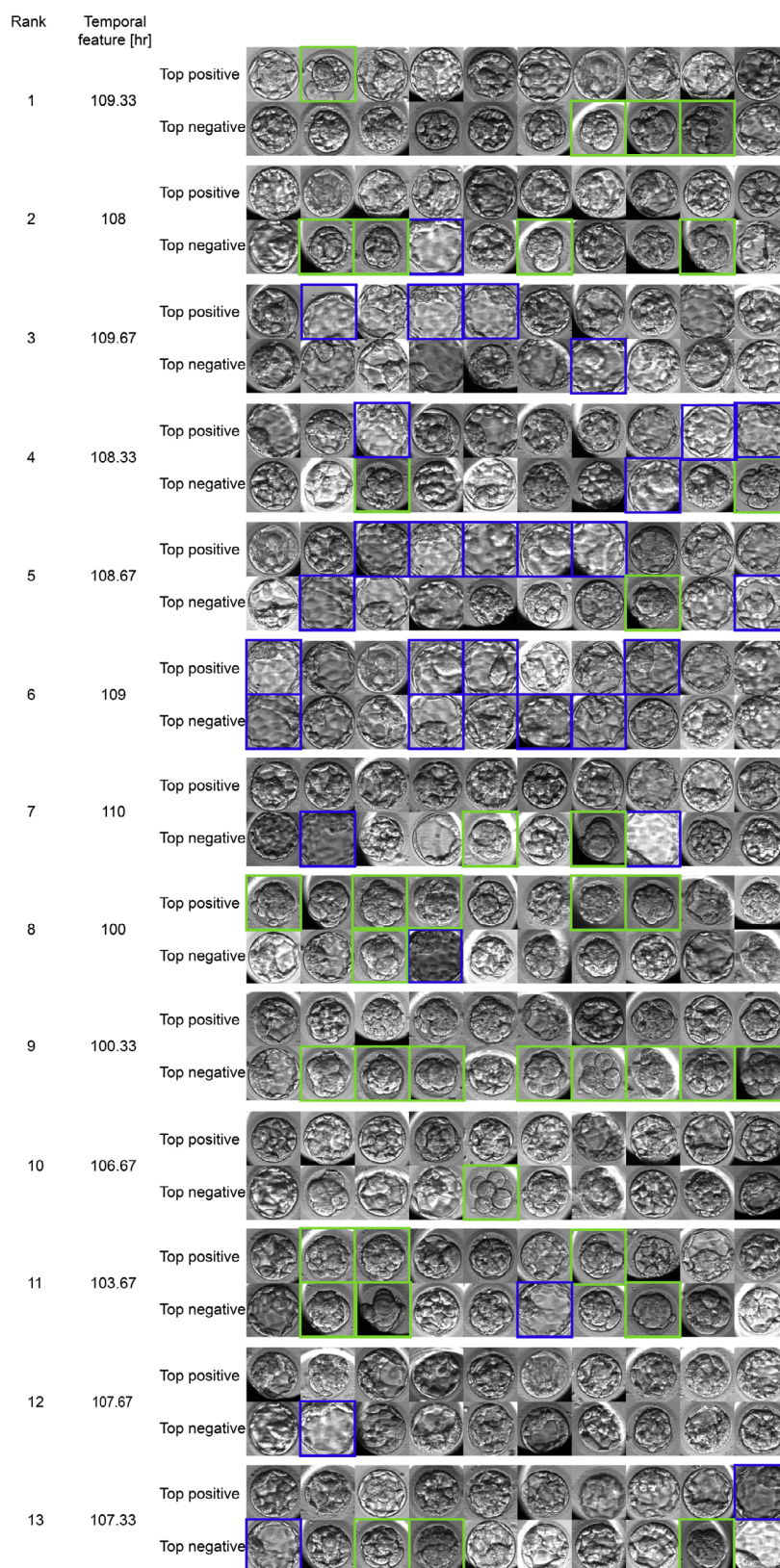


**Figure 4.** SHIFRA<sub>B</sub> and SHIFRA<sub>K</sub> are optimized for prediction of embryo blastulation and implantation. a,b) The temporal features of a-i) SHIFRA<sub>B</sub> (72 h) and b-i) SHIFRA<sub>K</sub> (110 h) are ranked according to their mean adjusted SHAP, which scores feature contribution to accurate BLAST and KID prediction. a-ii) The ten top-ranked BLAST features (>0.004) and b-ii) the 13 top-ranked KID features (>0.007) were derived from the latest time-lapse images (BLAST > 66 h and KID > 100 h; color coded). SHAP values and feature values of the top-ranked temporal features were calculated for a-ii) BLAST-labeled and (b-ii) KID-labeled train set embryos. c) ROC curves and AUC obtained by i) BLAST and ii) KID classifiers that were trained only on the top ten BLAST features and the top 13 KID features, respectively.



**Figure 5.** Top positive versus negative SHAP-scored embryo frames directing BLAST prediction. Top ten SHAP-positive versus top ten SHAP-negative embryo frames are shown for the selected temporal features by mean adjusted SHAP for SHIFRA<sub>B</sub> BLAST prediction at 72 h. Cleavage stage embryos with  $\leq 4$  cells (blue frames), blastomere asymmetry (green frames), and high fragmentation embryos (red frames) are marked.





**Figure 6.** Top positive versus negative SHAP-scored embryo frames directing KID prediction. Top ten SHAP-positive versus top ten SHAP-negative embryo frames are shown for the selected temporal features by mean adjusted SHAP for SHIFRA<sub>K</sub> KID prediction at 110 h. Cleavage stage or morula stage embryos exposing noncompacted blastomeres (blue frames) and morula stage embryos (green frames) are marked.



(Figure 5). Similarly, we marked the cleavage and morula stage embryos with noncompacted blastomeres (blue frames) and compacted morulae (green frames) in the images that contributed the most to positive and negative KID prediction by SHIFRA<sub>K</sub> (Figure 6). We found that temporal features 10 (106.7 h) and 13 (107.3 h) were most sensitive to the appearance of noncompacted blastomeres and temporal features 3 (109.7 h), 9 (100.3 h), and 11 (103.7 h) were most sensitive to morula. In both cases, these morphological characteristics directed negative-SHAP KID prediction (Figure 6). The morphokientic and morphological characteristics that we discuss herein are depicted only by a small fraction of the images of the temporal features, indicating that the determinant visual elements that direct blastulation and implantation prediction are not distinguished by human-level perception.

### 3. Discussion

The advantage of using retrospective over prospective embryo transfer datasets for DNN training is their ethical feasibility and far greater size. However, prediction accuracy is limited due to lacking critical information about endometrial receptivity and using a homogenous dataset of embryos that were retrospectively preselected for transfer according to established morphological and/or morphokientic criteria. To overcome these limitations, we first trained a DNN on blastulation outcome using a heterogeneous set of labeled embryos and integrated it with three additional DNNs that were trained separately on implantation outcome. Indeed, accuracy of day-3 implantation prediction was equally low by SHIFRA<sub>K</sub> and KIDScore-D3, but it was significantly higher on day-5 by SHIFRA<sub>K</sub>, proving to be more accurate than implantation prediction by KIDScore-D5 (Figure 2b).

Unlike the current state-of-the-art embryo classification algorithms,<sup>[8,10]</sup> deep learning provides full automation and standardization of embryo classification, which are both important for clinical adoption.<sup>[25,26]</sup> The SHIFRA classifiers, which provide early evaluation of blastulation and implantation potential, are the first step toward the development of a decision-making tool that will provide a personalized, multistep, embryo transfer strategy.<sup>[27,28]</sup> Namely, given a finite number of embryos obtained from a patient and their assessed quality, this tool will specify the multistep order and timing of embryo transfers (including transfers of multiple embryos), as well as which embryos are to be cryopreserved for subsequent transfers. The general framework presented herein, together with larger datasets of embryo data, open the door for the implementation of such personalized clinical tools that will optimize conception rates while shortening time to pregnancy in IVF treatments.

### 4. Experimental Section

**The SHIFRA Database, Embryo Annotation, and QA:** We developed a PostgreSQL database with a front-end website that supported display, query, and data annotation (Hebrew University IT). Maintenance of the SHIFRA database was outsourced (CHELEM LTD), and data curation was conducted by a full-time trained embryologist. Anonymized time-lapse video files and the corresponding metadata were imported from five hospitals (Table S1, Supporting Information). Data were imported under the approval of the Helsinki ethical committee in each hospital.

Only embryos that were fertilized via intracytoplasmic sperm injection (ICSI) and showed two-pronuclei appearance (2PNa) inside the incubator were included. In this manner, we accurately defined time of fertilization, discarded nonfertilized embryos, and obtained a full morphokientic profile starting from tPNa. Preimplantation genetic screening/diagnosis (PGS/PGD) tested embryos were discarded as well. Morphokientic annotations were performed by qualified and experienced embryologists in each IVF clinic adhering to established protocols.<sup>[2,21]</sup> QA was carried out by comparing the morphokientic annotations of 253 randomly selected embryos with the annotations of an expert embryologist in a blinded manner (Dr. IHV, Soroka Medical Center, Figure S2a, Supporting Information).

**Labeling of Embryo Blastulation:** Identification of arrested embryos that cannot advance toward blastulation was based on their morphokientic profiles (Figure S2b-i,ii, Supporting Information) and the time windows for advancing from one morphokientic event to the next (Table S2 and S3, Supporting Information).

**Morphokientic Events:** Each embryo was represented by its total time of incubation measured from fertilization,  $t_{inc}$ , and its latest developmental state reached inside the incubator,  $S_n$  (Figure S3a-i, Supporting Information). Embryos, which have the capacity to advance to the next developmental state if incubation was extended, are located within the orange regions, which were bound between the 1st percentile of the temporal distribution of the morphokientic event  $S_n$  and the 99th percentile of the temporal distribution of the consecutive morphokientic event  $S_{n+1}$  (Table S2, Supporting Information). Embryos that were located in the red regions, which were bottom bounded by the 99th percentile of the consecutive morphokientic event  $S_{n+1}$ , missed the permitted time window for advancing toward the next developmental state. For example, an embryo that was incubated for 96 h and reached four-cell state, was arrested and was labeled 4-cells positive 5-cells negative (4Cp5Cn).

**Morphokientic Intervals:** A similar statistical analysis was performed to identify developmentally arrested embryos based on the temporal distributions of the morphokientic intervals. Each embryo was represented by the time that lapsed between time of last morphokientic event and time of incubation,  $t_{inc} - t_n$ , which was plotted versus the latest developmental state reached inside the incubator,  $S_n$  (Figure S3a-ii, Supporting Information). The orange regions were bound between the 1st percentile and the 99th percentile (Table S3, Supporting Information). Embryos that were located in these regions reached developmental state  $S_n$  and still hold the potential to advance to morphokientic stage  $S_{n+1}$  if incubation was extended. In contrast, embryos that were located in the red regions missed their interval time window and were thus arrested in developmental state  $S_n$ . For example, an embryo that had completed four-cell cleavage and that 36 h had lapsed as without advancing to the next cleavage event was arrested.

**BLAST-Labeling:** Adhering to a constraining methodology, embryos that were found to be developmentally arrested either based on their morphokientic profiles or based on their interval profiles are classified BLAST<sub>n</sub> (Figure S3c-ii, Supporting Information). Embryos that reached start-of-blastulation (SB) inside the incubator were labeled BLAST<sub>p</sub> and were located in the green regions (Figure S3c-i, Supporting Information). Obtaining BLAST<sub>p</sub> embryos required lengthy incubations, which was facilitated by clinics that transfer embryos on day-5 (H3; Figure S3d and S1b, Supporting Information). The remaining embryos that were located within the orange regions were classified BLAST-unknown (BLAST<sub>u</sub>, Figure S3c-i, Supporting Information).

**Labeling of Embryo Implantation and Clinical Pregnancy:** Implantation of transferred embryos in the uterus was labeled by KID. KID status was determined based on established protocols by comparing the number of transferred embryos with the number of implanted embryos as determined by the measured number of gestational sacs on week five of pregnancy. In the case that the number of transferred and implanted embryos was equal, the embryos were labeled KID-positive (KID<sub>p</sub>). KID-negative (KID<sub>n</sub>) corresponds to the case of no implanted embryos. KID-unknown (KID<sub>u</sub>) marks embryos whose implantation outcome cannot be determined, for example, when three embryos were transferred and only one or two were implanted. Clinical pregnancy (CP) accounts for the implantation of a viable embryo as determined by fetal heart beat

ultrasound measurement on week 7 of pregnancy. Positive CP (CP<sub>p</sub>) accounts for the case that the number of transferred embryos, gestational sacs, and fetal heart beats were the same. Negative CP (CP<sub>n</sub>) includes transferred embryos that failed to implant (KID<sub>n</sub>) and KID<sub>p</sub> embryos with no fetal heartbeat.

**KID Classification by KIDScore-D3 and KIDScore-D5 Decision Support Tools:** Prediction of embryo implantation was performed by KIDScore-D3 on day-3 (66 h prediction time) and by KIDScore-D5 on day-5 (110 h prediction time) according to the manufacturer's protocols (Vitrolife). Specificity and sensitivity were evaluated as a function of KIDScore values.

**Statistical Analysis and Software:** Statistical analyses were performed using NumPy, Scikit-Learn, and SciPy Python packages (Python Software Foundation) and R (The R Foundation). Graphs were generated using Matplotlib and Seaborn Python packages (Python Software Foundation).

**Preprocessing of Embryo Video Files:** The convergence of supervised neural networks depended on the balance between the number of labeled objects and their total size. We performed preprocessing of the input video ( $\approx 100$  MB) and decreased their size 16-fold while capturing the dynamic nature of preimplantation embryo development. Preprocessing included 4 $\times$ -resizing of the embryo images (2 $\times$  biaxially) and segmentation of the embryo region of interest. Later, we described the steps for performing automated image preprocessing using a U-Net classifier.<sup>[29]</sup>

**Semiautomated Interactive Labeling of Embryo Segmentation:** The first goal was to generate a large set of validated embryo images labeled and by binary masks of the embryo region of interest that will be used for training a fully automated classifier. Using the semiautomated interactive GrabCut algorithm,<sup>[30]</sup> the embryos were framed manually by a bounding box, allowing for an initial segmentation of the embryo (Figure S4a, Supporting Information). Next, embryo segmentation was adjusted interactively by scribbling the regions to include within the embryo region of interest (ROI) and regions to exclude from the embryo ROI, allowing fine tuning of embryo segmentation. Multiple segmentation—scribbling—fine-tuning iterations were allowed. Once embryo segmentation was approved, morphological operations were executed (filling holes, contour smoothing, and dilation), and a binary mask was defined. In total, we generated binary masks for 2700 images spanning all relevant developmental states of the embryos.

**Automated Embryo Segmentation Using An Automated U-Net Pixel Classifier:** A U-Net classifier was trained using the images that were labeled interactively via GrabCut. Segmented images were resized to 256  $\times$  256 pixels and divided into train set (2350 images), validation set (200 images), and uncontaminated test set (150 images). Training was performed using randomly selected 100 images batch size at 1500 steps per Epoch. Each batch of images underwent random augmentation that included 0°–180° rotations, horizontal flips, 0–0.1 shearing, and 0–0.1 zooming. In this manner, training on exactly the same images on different steps was prevented. Network convergence was reached within 20 epochs. The U-Net classifier obtained embryo images as an input and generated a binary mask output of the embryo ROI resized to 500  $\times$  500 pixels (Figure 2b-i). Accuracy of embryo segmentation was evaluated on test-set embryos using the Dice Coefficient, which measures the overlap between the segmented pixels and the binary mask labels. Dice coefficient ranged between 0.91 and 0.94 across all developmental states (Figure S4b-ii, Supporting Information). Embryo ROIs were evaluated for all images in the database. Empty well images were identified based on small mask area and discarded.

**Grouping of Five-Frame Packets:** All segmented embryo frames were resized 2 $\times$  along each axis into a 128  $\times$  128 pixel images. Five consecutive resized and segmented frames of the same focal plane and the same embryo were grouped together into packets,  $P_n^m$ , of embryo  $m$  and first frame time index  $n$  (Figure S5a, Supporting Information). Each packet was associated with a blastulation and/or implantation label  $y_m$ . These packets served as the input objects of the packet learning neural networks (see in the following sections).

**Packet Learning—Neural Network Design and Training Parameters:** The network was implemented using the PyTorch framework,<sup>[31]</sup> and trained using Stochastic Gradient Descent with Nesterov of 0.9. The input objects of the network were packets of five preprocessed frames  $P_n^m$

(Figure S5a, Supporting Information). Training batches included  $k = 4$  packets obtained from  $K = 8$  embryos within 12 h windows. Training sets were balanced by selecting positive- and negative-labeled embryos (both blastulation and implantation) opposite of their respective frequencies. Packets of the three central focal planes ( $-15$ ,  $0$ , and  $+15$   $\mu\text{m}$ ) were used for training, whereas validation and test-set embryos contributed packets only of the middle focal plane. We used a residual network architecture<sup>[32]</sup> consisting of 13 layers that included seven residual blocks and two fully connected layers (Figure S5b, Supporting Information). The last layer consists of  $w$  input neurons (see below) that were associated with nonoverlapping time windows as defined for blastulation and implantation prediction networks (see in the following paragraphs). Packet score output was determined by selection of one of the input neurons according to the time index  $n$  of the first frame of the packet (Figure S5b, Supporting Information).

**Packet Learning—Weighted Logistic Loss:** We postulate that embryo developmental potential was marked by scarce dynamic events that last 30–60 min and were thus captured by individual packets. Hence, a high-quality embryo will have only a few packets that were scored high, whereas all packets of a low-quality embryo will be scored low. This principle was implemented by weighing logistic loss as follows.<sup>[33]</sup> The weighted loss of embryo  $m$  is calculated based on all  $k$  packets

$$I^m = \sum_{n=1}^k w_n^m \log(1 + e^{-\gamma_m s_n^m}) \quad (1)$$

where  $s_n^m$  are the packet scores that are further used for calculating the softmax weights  $w_n^m$

$$w_n^m = \frac{e^{\gamma s_n^m}}{\sum_{n=1}^k e^{\gamma s_n^m}} \quad (2)$$

Note that the sum of weights  $w_n^m$  across  $k$  packets of embryo  $m$  is 1.  $\gamma$  is the softness parameter. At the limit  $\gamma \rightarrow 0$ , the weight becomes  $\frac{1}{k}$  independent of the scores of the packets. At the opposite limit of large  $\gamma$ ,  $w_n^m$  approaches 1 only for the packet of maximal score. The problem of approaching this limit was that it will be increasingly difficult for the network to converge. The batch loss  $L$  is

$$L = \frac{1}{K} \sum_{m=1}^K I^m \quad (3)$$

The DNN weights were thus optimized to minimize  $L$  over all  $K$  embryos in the batch. Performances were optimized by setting  $\gamma = 3$ . For a negative-labeled embryo ( $y_i = -1$ ), even if a single packet will obtain a positive score, its weight  $w_n^m$  will be highest and the loss of the embryo  $I^m$  will be large. As a result, convergence will be approached only if all packets of a negative embryo will obtain negative scores. In contrast, one packet with a high-positive score was sufficient for obtaining a small loss for a positive embryo ( $y_i = +1$ ). For these embryos, the packets with low scores will have small weights and the packet with the highest score will have the highest weight and a small loss will be obtained.

**Packet Learning of Blastulation- and Implantation-Labeled Packets:** Packet learning for blastulation prediction and for implantation prediction was performed using the same DNNs as described earlier (Figure S5b, Supporting Information). Network training for blastulation prediction was performed using  $w = 25$  nonoverlapping time windows (0–4; 4–8; 8–12; 12–16; 16–20; 20–24; 24–27; 27–30; 30–33; 33–36; 36–42; 42–46; 46–48; 48–51; 51–56; 56–64; 64–72; 72–76; 76–80; 80–85; 85–90; 90–95; 95–105; 105–115; >115). Network training for KID prediction was performed using  $w = 16$  nonoverlapping time windows (0–12; 12–24; 24–30; 30–36; 36–42; 42–48; 48–56; 56–64; 64–72; 72–76; 76–80; 80–85; 85–95; 95–115; 115–120; >120). DNN training using BLAST-labeled and KID-labeled embryos typically converged within 20-to-60 epochs.

Unlike blastulation, implantation outcome of transferred embryos depended not only on their developmental competence, but also on endometrial receptivity, which was not considered in the learning process. Therefore, packet-learning for KID-prediction was performed

using an ensemble of three DNNs that were trained on KID-labeled packets and one DNN that was trained on BLAST-labeled packets (four networks in total). The three KID-networks were trained using different hyperparameters (initial weights and learning rates). In this manner, blastulation prediction packet learning generates one frame score, whereas KID prediction packet learning generates four scores for the first image of each packet that are summed into one final frame score.

**Embryo Learning:** Scoring the developmental potential of embryos was performed by training a second classifier. Each embryo was represented by a vector of frame scores obtained by packet learning as described earlier. Next, temporal features were generated by interpolation of the vectors of frame scores, thus obtaining a synchronized representation of all embryos. To allow embryo prediction at different time points (time of prediction,  $t_p$ ), different classifiers were trained independently. Given  $t_p$ , a classifier was trained on all train-set labeled embryos of video length greater than  $t_p$  using the temporal features earlier than  $t_p$ . BLAST prediction was performed using a Random Forest classifier, whereas KID prediction was performed using logistic regression. This choice was made based on the performance of multiple classifiers that were tested for BLAST and KID prediction. In both cases, training parameters were optimized via grid-search fivefold cross validation.

## Data and Materials Availability

The copyrights of the code are owned by Yisum—the technology transfer company of The Hebrew University of Jerusalem. Requests can be sent to AB. The clinical data are owned by Hadassah Medical Center and by Clalit Health Services. Restrictions apply to the availability of these data, which were used anonymously under ethical agreements with each clinic separately for this study, and so are not made publically available. Access requests can be directed to:

Dr. Assaf Ben-Meir, Fertility and IVF Unit, Department of Obstetrics and Gynecology, Hebrew University Hadassah Medical Center, Jerusalem Israel.

Dr. Yuval Or, Fertility and IVF Unit, Kaplan Medical Center, Rehovot, Israel, Hebrew University, Jerusalem, Israel.

Dr. Iris Har-Vardi, Fertility and IVF Unit, Department of Obstetrics and Gynecology, Soroka University Medical Center, Ben-Gurion University of the Negev, Beer Sheva, Israel.

Dr. Yoel Shufaro, Women's Hospital, Rabin Medical Center, Petah Tikva, Israel.

## Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

## Acknowledgements

The authors greatly appreciate support from the European Research Council (ERC-StG 678977), Israel Innovation Authority (KAMIN grant no. 55326), Israel Science Foundation (grant no. 1246/14), and Niedersächsisches Vorab (grant no. ZN 3183). The authors thank Y. Hoshen and Y. Kochman (Hebrew University of Jerusalem) for fruitful discussions.

## Conflict of Interest

The authors declare no conflict of interest.

## Author Contributions

A.B.M. and A.B. formulated and designed research. Y.K.T. assembled the SHIFRA database. Y.K.T. and N.Z. trained the classifiers. A.Z., Y.K.T., and

T.A. performed data preprocessing. D.R. and I.H.V. managed morphokinetic annotation and embryo labeling. Y.O., Z.S., A.H., and I.H.V. provided clinical data. M.G. provided statistical support. A.B.M. and A.B. supervised research. Y.K.T. and A.B. designed and prepared the graphical material. A.B. wrote the text.

## Keywords

automated embryo classification, deep learning, embryo transfers, in vitro fertilization

Received: April 18, 2020

Revised: May 18, 2020

Published online:

- [1] American College of Obstetricians and Gynecologists Committee on Practice Bulletins-Obstetrics, Society for Maternal-Fetal Medicine, ACOG Joint Editorial Committee, *Obstet. Gynecol.* **2004**, *104*, 869.
- [2] Alpha Scientists in Reproductive Medicine, Embryology ESIgo, *Hum. Reprod.* **2011**, *26*, 1270.
- [3] L. R. Goodman, J. Goldberg, T. Falcone, C. Austin, N. Desai, *Fertil. Steril.* **2016**, *105*, 275.
- [4] X. Wang, C. Chen, L. Wang, D. Chen, W. Guang, J. French, *Fertil. Steril.* **2003**, *79*, 577.
- [5] D. J. Kaser, C. Racowsky, *Hum. Reprod. Update* **2014**, *20*, 617.
- [6] F. Guerif, A. Le Gouge, B. Giraudeau, J. Poindron, R. Bidault, O. Gasnier, D. Royere, *Hum. Reprod.* **2007**, *22*, 1973.
- [7] Y. Motato, M. J. de los Santos, M. J. Escriba, B. A. Ruiz, J. Remohí, M. Meseguer, *Fertil. Steril.* **2016**, *105*, 376.
- [8] R. Milewski, P. Kuć, A. Kuczyńska, B. Stankiewicz, K. Łukaszuk, W. Kuczyński, *J. Assist. Reprod. Genet.* **2015**, *32*, 571.
- [9] M. Meseguer, M. Meseguer, J. Herrero, A. Tejera, K. M. Hilligsøe, N. B. Ramsing, J. Remohí, *Hum. Reprod.* **2011**, *26*, 2658.
- [10] B. M. Petersen, M. Boel, M. Montag, D. K. Gardner, *Hum. Reprod.* **2016**, *31*, 2231.
- [11] C. Blank, R. R. Wildeboer, I. DeCroo, K. Tillemann, B. Weyers, P. De Sutter, M. Mischi, B. C. Schoot, *Fertil. Steril.* **2019**, *111*, 318.
- [12] K. H. Yu, A. L. Beam, I. S. Kohane, *Nat. Biomed. Eng.* **2018**, *2*, 719.
- [13] E. J. Topol, *Nat. Med.* **2019**, *25*, 44.
- [14] A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun, J. Dean, *Nat. Med.* **2019**, *25*, 24.
- [15] M. Simopoulou, K. Sfakianoudis, E. Maziotis, N. Antoniou, A. Rapani, G. Anifandis, P. Bakas, S. Bolaris, A. Pantou, K. Pantos, M. Koutsilieris, *J. Assist. Reprod. Genet.* **2018**, *35*, 1545.
- [16] D. Dirvanauskas, R. Maskeliunas, V. Raudonis, R. Damasevicius, *Comput. Methods Programs Biomed.* **2019**, *177*, 161.
- [17] P. Khosravi, E. Kazemi, Q. Zhan, J. E. Malmsten, M. Toschi, P. Zisimopoulos, A. Sigaras, S. Lavery, L. A. Cooper, C. Hickman, M. Meseguer, *npj Digital Med.* **2019**, *2*, 1.
- [18] D. Tran, S. Cooke, P. J. Illingworth, D. K. Gardner, *Hum. Reprod.* **2019**, *34*, 1011.
- [19] M., VerMilyea, J. M., Hall, S. M., Diakiw, A., Johnston, T., Nguyen, D., Perugini, A., Miller, A., Picou, A. P., Murphy, M., Perugini, *Hum. Reprod.* **2020**, *35*, 770.
- [20] N. Lewin-Epstein, Y. Cohen, *J. Ethn. Migr. Stud.* **2019**, *45*, 2118.
- [21] H. N. Ciray, A. Campbell, I. E. Agerholm, J. Aguilar, S. Chamayou, M. Esbert, S. Sayed, *Hum. Reprod.* **2014**, *29*, 2650.
- [22] S. Fernandez-Shaw, R. Cercas, C. Brana, C. Villas, I. Pons, *J. Assist. Reprod. Genet.* **2015**, *32*, 177.

- [23] Q. Wu, H. Li, Y. Zhu, W. Jiang, J. Lu, D. Wei, J. Yan, Z. J. Chen, *Hum. Reprod.* **2018**, 33, 1875.
- [24] S. Lundberg, S. I. Lee, in *Advances in Neural Information Processing Systems*, **2017**, pp. 4765–4774.
- [25] S. Harrer, P. Shah, B. Antony, J. Hu, *Trends Pharmacol. Sci.* **2019**, 40, 577.
- [26] K. Y. Ngiam, I. W. Khor, *Lancet Oncol.* **2019**, 20, e262.
- [27] Z. Pandian, J. Marjoribanks, O. Ozturk, G. Serour, S. Bhattacharya, *Cochrane Database Syst. Rev.* **2013**, CD003416.
- [28] F. M. Ubaldi, A. Capalbo, S. Colamaria, S. Ferrero, R. Maggiulli, G. Vajta, F. Sapienza, D. Cimadomo, M. Giuliani, E. Gravotta, A. Vaiarelli, *Hum. Reprod.* **2015**, 30, 2097.
- [29] O. F. Ronneberger, P. Fischer, T. Brox, in *International Conference on Medical image computing and computer-assisted intervention*, Springer, Cham **2015**, pp. 234–241.
- [30] C. Rother, V. Kolmogorov, B. Blake, *ACM Trans. Graph.* **2004**, 23, 309.
- [31] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimeshein, L. Antiga, A. Desmaison, *Adv. Neural Inf. Process. Syst.* **2019**, 8024.
- [32] K. M. He, X. Zhang, S. Ren, J. Sun, in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition* **2016**, pp. 770–778.
- [33] S. Ben-David, S. Shalev-Shwartz, *Understanding Machine Learning: From Theory to Algorithms*, Cambridge University Press, Cambridge **2014**.