

AN IMAGE PROCESSING PROTOCOL TO EXTRACT VARIABLES PREDICTIVE OF HUMAN EMBRYO FITNESS FOR ASSISTED REPRODUCTION

RUNNING TITLE

Image processing for human blastocyst selection

Dóris Spinosa Chéles^{1,2,5,§}, André Satoshi Ferreira^{1,§}, Isabela Sueitt de Jesus¹, Eleonora Inácio Fernandez¹, Gabriel Martins Pinheiro¹, Eloiza Adriane Dal Molin¹, Wallace Alves¹, Rebeca Colauto Milanezi de Souza^{1,2}, Lorena Bori³, Marcos Mesequer^{3,4}, José Celso Rocha^{1,2,§}, Marcelo Fábio Gouveia Nogueira^{2,5,§,*}

¹Laboratory of Applied Mathematics, Department of Biological Sciences, São Paulo State University (UNESP), Assis, Brazil.

²Graduate Program in Pharmacology and Biotechnology, Institute of Biosciences, São Paulo State University (UNESP), Botucatu, Brazil.

³IVF Laboratory, IVI Valencia, Valencia, Spain.

⁴Health Research Institute la Fe, Valencia, Spain.

⁵Laboratory of Embryonic Micromanipulation, Department of Biological Sciences, São Paulo State University (UNESP), Assis, Brazil.

§These authors contributed equally: D.S.C., A.S.F., J.C.R., and M.F.G.N.

*Corresponding author (marcelo.fabio@unesp.br)

CAPSULE

The image processing protocol can segment human blastocyst images and automatically extract 33 variables that describe quantitative aspects of the blastocyst's regions, with potential utility in embryo selection for ART.

ABSTRACT

OBJECTIVE

To study the application of image processing for segmentation of blastocysts images and extraction of potential variables for prediction of embryo fitness.

DESIGN

Retrospective study.

SETTING

Single reproductive medical center. IVI-RMA (Valencia, Spain) between 2017 and 2019.

PATIENTS

An initial dataset including 353 images from EmbryoScope[®] and 474 images from Geri[®] incubators was acquired, of which 320 images from EmbryoScope[®] and 309 images from Geri[®] incubators were used in this study.

INTERVENTION(S)

None.

MAIN OUTCOME MEASURE(S)

Successful segmentation of images into trophectoderm (TE), blastocoel, and inner cell mass (ICM) using the proposed processing steps.

RESULTS

A total of 33 variables were automatically generated by digital image processing, each representing a different aspect of the embryo and describing a different characteristic of the expanding blastocyst (EX), ICM, or TE. These variables can be categorized into texture, gray level average, gray level standard deviation, modal value, relations, and light level. The automated and directed steps of the proposed processing protocol exclude spurious results, except when image quality (e.g., focus) prevents correct segmentation.

CONCLUSIONS

The proposed image processing protocol that can successfully segment human blastocyst images from two distinct sources and extract 33 variables with potential utility in embryo selection for ART.

KEY WORDS

digital image processing, blastocyst, morphology-derived variables, segmentation, embryo selection.

INTRODUCTION

Digital image processing and artificial intelligence (AI) techniques are rapidly gaining acceptance in medical sciences, including reproductive medicine (1, 2), as tools for objective identification of lesions on medical images and even for disease diagnosis and prognosis. In the field of assisted reproduction technology (ART), these techniques are currently under investigation for identifying embryological parameters predictive of fitness for freezing and implantation (3). Further, automated image processing is now being used to improve the standardization and accuracy of diagnostics and prediction (4).

In the field of ART, computer-aided analyses are especially suitable for evaluation and classification of gametes and embryos (5), and when combined with other methods, to predict embryo fate for ART (6). Considering the complexity of each step and the vast number of variables to be considered, ART is an ideal field to

test the potential of AI techniques (7). Indeed, several algorithms have been developed to classify embryos or predict reproductive success that applies classical statistical methods such as logistic regression or various AI techniques (8).

The raw data for embryo evaluation is often acquired using a time-lapse system (TLS), which yields dynamic information on embryo morphology (9). Digital image processing of these TLS-acquired image series has improved the accuracy of human embryo evaluation (10, 11). For instance, selection based on TLS has resulted in a 20% increase in pregnancy rate using ART (12). However, embryo evaluation is still largely subjective and based on the embryologist's experience, resulting in inconsistencies that contribute to the variability in clinical outcome (13, 14). Therefore, automation of embryo evaluation, although challenging due to the complexity of embryological variants (15), may lead to more consistent ART success.

Digital image processing as applied to ART includes multiple computational techniques that can extract mathematical variables from either conventional or TLS-acquired embryo images. Further, AI techniques can automate this parameter selection process to obtain objective metrics associated with embryo morphology and quality (16, 17). Currently, the most widely used embryo classification system in Brazil and Argentina is that proposed by Gardner and Schoolcraft (18) and recently detailed by Puga-Torres *et al.* (19), which evaluates embryo fitness according to expansion of the blastocyst (EX), quality of the inner cell mass (ICM), and quality of the trophectoderm (TE) (20). Image processing can evaluate embryonic quality by combining metrics for each region of the blastocyst (including the zona pellucida and the segmented ICM and TE) as described for bovine embryos (21). In conjunction with AI tools, such image processing methods may help embryologists select the best oocyte or embryo for freezing, in addition to helping to distinguish euploid from aneuploid embryos (8, 22).

Several previous studies have used computational techniques to extract various quantitative parameters describing the blastocyst stage. Santos Filho *et al.* (23) developed a method that yielded segmented images of the ICM and TE, while Singh *et al.* (24) applied identification and segmentation algorithms to isolate the TE. Matos *et al.* (25) even used digital image processing techniques for semiautomated selection of mouse blastocyst quality.

The inclusion of more objective embryo features, such as blastocyst width and/or area, may further improve selection for successful pregnancy (26). The current study aimed to provide an objective and comprehensive assessment of human blastocyst morphology based on images captured by two distinct sources, EmbryoScope[®] and Geri[®] incubators, using various image processing techniques from the widely available Matlab[®] platform.

MATERIALS AND METHODS

Study Design

The study was approved by the Institutional Review Board (IRB) of IVI Foundation, whose code corresponds to 1709-VLC-094-MM. In this research project encrypted data was used, i.e., it was not possible to associate identified persons because the link with all information that identifies the subjects has been destroyed. Consequently, it was estimated that it would neither be necessary nor feasible to obtain the Informed Consent of the subjects, as it is a study with anonymized data. In addition, there was not express opposition from the subjects for the use of the data for the intended purposes.

Image Dataset

All raw images were of blastocysts collected from patients receiving assisted reproduction treatment between 2017 and 2019 at IVI-RMA (Valencia, Spain) and grown in either the EmbryoScope[®] time-lapse incubator from Vitrolife or the Geri[®] time-lapse incubator from Genea Biomedx (Figure 1).

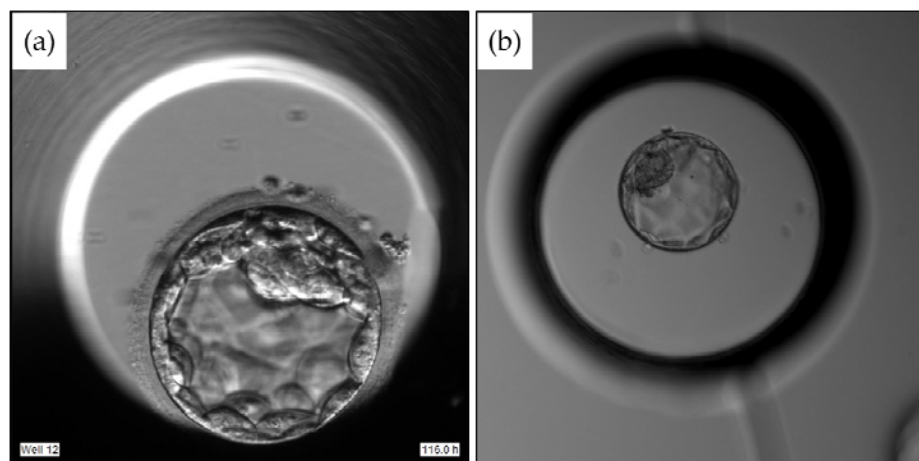


Figure 1 Human blastocysts maintained in the EmbryoScope[®] incubator (a) or the Geri[®] incubator (b).

The original dataset comprised 353 images of embryos growing in an EmbryoScope[®] incubator and 474 images of embryos growing in a Geri[®] incubator, of which 33 images from the EmbryoScope[®] incubator and 165 from the Geri[®] incubator were excluded. The 33 EmbryoScope[®] images were excluded due to poor focus ($n = 16$), display of a nonblastocyst stage embryo ($n = 14$), incomplete visualization of the entire blastocyst ($n = 2$), or insufficient light to recognize individual regions ($n = 1$). The 165 Geri[®] images were excluded due to display of a nonblastocyst stage embryo ($n = 79$), blurred inner cell mass ($n = 82$), or display of a hatched blastocyst ($n = 4$). Therefore, 320 EmbryoScope[®] and 309 Geri[®] images were included (629 images in total). Of these, 289 images from the EmbryoScope[®] incubator (90.3%) and 215 images from the Geri[®] incubator (69.6%) were correctly processed by the methods proposed in the following sections, while the other 125

images were not correctly segmented, mainly due to the low quality of the raw image.

Image Processing and Variable Extraction

The images provided by each incubator have unique characteristics (Figure 1) requiring different approaches for processing and feature extraction. The principal differences in processing relevant to image segmentation are presented in Table 1.

Table 1 Different segmentation steps required for images provided by Geri[®] and EmbryoScope[®] incubators, and the variables extracted by each step

	Segmentation Step	EmbryoScope[®]	Geri[®]	Variable Extraction
1	Conversion of the input image to grayscale	X	X	-
2	Edge detection using a Canny filter	X	X	-
3	Partial embryo isolation	X	X	-
4	Histogram adjustment (contrast, stretching, and tone)	X	X	-
5	Prewitt adjustment	X	X	-
6	Binary conversion	X	X	-
7	Isolation of the whole blastocyst	X	X	Measurement of the area and radius of the blastocyst by equations using the Matlab [®] Image Processing Toolbox [™]
8	TE and Blastocoel + ICM isolation	X	X	Variables describing TE and the Blastocoel + ICM using the Matlab [®] Image Processing Toolbox [™] , Local Binary Pattern (LBP) algorithm, and Gray level Cooccurrence matrix (GLCM)
9	Threshold adjustment	X	X	-
10	Segmentation based on the grayscale		X	-
11	Determination of the binary distance		X	-
12	ICM isolation	X	X	Variables describing the ICM using the Matlab [®] Image

Processing Toolbox™, LBP algorithm, and GLCM

ICM: inner cell mass; TE: trophoctoderm

Segmentation of embryo images required between 10 and 12 steps to completely isolate the blastocyst, ICM, and TE. The steps used for segmentation of images from each incubator and typical changes to the image following each step are presented in Figures 2 and 3.

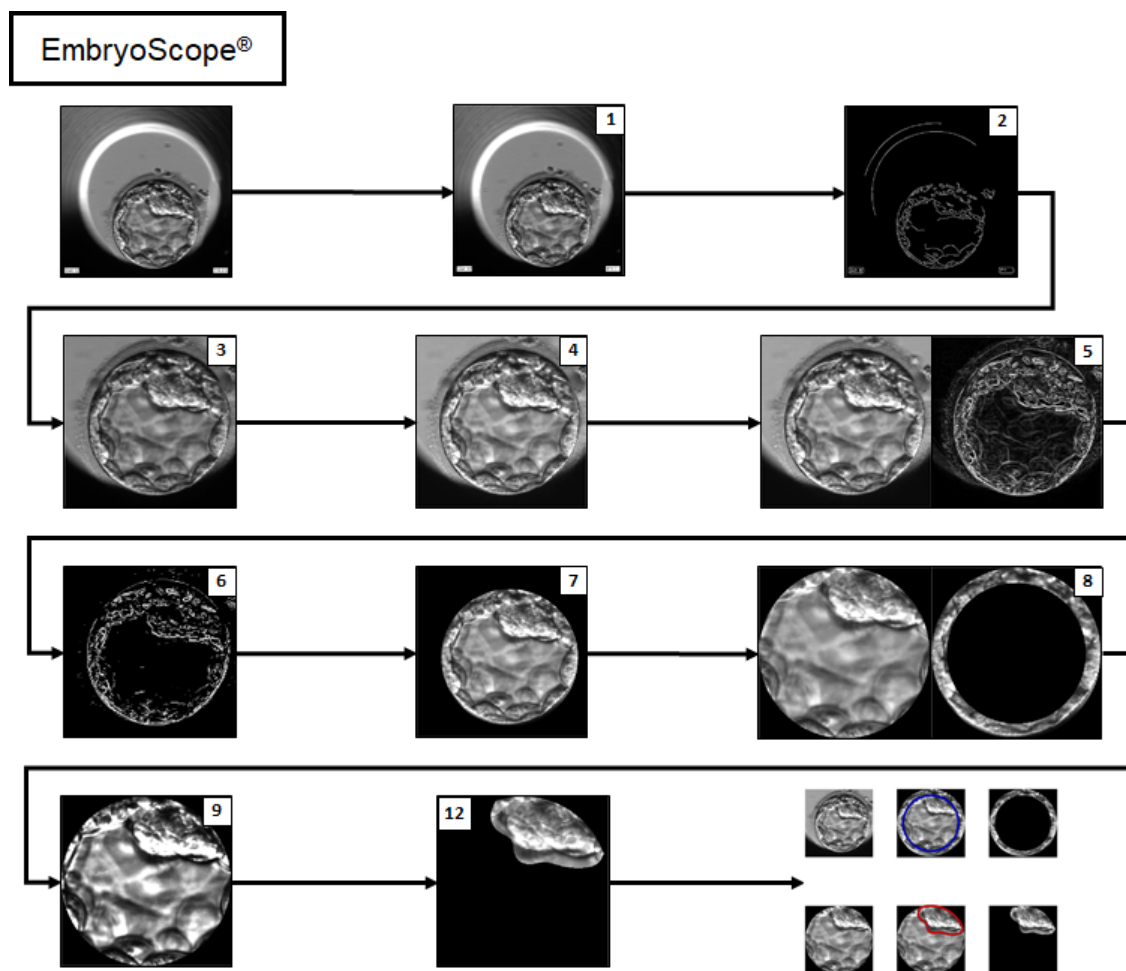


Figure 2 Flowchart of embryo segmentation from an image provided by the EmbryoScope® incubator. The first panel is the raw original image and the last panel presents each segment isolated by the software. The intermediate images are identified by the respective step numbers presented in Table I. Steps 10 and 11 are not represented in the figure since they are exclusive to Geri® image segmentation.

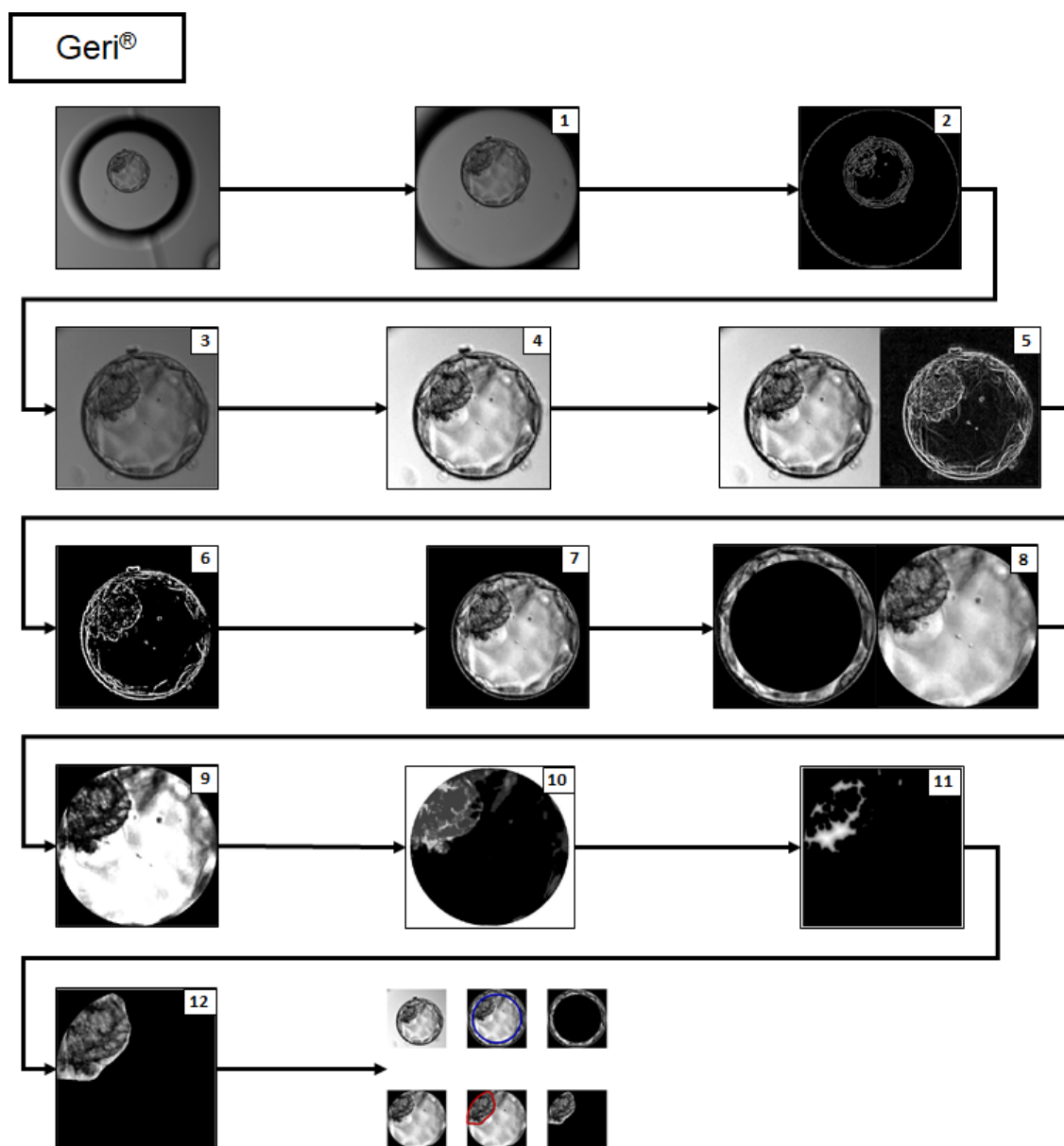


Figure 3 Flowchart of embryo segmentation from an image provided by the Geri[®] incubator. The first panel is the original raw image and the last panel shows each segment isolated by the software. The intermediate images are identified by the respective step numbers presented in Table I.

Grayscale Conversion and Image Resizing

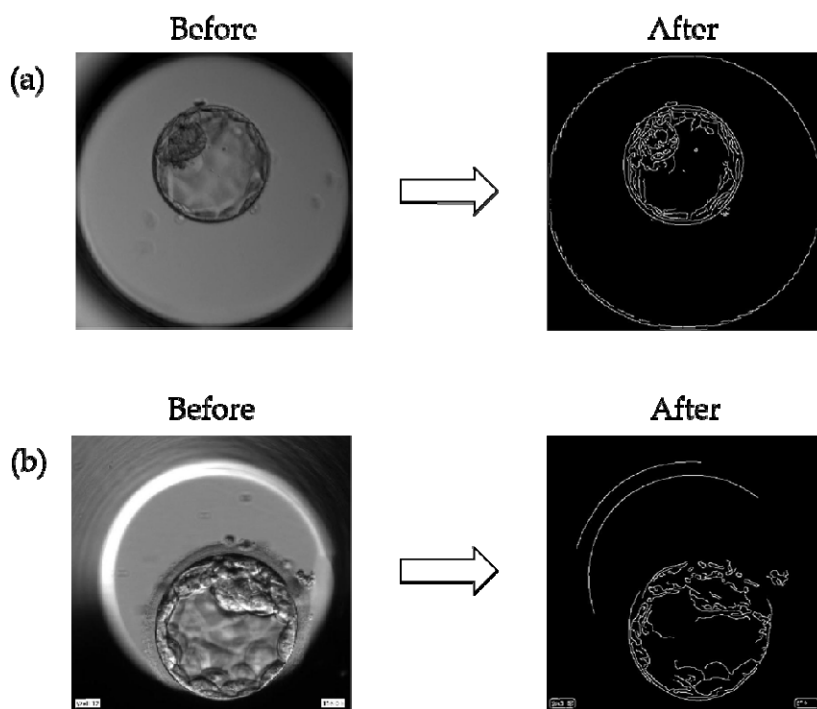
To reduce the computational load associated with image processing, raw images were first converted to a grayscale (27, 28) and adjusted to a standard size using the Matlab[®] Image Processing Toolbox[™] grayscale conversion and resizing functions (29).

Canny Filter Application for Edge Detection

In this work, a Canny filter (CF) was applied to detect image edges (Supplemental Figure 1). In addition to a CF, Santos (30) added a Gaussian filter to reduce noise for improved location of edges. Therefore, the Gaussian filter described by Equation 1 was also applied, where σ_x and σ_y represent the standard deviations of pixels along the X and Y axes, respectively.

$$G(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}} \quad (1)$$

This filter application is necessary to obtain information for isolating the region of interest (ROI), specifically the image of the blastocyst without surrounding background. A CF application can be found in the Matlab[®] Image Processing Toolbox[™] (29).



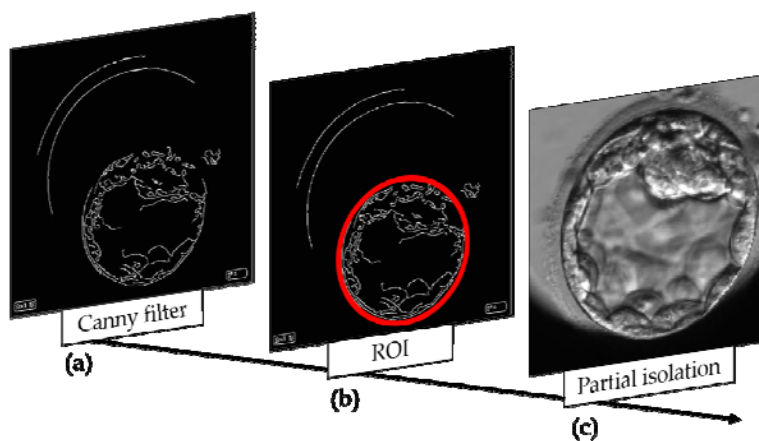
Supplemental Figure 1 Application of the Canny Filter (CF) on blastocyst images from (a) a Geri[®] incubator and (b) an EmbryoScope[®] incubator for edge detection, thereby distinguishing the blastocyst area from the background.

Detection of Circles by the Hough Transform

The Hough transform (HT) algorithm is used to detect forms such as lines, circles, and ellipses in digital images (31). According to Seifozakerini *et al.* (32), this technique is useful for the recognition of straight lines or circular forms, and so is appropriate for ROI detection since embryos present a circular form. Circle parameterization described by Hough (33) is presented in Equation 2, where (a, b) represents the central coordinate and r represents the radius.

$$f_c(x, y) = (x - a)^2 + (y - b)^2 - r^2 = 0 \quad (2)$$

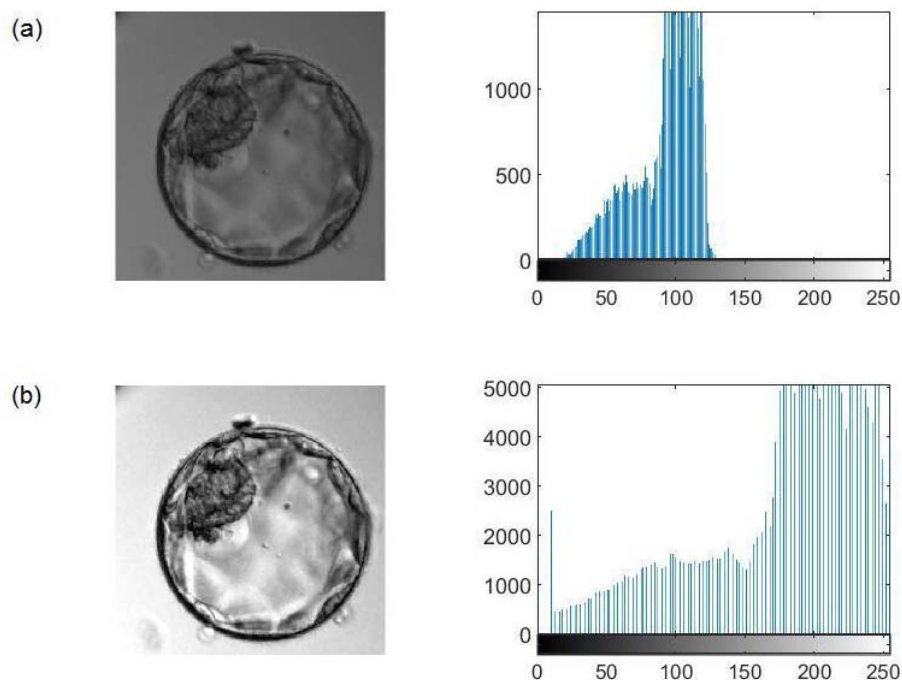
These functions yield parameters such as the circle center and radius length that in turn allow for isolation of the ROI (Supplemental Figure 2). Hough transformation was then followed by a final contrast adjustment to enhance ROI boundaries. The Matlab® Image Processing Toolbox™ has all the functions needed to apply the HT as well as the contrast adjustment (29).



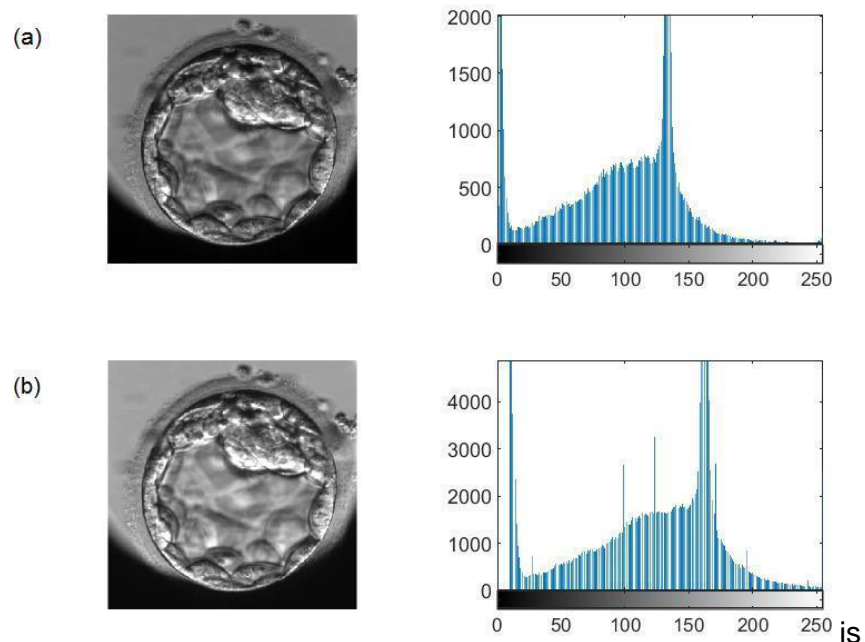
Supplemental Figure 2 Partial detection of the embryo by the Hough transform (HT). (a, b) From the edges of the image obtained by the Canny filter, circles were detected by the Hough transform to isolate the ROI (red circumference). (c) Thus, the embryo is partially isolated from the rest of the image.

Histogram Adjustment

Contrast is an essential parameter determining image resolution (34). To further improve image quality, contrast was optimized by histogram adjustment using the Matlab® Image Processing Toolbox™ (29). However, different contrast adjustments are required for images obtained by Geri® (Supplemental Figure 3) and EmbryoScope® (Supplemental Figure 4) to maintain an adequate standard for the desired variable acquisition.



Supplemental Figure 3 Histogram adjustment for an image from the Geri[®] incubator. The histogram of the original image (a) is stretched to yield a higher contrast image (b).



Supplemental Figure 4 Histogram adjustment for an image from the EmbryoScope[®] incubator. The histogram of the original image (a) is stretched to yield a higher contrast image (b).

Maxima of Gradient Magnitude

In a grayscale image, gradient magnitude (GM) is defined as the root mean square of the directional gradients along two orthogonal directions (35). In general, GM maxima can be detected at the boundaries (edges) of a region (36). For extracting the GM, several operators can be used, such as the Prewitt method, which is efficient for detecting edges. According to Yang *et al.* (37), this method can detect edges in the vertical (G_y) and horizontal (G_x) directions and calculate the GM using a pair of 3×3 convolution masks (Supplemental Figure 5). This technique is applied using the Matlab[®] Image Processing Toolbox[™].

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}$$

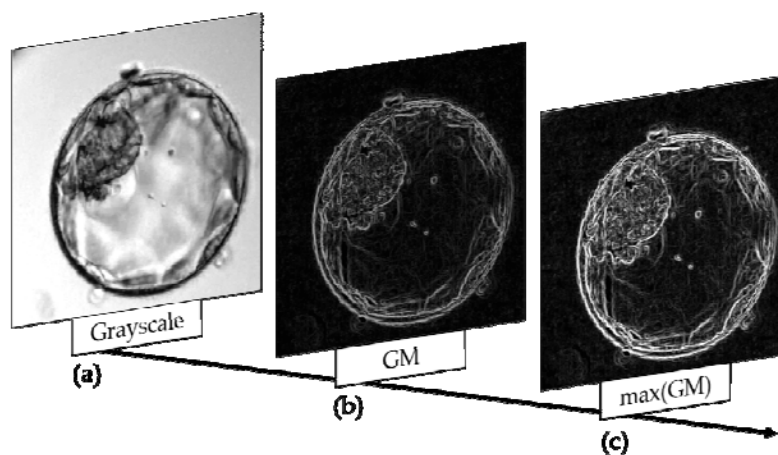
Vertical directions (G_y)

$$\begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix}$$

Horizontal directions (G_x)

Supplemental Figure 5 Directions of the Prewitt gradient operator. From the convolution masks, the Prewitt differentiation operator calculates the gradient in the vertical direction through the G_y matrix, while G_x calculates the gradient in the horizontal direction. G_y and G_x matrixes are also called derivative masks and are applied to calculate the difference of pixel intensities in an edge region of the image.

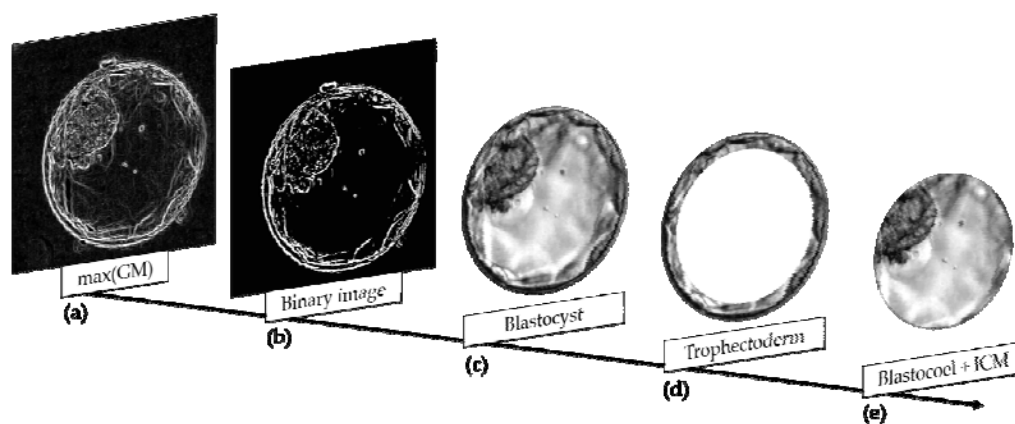
Using this technique, GMs are obtained (Supplemental Figure 6b), which can later be converted into maximum values (Supplemental Figure 6c). This process allows brightness invariant areas (i.e., regions of the image with similar characteristics) to be captured and grouped. Based on the maximum magnitude values, pixels within a specific range can be extracted.



Supplemental Figure 6 Extraction of gradient magnitude (GM) maxima. (a) Grayscale image of the blastocyst. (b) Extraction of the GM. (c) Extraction of GM maxima.

Binary Image Construction

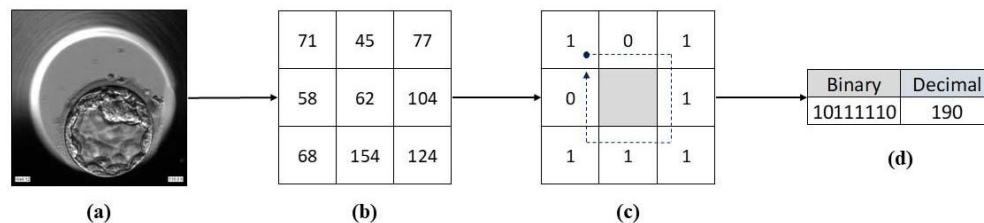
A binary image is composed of a matrix containing only 0s and 1s, where the 1s represent the object of interest (38). To obtain a binarized image, pixel values within a defined range are converted to 1s, while all other pixels are set to 0 (Supplemental Figure 7b). This process is important for applying the HT and allows total isolation of the blastocyst (Supplemental Figure 7c), TE (Supplemental Figure 7d), and blastocoel plus ICM (Supplemental Figure 7e).



Supplemental Figure 7 Isolation of the blastocyst, trophectoderm, and blastocoel plus inner cell mass (ICM). (a) Maxima of gradient magnitude. (b) Binary transformation using a defined range. (c–e) Isolation of the blastocyst (c), trophectoderm (d), and Blastocoel + ICM (e).

Segmentation of the image into blastocyst, TE, and blastocoel plus ICM allows the extraction of several variables. In this process, variation and similarity of texture, uniformity in gray level distribution, proximity of the grayscale to the GLCM diagonal in ICM and Blastocoel, mean gray level, and standard deviation of gray level are extracted. In addition, brightest and darkest regions, average brightness, most frequent light intensity (modal value), binary image sum, blastocyst radius, and area measurements are calculated.

The textures were examined by the local binary pattern (LBP) algorithm and by constructing a gray level cooccurrence matrix (GLCM). The LBP algorithm is a widely applied approach to recognize and analyze texture in images (39). In this study, the LBP technique was applied to determine the textures of the ICM, TE, and EX. In Matlab®, the LBP algorithm is included in the Computer Vision Toolbox™ (40). Feature extraction using the LBP algorithm consists of two main steps (Supplemental Figure 8), thresholding and encoding. In the first step, all neighboring pixels in each defined area (e.g., pattern) are compared to the value of the central pixel. Then, all values of the neighboring pixels are changed to a binary value (0 if lower than the value of the central pixel or 1 if larger). After this transformation, the encoding step involves conversion of the binary number obtained to a decimal value (39).



Supplemental Figure 8 Representation of the Local Binary Pattern (LBP) algorithm. (a) The initial blastocyst image. (b) The matrix of pixel values obtained from an area of the image. (c) The binary matrix derived using the LBP algorithm. The values of this matrix (0s and 1s) are read clockwise starting from the left-most upper quadrant following the arrow. (d) Through this process, a binary number is obtained and converted to a decimal number. This process is repeated until the entire image is processed.

The GLCM algorithm was also applied to examine the textures of the ICM, TE, and the blastocoel plus ICM images obtained by segmentation. This statistical method, included in the Image Processing Toolbox™, essentially examines texture by analyzing the spatial relationships among pixels, specifically how often pairs of pixels with specific values and in a specified spatial relationship occur in an image. This analysis yields the GLCM, from which statistical measures are extracted (29). Using this algorithm, it is possible to extract information such as variation of the

texture level, texture similarities, uniformity of the gray level distribution, and proximity of the grayscale to the GLCM diagonal.

The brightness and darkness variables of the image were then extracted from the TE and Blastocoel + ICM images. The darkness of the image was measured as the sum of the top 10% darkest pixels, and the brightness as the sum of the remaining pixels. The mean brightness of each region was estimated using the pixels with values between the 10% darkest and the 10% brightest. Image metrics such as radius and area were extracted from the distribution of pixels obtained by the Matlab[®] Image Processing Toolbox[™] (29).

Grayscale Intensification

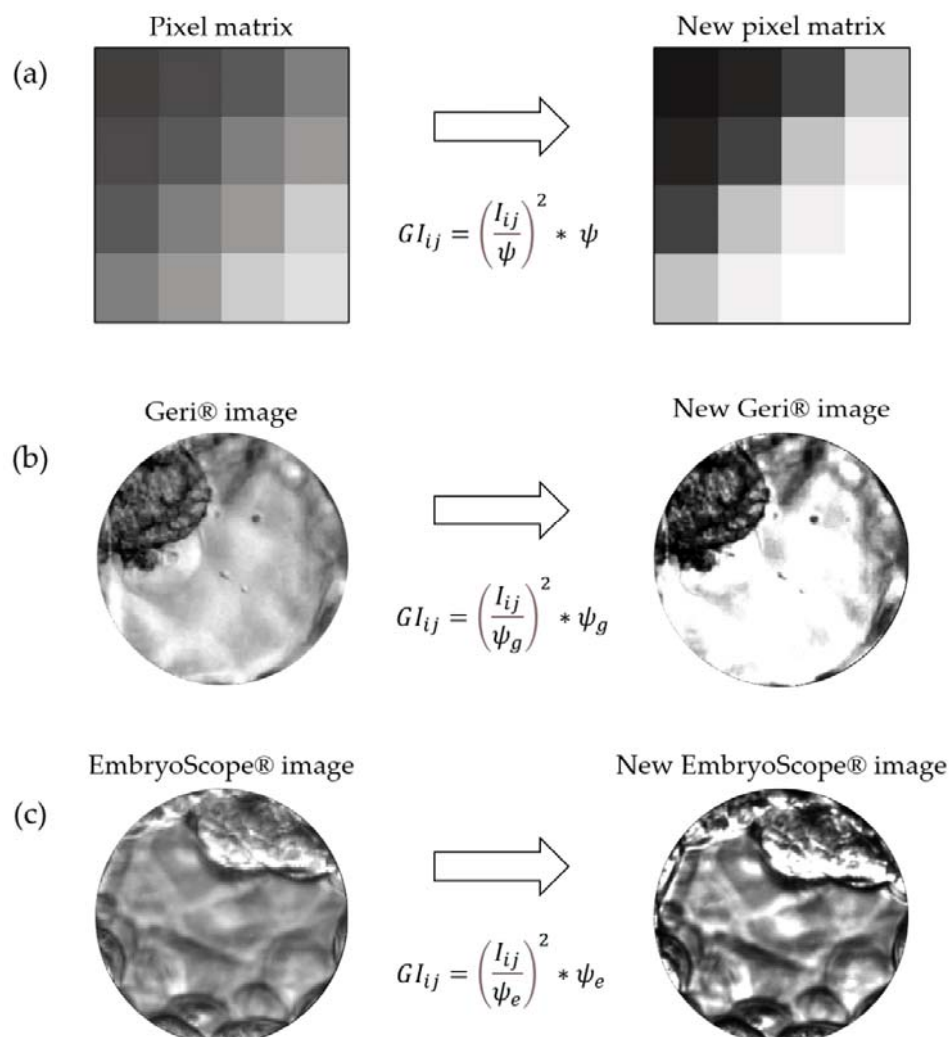
Grayscale standardization alone cannot provide good texture segmentation results. In addition, it is important to intensify the light and dark tones of the image using an optimal threshold (ψ) specific for Geri[®] and EmbryoScope[®] incubators. Using Equation 3, it is possible to convert the values of a pixel matrix (I_{ij}) and increase the frequency of maxima and minima, yielding a new higher contrast matrix (GI_{ij}) as depicted in Supplemental Figure 9.

$$GI_{ij} = \left(\frac{I_{ij}}{\psi}\right)^2 * \psi \quad \mathbf{(3)}$$

I_{ij} : grayscale pixel matrix.

ψ : threshold.

GI_{ij} : matrix containing the converted values of I_{ij} .



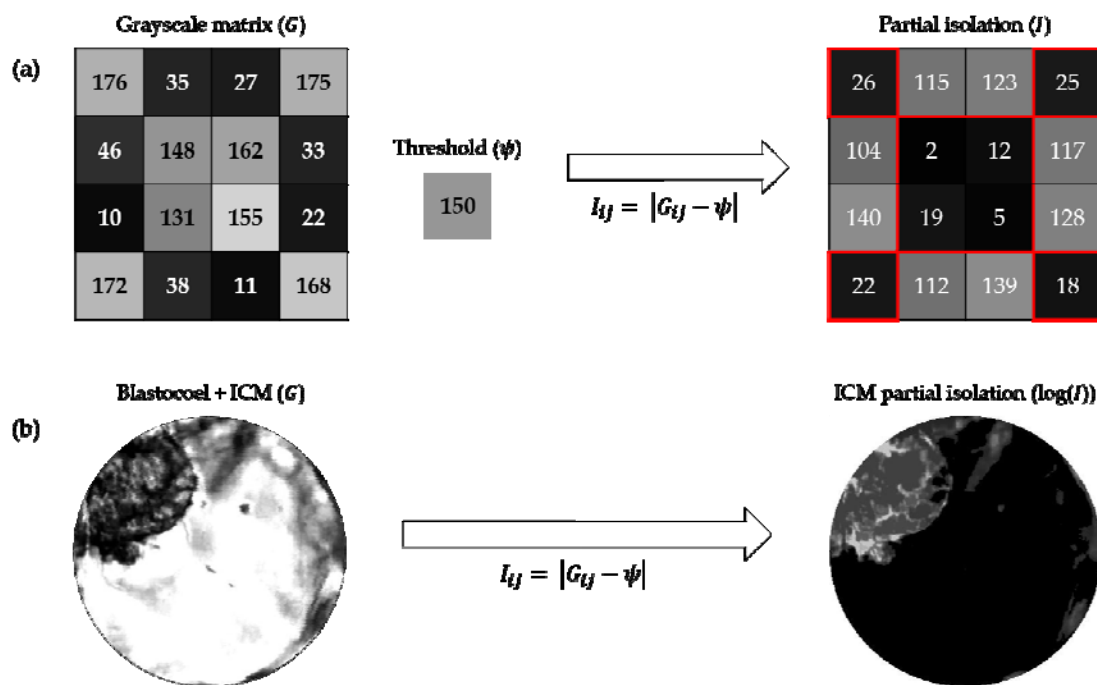
Supplemental Figure 9 Grayscale intensification. (a) Representation of the intensification process by applying a threshold ψ to the lightest and darkest tones (top 10% highest and lowest pixel values). (b) Grayscale intensification based on a defined ψ_g threshold for images from the Geri® incubator. (c) Grayscale intensification based on a defined ψ_e threshold for images from the EmbryoScope® incubator.

ICM Partial Isolation Based on the Grayscale

Segmentation based on grayscale is a step specific to images from the Geri® incubator, which have greater contrast between light and dark tones. In this process, the weight of each pixel in the image is calculated as the absolute difference between the pixel value (G_{ij}) and a previously defined threshold (ψ) (Equation 4). Using these calculated weights, it is possible to generate a new image (I) containing the partially segmented ICM. Resources from Matlab® Image Processing Toolbox™ (29) were used for this process. The method is illustrated in Supplemental Figure

10a, and the results of partial ICM segmentation are shown in Supplemental Figure 10b.

(4)

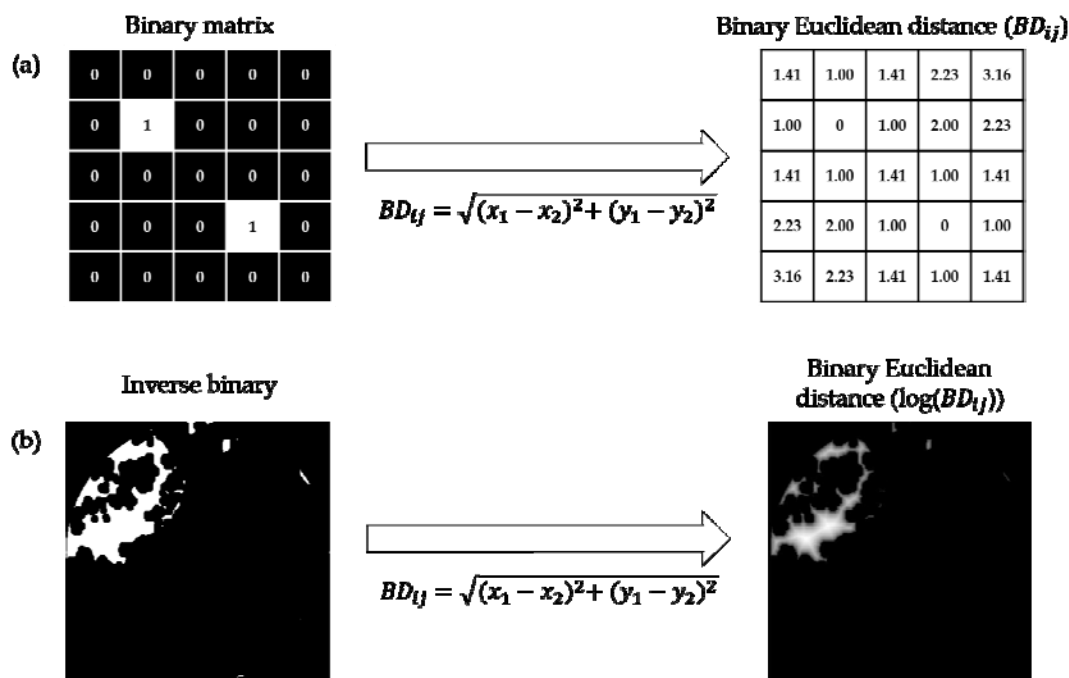


Supplemental Figure 10 Grayscale segmentation. (a) A grayscale pixel matrix () can be partially segmented by the difference between each pixel value and a set threshold (). The new partially segmented matrix is termed (). (b) Process applied to the blastocoel and ICM. It is possible to visualize the result of partial segmentation using a weight matrix logarithm.

Determination of Binary Distance

Another technique used only for Geri[®] images is binary Euclidean distance () calculation (Equation 5). The binary distance is defined as the distance from every pixel to the nearest nonzero pixel (41). After binary transformation of the partially segmented ICM, the inverse of this image is submitted to the distance calculation process (Supplemental Figure 11). This improves the quality of the final segmentation by eliminating noise present in the image.

(5)

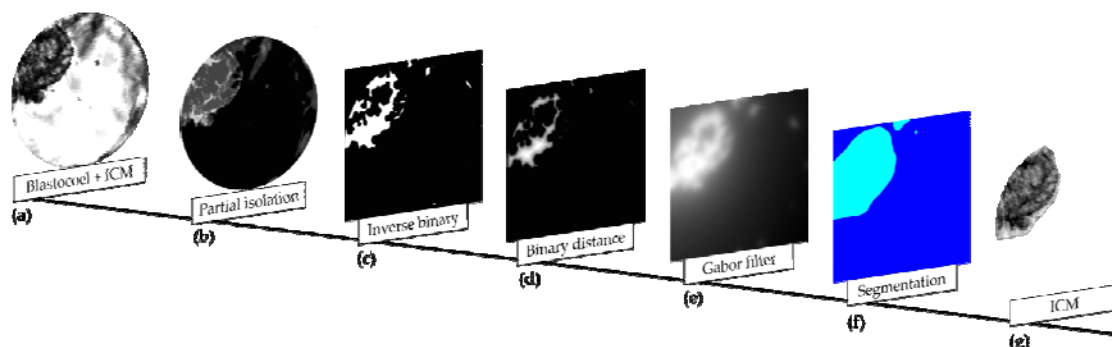


Supplemental Figure 11 Noise removal by calculating binary Euclidean distance. (a) Euclidean distance is calculated for the binary matrix. The new matrix obtained () contains the distance from each pixel to the nearest nonzero pixel. (b) After binary conversion and inversion of 0s and 1s (yielding an inverse binary), this process is applied to the pixel matrix and an image with reduced noise is obtained. It is possible to visualize the new image after applying a logarithm to the pixel matrix [$\log(BD_{ij})$].

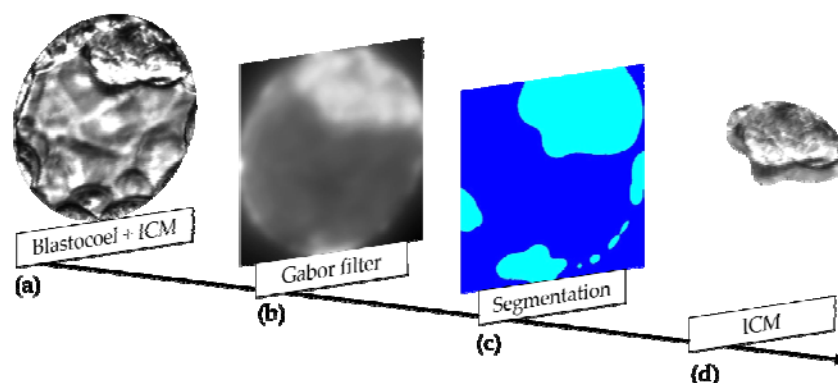
ICM Isolation by Gabor Filter

Segmentation by Gabor multichannel filters is an established method in image processing. This technique allows characterization and separation of textures by creating channels for different frequencies and orientations, similar to the human visual system. To segment the ICM and obtain the Gabor magnitude, we initially used the combinations of frequency and orientation presented by Jain and Farrokhnia (42). After this process, Gaussian low-pass filters were applied (Supplemental Figure 12e and Supplemental Figure 13b), an important postprocessing step to normalize the different outputs and decrease the magnitude variations according to Kim and Kang (43). Gaussian low-pass filters enhance smoothness and reduce noise (the difference between neighboring gray levels). In this way, it is possible to differentiate image regions in different planes (Supplemental Figure 12f and Supplemental Figure 13c) and fully segment the ICM (Supplemental Figure 12g and Supplemental Figure 13d). In the last steps,

clustering is performed using the Matlab[®] Statistics and Machine Learning Toolbox[™] (44).



Supplemental Figure 12 Isolation of the ICM by Gabor filtering of an image from the Geri[®] incubator. (a) Isolation of the Blastocoele + ICM. (b) Partial isolation of the ICM. (c) After partial isolation, an inverse binary matrix was obtained. (d) Binary distance matrix. (e) When applying the Gabor filter, low-pass Gaussian filtering is also used to differentiate regions of the previous image. (f) Two main planes are obtained, each representing the different textures of the image. (g) Isolation of the ICM.



Supplemental Figure 13 Isolation of the ICM by Gabor filtering of an image from the EmbryoScope[®] incubator. (a) Isolation of the Blastocoele + ICM. (b) Application of the Gabor filter and low-pass Gaussian filter. (c) Two main planes representing different textures of the image. (d) Isolation of the ICM.

Additional variables were extracted by isolation of the ICM. The textures were then analyzed by the LBP algorithm and the GLCM. The area was calculated by the distribution of the pixels in this segmentation. The blastocoele area was calculated by subtraction of the area of the ICM (Supplemental Figure 12g and Supplemental Figure 13d) from the total area of the Blastocoele + ICM image (Supplemental Figure

12a and Supplemental Figure 13a). Expansion of the embryo was estimated indirectly by calculating the area of embryo (without the TE) occupied by the ICM.

RESULTS

The image processing steps summarized in Table I automatically generated a total of 33 variables, each representing a different feature of the expanding embryo (EX) or a subregion (ICM or TE). These variables can be divided into three main categories, EX, ICM, and TE, according to the ROI represented (Supplemental Table 1).

Supplemental Table 1 Variables obtained through image processing categorized according to region of interest represented: Expansion of the blastocyst (EX), inner cell mass (ICM), and trophectoderm (TE)

EX	ICM	TE
Local Texture Descriptor	Local Texture Descriptor	Local Texture Descriptor
Brightest region in ICM and Blastocoel	Texture variation	Brightest region
Mean luminosity in ICM and Blastocoel	Texture similarities	Mean luminosity
Blastocyst radius	Uniformity of the gray level distribution	Darkest region
Blastocyst sum of binary image	Proximity of the grayscale to the GLCM diagonal	Modal value
Blastocyst gray level average	ICM Area	Gray level average
Gray level standard deviation in ICM and Blastocoel		Gray level standard deviation
ICM and Blastocoel gray level average		Texture variation

Modal value of ICM and Blastocoel		Texture similarities
Darkest region in ICM and Blastocoel		Uniformity of the gray level distribution
Texture variation in ICM and Blastocoel		Proximity of the grayscale to the GLCM diagonal
Texture similarities in ICM and Blastocoel		
Uniformity of gray level distribution in ICM and Blastocoel		
Proximity of the grayscale to the GLCM diagonal in ICM and Blastocoel		
Blastocoel area		
Ratio between ICM and Blastocoel area		

These variables can also be divided into six categories according to the image property represented: texture, gray level average; gray level standard deviation, modal value (most frequent pixel value), relations, and light level (Supplemental Table 2).

Texture variables denote repeating random regular patterns that provide measures of structural arrangements on surfaces. They represent the different interactions among pixels, from the differences in gray level from pixel to pixel in local regions of the image to the spatial arrangement of gray levels throughout the image. Gray level average, gray level standard deviation, and modal value represent the overall brightness/darkness and variation in brightness of the embryo. Relations refer to associations between otherwise distinct features, such as radius and area of the blastocyst. Finally, the light level variables represent the brightness variation in different regions of the embryo.

Supplemental Table 2 Variables derived from image processing categorized according to image property

Texture	Gray level average	Gray level standard deviation	Modal value	Relations	Light level
Texture variation in ICM and Blastocoel	Gray level average in ICM and Blastocoel	Gray level standard deviation in ICM and Blastocoel	Modal value in ICM and Blastocoel	Blastocyst sum of binary image	Mean luminosity in ICM and Blastocoel
Texture variation in ICM	Gray level average in TE	Gray level standard deviation in TE	Modal value in TE	Blastocyst radius	Mean luminosity in TE
Texture variation in TE	Blastocyst gray level average			ICM area	Brightest region in ICM and Blastocoel
Texture similarities in ICM and Blastocoel				Blastocoel area	Brightest region in TE
Texture similarities in ICM				Ratio between ICM and Blastocoel area	Darkest region in ICM and Blastocoel
Texture similarities in TE					Darkest region in TE
Uniformity of the gray level distribution in ICM and Blastocoel					

Uniformity of the gray level distribution in ICM					
Uniformity of the gray level distribution in TE					
Proximity of the grayscale to the GLCM diagonal in ICM and Blastocoel					
Proximity of the grayscale to the GLCM diagonal in ICM					
Proximity of the grayscale to the GLCM diagonal in TE					
Local texture descriptor in EX					
Local texture					

descriptor in ICM					
Local texture descriptor in TE					

Variable definitions

Texture

1. Texture variation in ICM and Blastocoel

Represents the texture gray level differences in ICM and Blastocoel. Reflects image sharpness and depth of grooves. Deeper grooves are related to greater variation and sharpness.

2. Texture similarities in ICM and Blastocoel

Represents the linear dependency of gray levels.

3. Uniformity of the gray level distribution in ICM and Blastocoel

Represents the uniformity of the gray level distribution through the sum of GLCM elements squared.

4. Proximity of the grayscale to the GLCM diagonal in ICM and Blastocoel

Represents the distribution of elements in the GLCM according to the diagonal reference values.

5. Texture variation in ICM; 6. Texture similarities in ICM; 7. Uniformity of gray level distribution in ICM; 8. Proximity of the grayscale to the GLCM diagonal in ICM

These variables correspond to variables 1, 2, 3, and 4, respectively, but are calculated from the ICM segmented image.

9. Texture variation in TE; 10. Texture similarities in TE; 11. Uniformity of the gray level distribution in TE; 12. Proximity of the grayscale to the GLCM diagonal in TE

These variables correspond to variables 1, 2, 3, and 4, respectively, but are calculated from the TE segmented image.

13. Local texture descriptor in EX

Texture pattern descriptor used to describe the local texture patterns of an image. The algorithm extracts information about the local texture by establishing a threshold for “n” neighbors relative to the value of the central pixel. The result is a binary number between 0 and 255.

14. Local texture descriptor in ICM; 15. Local texture descriptor in TE

These variables correspond to variable 13, but are calculated from ICM and TE segmented images, respectively.

Gray level average

16. Gray level average in ICM and Blastocoel

Represents the average pixel intensity (gray level value) within the segmented ICM plus Blastocoel image.

17. Gray level average in TE

Represents the average pixel intensity (gray level value) within the segmented TE.

18. Blastocyst gray level average

Represents the average pixel intensity (gray level value) for the entire blastocyst with zona pellucida.

Gray level standard deviation

19. Gray level standard deviation in ICM and Blastocoel

Represents the standard deviation of pixel intensity within the segmented ICM and Blastocoel.

20. Gray level standard deviation in TE

Represents the standard deviation of pixel intensity within the segmented TE.

Modal value

21. Modal value in ICM and Blastocoel

Represents the most frequent light intensity value in the segmented ICM and Blastocoel.

22. Modal value in TE

Represents the most frequent light intensity value in the segmented TE.

Relations

23. Blastocyst sum of binary image

The image threshold is given by the sum of the binary image divided by the area of the isolated blastocyst calculated from the radius.

24. Blastocyst radius

Radius of the blastocyst.

25. ICM area

Area of the segmented ICM.

26. Blastocoel area

Area of the segmented blastocoel.

27. Ratio between ICM and Blastocoel

Represents the expansion of the blastocyst.

Light level

28. Mean luminosity in ICM and Blastocoel

All pixels with a luminous intensity between 10 greater than and 10 less than the average intensity are counted, and this value is then divided by the total area of the blastocyst.

29. Mean luminosity in TE

Calculated as described for variable 28 but using values from the segregated TE.

30. Brightest region in ICM and Blastocoel

Brightest area in the segmented ICM and Blastocoel.

31. Brightest region in TE

Brightest area in the segmented TE.

32. Darkest region in ICM and Blastocoel

Darkest area in the segmented ICM and Blastocoel.

33. Darkest region in TE

Darkest area in the segmented TE.

While 33 variables were extracted, it is unclear if all can be modulated independently. Collinearity refers to a strong association between two independent variables and multicollinearity as a strong association among more than two independent variables. Multicollinearity alters the grouping because the collinear variables end up implicitly receiving a greater weight. A collinearity analysis can be performed to identify if any of these 33 variables are related (45).

DISCUSSION

In this work, we present a sequence of digital processing steps to segment human embryo images obtained from two distinct sources. The method produced 33 mathematical variables from the whole embryo and its specific regions (ICM, TE, and blastocoel). We speculate that these variables may help in identifying embryos most suitable for freezing and ultimate implantation for ART.

This sequence of steps was adapted from previous studies by our group on murine embryos (25) and bovine embryos (21, 45, 46). The relatively high success rate of segmentation, especially for embryos photographed in the EmbryoScope[®] incubator (90.3%), suggests that these previous findings were beneficial for analysis human of embryo images, a more challenging application due to both greater variability in embryo morphology compared to mice and cattle (47) and the more rigorous standards that must be met for ART.

In addition to work on nonhuman species, preliminary results have been published on digital processing of human blastocyst images from EmbryoScope[®] for implantation selection (48, 49). In contrast, Geri[®] images have been used much less extensively for AI development (50). This paucity of research may reflect less ideal imaging conditions. Indeed, the rate of successful segmentation was much lower using Geri[®] images (69.6%) and required extra steps.

Several of these proposed steps, such as application of the HT algorithm, have already been applied in previous image processing studies. Van den Heuvel *et al.* (51) used the HT to detect the center of the fetal skull and a dynamic programming algorithm to detect the outside of the fetal skull. They also fitted an ellipse to the dynamic programming result to measure head circumference, estimate gestational age, and monitor fetal growth. Huang *et al.* (52) also used the GLCM approach to identify textural features on ovarian adenocarcinoma cells indicative of chemoresistance. Specifically, the GLCM was used to calculate contrast, energy, entropy, and homogeneity, variables that collectively can reveal the disordered surface morphology characteristic of cancer cells. Alternatively, Nanni *et al.* (53) used the LBP algorithm to derive texture descriptors for classifying 2D HeLa images, cells in pap smear datasets, and pain levels from the facial images of newborns.

Thus, our chosen algorithms constitute a set of mathematical tools already used for a wide range of image processing applications.

Some of the variables extracted may have little or no biological significance despite conveying information on texture, intensity of edges, and other features. However, these quantifiable features may still be advantageous over the subjective imaging criteria used to grade embryos (18, 54). As proof of concept for the potential utility of these variables in embryo selection, we are performing a pilot study (in progress) on prediction of pregnancy and aneuploidy. Briefly, the same 33 extracted variables were used as inputs for artificial neural network (ANN)-based programs to predict aneuploidy or pregnancy (fetal heartbeat) from blastocyst images. The ANNs were chosen based on a genetic algorithm search for the fittest model, like that described by Rocha *et al.* (21, 45). We recently obtained 96% predictive accuracy for the presence of a fetal heartbeat based on processing of 172 images from EmbryoScope[®] and receiver operating characteristic curve analysis as well 95% accuracy for prediction of aneuploidy based on 277 images using the same analytic processes. Fine tuning the image processing protocol by creating high-resolution mosaic images of the whole embryo in segmentation masks to provide additional information on certain structures (e.g., ICM or TE) or features (texture, contrast, brightness, or edges) inscrutable by visual inspection may further enhance predictive efficacy.

In summary, we describe an image processing protocol that can successfully segment human blastocyst images from two distinct sources and extract 33 variables with potential utility in embryo selection for ART.

DATA AVAILABILITY STATEMENT

The human blastocyst images were provided by the IVI-RMA (Valencia, Spain). The steps, algorithms, and the flowcharts used in this work are all clearly described in the main text. Any additional information not restricted by the confidential agreement can be provided by the authors on reasonable request.

AUTHORS' ROLES

D.S.C., M.M., J.C.R., and M.F.G.N. conceived the idea and planned the study. A.S.F., E.I.F., G.M.P., W.A., I.S.J., R.C.M.S., L.B., and J.C.R. conducted the experiments. D.S.C., A.S.F., E.I.F., G.M.P., W.A., R.C.M.S., J.C.R., and M.F.G.N. conducted the analysis and interpretation of all experiments. D.S.C., A.S.F., E.I.F., G.M.P., E.A.D.M., J.C.R., and M.F.G.N. contributed to critical discussions. D.S.C., A.S.F., E.I.F., G.M.P., W.A., E.A.D.M., I.S.J., R.C.M.S., L.B., M.M., J.C.R., and M.F.G.N. wrote and revised the manuscript.

FUNDING

This work is supported by grants #2017/19323-5, #2018/24252-2, #2019/26684-0, #2019/26749-4, and #2020/07634-9, São Paulo Research Foundation (FAPESP); this study was financed in part by the Coordenação de

Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001; and National Council for Scientific and Technological Development (CNPq PIBIC/PIBITI), process #2020/981 and #2020/979.

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest relevant to this study.

REFERENCES

1. Swain J, Vermilyea MT, Meseguer M, Ezcurra D, Fertility AI Forum Group. AI in the treatment of fertility: key considerations. *J Assist Reprod Genet.* 2020 Nov;37(11):2817-24.
2. Bormann CL, Curchoe CL, Thirumalaraju P, Kanakasabapathy MK, Gupta R, Pooniwala R, et al. Deep learning early warning system for embryo culture conditions and embryologist performance in the ART laboratory. *J Assist Reprod Genet.* 2021 Apr 27;38:1641-6.
3. Letterie G. Three ways of knowing: the integration of clinical expertise, evidence-based medicine, and artificial intelligence in assisted reproductive technologies. *J Assist Reprod Genet.* 2021 Apr 19;38:1617-25.
4. Louis CM, Erwin A, Handayani N, Polim AA, Boediono A, Sini I. Review of computer vision application in in vitro fertilization: the application of deep learning-based computer vision technology in the world of IVF. *J Assist Reprod Genet.* 2021 Apr 03;38:1627-39.
5. Zaninovic N, Rosenwaks Z. Artificial intelligence in human in vitro fertilization and embryology. *Fertility and Sterility.* 2020 Nov 1;114(5):914-20.
6. Giscard d'Estaing S, Labrune E, Forcellini M, Edel C, Salle B, Lornage J, et al. A machine learning system with reinforcement capacity for predicting the fate of an ART embryo. *Syst Biol Reprod Med.* 2021 Mar 10;67(1):64-78.
7. Chéles DS, Molin EAD, Rocha JC, Nogueira MFG. Mining of variables from embryo morphokinetics, blastocyst's morphology and patient parameters: an approach to predict the live birth in the assisted reproduction service. *JBRA Assist Reprod.* 2020 Oct 6;24(4):470-9.
8. Fernandez EI, Ferreira AS, Cecílio MHM, Chéles DS, de Souza RCM, Nogueira MFG, et al. Artificial intelligence in the IVF laboratory: overview through the application of different types of algorithms for the classification of reproductive data. *J Assist Reprod Genet.* 2020 Jul 11;37:2359-76.
9. Yang SH, Wu CH, Chen YC, Yang CK, Wu TH, Chen PC, et al. Effect of morphokinetics and morphological dynamics of cleavage stage on embryo

- developmental potential: a time-lapse study. *Taiwan J Obstet Gynecol.* 2018 Feb;57(1):76-82.
10. Strouthopoulos C, Anifandis G. An automated blastomere identification method for the evaluation of day 2 embryos during IVF/ICSI treatments. *Comput Methods Programs Biomed.* 2018 Mar;156:53-9.
 11. Patil SN, Wali UV, Swamy MK. Application of vessel enhancement for automated classification of human in vitro-fertilized (IVF) images. *International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT)*; 2016 Dec 9-10; Mysuru, India. *IEEE.* 2017. p. 27-36.
 12. Montag M, Toth B, Strowitzki T. New approaches to embryo selection. *Reprod Biomed Online.* 2013 Nov;27(5):539-46.
 13. Storr A, Venetis CA, Cooke S, Kilani S, Ledger W. Inter-observer and intra-observer agreement between embryologists during selection of a single Day 5 embryo for transfer: a multicenter study. *Hum Reprod.* 2017 Feb 1;32(2):307-14.
 14. Friedenthal J, Hernandez-Nieto C, Roth RM, Slifkin R, Gounko D, Lee JA, et al. Clinical implementation of algorithm-based embryo selection is associated with improved pregnancy outcomes in single vitrified warmed euploid embryo transfers. *J Assist Reprod Genet.* 2021 May 01;38:1647-53.
 15. Bormann CL, Thirumalaraju P, Kanakasabapathy MK, Kandula H, Souter I, Dimitriadis I, et al. Consistency and objectivity of automated embryo assessments using deep neural networks. *Fertil Steril.* 2020 Apr;113(4):781-7.
 16. Gonzalez RC, Woods RE. *Digital Image Processing.* 3rd ed. New Jersey: Pearson; 2007.
 17. Rocha JC, Passalia FJ, Matos FD, Maserati MP, Alves MF, Almeida TG, et al. Methods for assessing the quality of mammalian embryos: how far we are from the gold standard?. *JBRA Assist Reprod.* 2016 Aug 1;20(3):150-8.
 18. Gardner DK, Schoolcraft WB. In vitro culture of human blastocysts. In: Jansen R and Mortimer D, editors. *Towards Reproductive Certainty: Fertility and Genetics Beyond.* Carnforth: Parthenon Press; 1999. p. 377-88.
 19. Puga-Torres T, Blum-Rojas X, Blum-Narváez M. Blastocyst classification systems used in Latin America: is a consensus possible?. *JBRA Assist Reprod.* 2017 Set 01;21(3):222-9.
 20. Machtinger R, Racowsky C. Morphological systems of human embryo assessment and clinical evidence. *Reprod Biomed Online.* 2013 Mar;26(3):210-21.

21. Rocha JC, Passalia FJ, Matos FD, Takahashi MB, Maserati MP, Alves MF, et al. Automatized image processing of bovine blastocysts produced in vitro for quantitative variable determination. *Sci Data*. 2017a Dez 19;4:170192.
22. Rocafort E, Enciso M, Leza A, Sarasa J, Aizpurua J. Euploid embryos selected by an automated time-lapse system have superior SET outcomes than selected solely by conventional morphology assessment. *J Assist Reprod Genet*. 2018 Jul 20;35(9):1573-83.
23. Santos Filho E, Noble JA, Poli M, Griffiths T, Emerson G, Wells D. A method for semi-automatic grading of human blastocyst microscope images. *Hum Reprod*. 2012 Sep;27(9):2641-8.
24. Singh A, Au J, Saeedi P, Havelock J. Automatic segmentation of trophoctoderm in microscopic images of human blastocysts. *IEEE Trans Biomed Eng*. 2015 Jan;62(1):382-93.
25. Matos FD, Rocha JC, Nogueira MFG. A method using artificial neural networks to morphologically assess mouse blastocyst quality. *J Anim Sci Technol*. 2014 Aug 30;56(15):1-10.
26. Sciorio R, Meseguer M. Focus on time-lapse analysis: blastocyst collapse and morphometric assessment as new features of embryo viability. *Reprod Biomed Online*. 2021 Nov;43(5):821-32.
27. Leung AS, Son WY, Dahan MH. Time-lapse imaging of embryos: current evidence supporting its use. *Expert Rev Med Devices*. 2016 Aug 31;13(10):881-3.
28. Virmani J, Singh GP, Singh Y. PNN-Based Classification of Retinal Diseases Using Fundus Image. *Sensors for Health Monitoring*. 2019;5:215-42.
29. The MathWorks Inc. *Image Processing Toolbox™ User's Guide*. Natick: The MathWorks; 2019. 1614 p.
30. Santos LA. [Aplicação do processamento de imagem digital para a detecção de câncer de mama]. In Simões AS, Colombini, EL, Tonidandel F, editors. *Mostra Nacional de Robótica 2018: Anais da Mostra Nacional de Robótica*; 2018 Nov 6-10; João Pessoa, Brazil. MNR; 2018. p. 580-2. Portuguese.
31. Sit B, Quraishi I. A review paper on Hough transform and its applications in image processing. *Int J Innov Res Sci Eng Technol*. 2016 Oct;5(13):206-13.
32. Seifozakerini S, Yau WY, Mao K, Nejati H. Hough transform implementation for event-based systems: concepts and challenges. *Front Comput Neurosci*. 2018 Dez 21;12:103.

33. Hough PV, inventor. U.S. Atomic Energy Commission, assignee. Method and Means for Recognizing Complex Patterns. United States Patent US n° 3,069,654. 1962 Dec 18.
34. Al-Ameen Z. Contrast enhancement of medical images using statistical methods with image processing concepts. 6th International Engineering Conference "Sustainable Technology and Development" (IEC); 2020 Feb 26-27; Erbil, Iraq. IEEE 2020. p. 169-73.
35. Xue W, Zhang L, Mou X, Bovik AC. Gradient magnitude similarity deviation: a highly efficient perceptual image quality index. IEEE Trans Image Process. 2014 Feb;23(2):684-95.
36. Faraji M, Shanbehzadeh J, Nasrollahi K, Moeslund TB. Extremal regions detection guided by maxima of gradient magnitude. IEEE Trans Image Process. 2015 Set 07;24(12):5401-15.
37. Yang L, Wu X, Zhao D, Li H, Zhai J. An improved Prewitt algorithm for edge detection based on noised image. 4th International congress on image and signal processing; 2011 Oct 15-17; Shanghai, China. IEEE 2011. p. 1197-1200.
38. Gonzalez RC, Woods RE, Eddins SL. Digital Image Processing Using MATLAB. 3rd ed. New Jersey: Prentice Hall; 2003.
39. Muslihah I, Muqorobin M. Texture characteristic of local binary pattern on face recognition with probabilistic linear discriminant analysis. Int J Comput Inf Syst. 2020 May;1(1):22-6.
40. The MathWorks Inc. Computer Vision Toolbox™ User's Guide. Natick: The MathWorks; 2019. 1504 p.
41. Chen Q, Yang X, Petriu EM. Watershed segmentation for binary images with different distance transforms. 3rd IEEE International Workshop on Haptic, Audio and Visual Environments and Their Applications; 2004 Oct 2-3; Ottawa, Canada. IEEE 2005. p. 111-6
42. Jain AK, Farrokhnia F. Unsupervised texture segmentation using Gabor filters. Pattern Recognit. 1991;24(12):1167-86.
43. Kim SC, Kang TJ. Texture classification and segmentation using wavelet packet frame and Gaussian mixture model. Pattern Recognit. 2007 Apr;40(4):1207-21.
44. The MathWorks Inc. Statistics and Machine Learning Toolbox™ User's Guide. Natick: The MathWorks; 2020. 9684 p.

45. Rocha JC, Passalia FJ, Matos FD, Takahashi MB, Ciniciato DdS, Maserati MP, et al. A method based on artificial intelligence to fully automatize the evaluation of bovine blastocyst images. *Sci Rep.* 2017b Aug 9;7(1):7659.
46. Nogueira MFG, Guilherme VB, Pronunciate M, Santos PH, Silva DLB, Rocha JC. Artificial intelligence-based grading quality of bovine blastocyst digital images: direct capture with juxtaposed lenses of smartphone camera and stereomicroscope ocular lens. *Sensors.* 2018 Dec 15;18(12):1-15.
47. Wamaitha SE, Niakan KK. Human pre-gastrulation development. *Curr Top Dev Biol.* 2018;128:295-338.
48. Rocha C, Nogueira MG, Zaninovic N, Hickman C. Is AI assessment of morphokinetic data and digital image analysis from time-lapse culture predictive of implantation potential of human embryos?. *Fertil Steril.* 2018 Oct 10;110(4):e373.
49. Zaninovic N, Rocha CJ, Zhan Q, Toschi M, Malmsten J, Nogueira MFG, et al. Application of artificial intelligence technology to increase the efficacy of embryo selection and prediction of live birth using human blastocysts cultured in a time-lapse incubator. *Fertility and Sterility.* 2018 Sep 1;110(4):e372-e3.
50. Alegre L, Bori L, de los Angeles Valera M, Nogueira MFG, Ferreira AS, Rocha JC, et al. First application of artificial neuronal networks for human live birth prediction on geri time-lapse monitoring system blastocyst images. *Fertil Steril.* 2020 Oct 17;114(3):e140.
51. Van Den Heuvel TLA, De Bruijn D, De Korte CL, Ginneken BV. Automated measurement of fetal head circumference using 2D ultrasound images. *PLOS ONE.* 2018 Aug 23;13(8):e0200412.
52. Huang CL, Lian MJ, Wu YH, Chen WM, Chiu WT. Identification of human ovarian adenocarcinoma cells with cisplatin-resistance by feature extraction of gray Level Co-occurrence matrix using optical images. *Diagnostics.* 2020 Jun 9;10(6):1-8.
53. Nanni L, Lumini A, Brahmam S. Local binary patterns variants as texture descriptors for medical image analysis. *Artif Intell Med.* 2010 Jun 49(2):117-25.
54. Van Den Abbeel E, Balaban B, Ziebe S, Lundin K, Cuesta MJG, Klein BM, et al. Association between blastocyst morphology and outcome of single-blastocyst transfer. *Reprod Biomed Online.* 2013 Oct 1;27(4):353-61.

