Detecting the Area of Bovine Cumulus Oocyte Complexes Using Deep Learning and Semantic Segmentation

Georgios ATHANASIOU [a], Jesus CERQUIDES [a], Annelies RAES [b],
Nima AZARI-DOLATABAD [b], Daniel ANGEL-VELEZ [b], Ann VAN SOOM [b], and
Josep-Lluis ARCOS [a]

[a] *Artificial Intelligence Research Institute (IIIA), CSIC, Campus UAB, 08193 Bellatera, Spain*
[b] *Department of Internal Medicine, Reproduction and Population Medicine, Ghent University, 9820 Merelbeke, Belgium*

Abstract. The cumulus-oocyte complex (COC) is an oocyte surrounded by specialized granulosa cells, called cumulus cells. The cumulus cells surrounding the oocyte ensure healthy oocyte and embryo development. The maturity of COCs at oocyte retrieval may be used as an indicator to predict outcome of assisted reproductive technology (ART). Segmenting COCs is a preliminary step in many image processing pipelines to evaluate maturity. However, acquiring well-annotated bright-field microscopy image datasets remains a time-consuming and inaccurate procedure, for most biological domains. Additionally, specialists often partially disagree on their annotations, not only among each other, but also among their own annotations, leading to an inconsistent outcome. Despite the recent advancements in deep learning and image segmentation tools for biological and biomedical images, there is limited usage of them for having more accurate and automated procedures. In this work, we propose an automated pipeline to segment bovine COCs in bright-field microscopy images. The results of our evaluation show that our pipeline is able to segment COCs with the same level of quality as provided by human experts.

Keywords. Deep Learning, Bright-Field Microscopy, Biomedical Imaging, Image Segmentation, Image Analysis

1. Introduction

Infertility is defined as a failure to achieve clinical pregnancy of 12 months or more of regular, unprotected intercourse, and is a big issue for medicine and society. Once the disease is diagnosed, the treatments involve the techniques of Assisted Reproductive Technology (ART). Methods of ART are considered the intracytoplasmic injection of sperm (ICSI) and the in-vitro fertilization (IVF). These methods require several substeps, among of which the characterization of morphological characteristics of oocyte and embryo biology elements.

Cumulus expansion is a key element for characterizing the quality of mammalian oocytes, for later use in in-vitro fertilization (IVF). There are several methods for mea-

suring cumulus expansion described in the literature (Chen et al. [1], Ploutarchou et al. [2]). All the methods available are time-consuming, and depend deeply on human subjectivity since the annotation might vary from one expert to another. Some of the methods for measuring the cumulus expansion rely on assessing the area of cumulus including the oocyte. With the aim to help in the automation of these methods, in this work, we propose a pipeline for segmenting the cumulus oocyte complex (COC), since after segmentation, measuring the size of the COC is a very simple task.
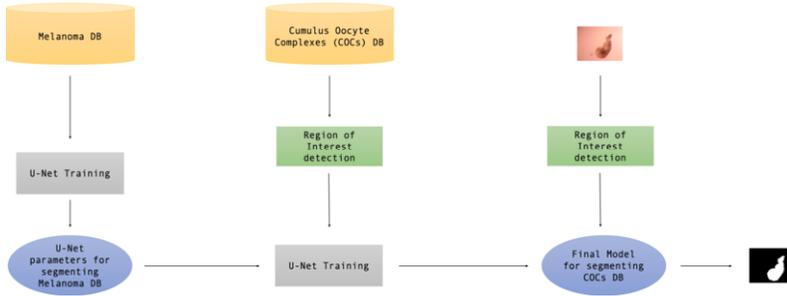
Deep learning and Convolutional Neural Networks (CNN) have seen great progress in the use of medical image segmentation in the recent years, offering a positive impact in medicine and healthcare. Image segmentation is a process of breaking an image into smaller parts, creating a representation more meaningful to be processed by machines. In this work, image segmentation is used to segment bright-field microscopy images of cumulus oocyte complexes in immature and mature oocytes, to later compute the relative cumulus expansion, using a U-Net network architecture.

Literature in image segmentation for oocyte microscopy images is not very broad. Firuzinia et al. [3] applied image segmentation methods on human metaphase II mature oocytes, focusing on several morphological characteristics at this stage, and using a total number of 1009 images. Targosz et al. [4] used image segmentation on human oocytes of different phases (MII, MI, PI, DYS, DEG). A dataset of 334 pictures with one or more oocytes was used. There is no clue for these two approaches that the annotation of the oocytes was performed by more than one specialist. Also, both approaches used already pre-trained networks, such as ResNet and MobileNet, and a variety of data augmentation techniques.

There are other applications of image segmentation for supporting Assisted Reproduction Technology techniques, using bright-field microscopy images. The main focus is on early-stage human embryo development to characterize morphological characteristics. Fukunaga et al. [5] developed a system of automating the detection of pronuclei on 900 embryos. Khan et al. [6] and Leahy et al. [7] applied segmentation techniques for counting the number of cells, while there are works (Dirvanauskas et al. [8], Liu et al. [9], Malmsten et al. [10][11][12], Lau et al. [13], Gingold et al. [14], Meseguer et al. [15]) on identifying the development stage.

To the best of our knowledge, this is the first research on image segmentation in bovine oocytes of bright-field microscopy images. The size of the dataset is just 100 oocytes in total, significantly smaller than any of the already mentioned ones. Last, it is the first approach of trying to measure cumulus expansion, and also exploring the effect of the inconsistency and disagreement among several experts. Our results show that the proposed deep learning model could replace humans in segmenting COCs, and that transfer learning is a key component in the training of our model.

The rest of the paper is organized as follows: in Section 2 we present our segmentation pipeline, with details on the network and the techniques used. In Section 3 we introduce the experiments carried out, with an insight on the dataset followed by the results. Section 4 contains a brief discussion and the conclusions, and we finish the paper with some future perspectives.

Figure 1. Proposed segmentation training architecture.
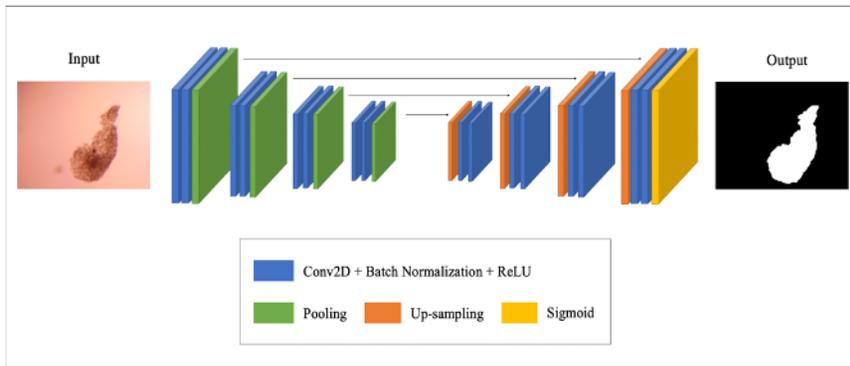
Figure 2. U-Net Architecture

2. Proposal

Our proposed pipeline is presented in Figure 1. The main segmentation model relies on convolutional neural netwoks (in particular, we rely on the U-Net [16] architecture).

The left hand side of Figure 1 shows our usage of *transfer learning* to overcome the lack of training data from the domain. Specifically, in a first stage, U-Net is pre-trained on a publicly available dataset of a related domain, containing bright-field microscopy images of a melanoma cells.

Furthermore, we split the COC segmentation task in two stages. In the first stage, we use local entropy to perform a very rough segmentation and use it to determine a region of interest (ROI), that is, a bounding box containing the COC. The second stage takes as input the image of the ROI and produces a fine segmentation using the U-net model.

2.1. Network Architecture

A U-Net architecture[16] is adopted for all the experiments. U-Net structure for convolutional neural networks have provided satisfying results in the last years for segmenting biomedical image datasets. In Figure 2, there is a representation of the architecture used for our approach.

The contraction path consists of four blocks of two 3x3 convolutional layers, followed by a ReLU layer and a 2x2 max-pooling layer of stride 2, followed by a same

block with an added dropout layer of p = 0.5. The expansive path consists of four blocks of a transposed convolutional layer for up-sampling, a concatenated layer, two 3x3 convolutional layers, a ReLU layer, and afterwards, a last convolutional layer. The proposed architecture was implemented using Keras open-source package and TensorFlow as a back-end platform.

2.2. Loss Functions and Evaluation Metrics

To determine the accuracy of the proposed segmentation we use the Dice Coefficient [17]. Dice coefficient is an indicator of the spatial overlap between two areas, ranging from 0 to 1, with 0 denoting no overlap at all, and 1 denoting perfect overlap. The equation is as follows (1):

$$Dice(f,x,y) = \frac{2\sum_{ij} f(x)_{ij} y_{ij}}{\sum_{ij} f(x)_{ij} + \sum_{ij} y_{ij}} \tag{1}$$

where y is the ground truth, x is the input image, f(x) is the prediction of the model.

Because Dice Coefficient was considered as the evaluation metric, we selected Dice Loss measure to train the weights of the U-Net architecture. Specifically, Dice Loss function can be expressed as the following equation:

$$loss_{Dice}(f,x,y) = 1 - Dice(f,x,y) \tag{2}$$

2.3. Transfer Learning and Data Augmentation

Transfer learning in machine learning is a technique of using knowledge that has been previously acquired from a model, trained to perform a specific task, to a different but somehow related task. The advantage of using this technique is the reduction of the required data size to train a new model, providing a way of building models without requiring large amounts of data, especially in domains that it is highly difficult to find available data, or the labeling of them is time-consuming. In the current approach, annotating the images takes long, and requires specialists with deep knowledge in the domain, while the required annotations are not available in the first place. For the purposes of this application, an open-source dataset of melanoma images is used[1], and then the models are fine-tuned with a small bunch of images and their corresponding annotations.

Data augmentation in Machine Learning is a set of techniques used to increase the amount of available data by creating modified copies of the given data. Since the available COCs dataset is relatively small, data augmentation is used to increase the randomness of the samples, by flipping images along the axes (horizontally, vertically) or rotating them (90, 180, 270 degrees). In that way, for each iteration, there was some percentage of the given images (and corresponding masks) modified, achieving a better generalization of the approach.

---

[1] https://challenge.isic-archive.com/data/
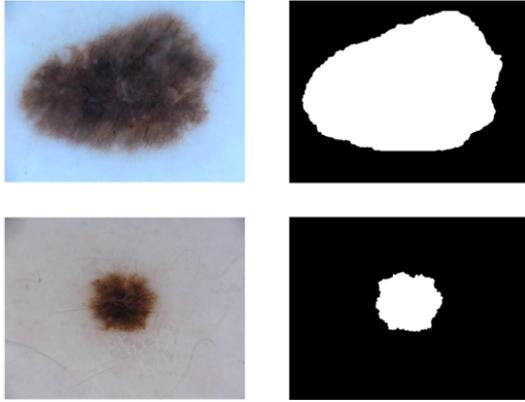
Figure 3. Melanoma dataset sample

3. Experiments

In this section we describe the experimental settings as well as the results of the segmentation pipeline. We start by describing the data used is Section 3.1, then we describe the procedure used for training in Section 3.2, and give the experimental results in Section 3.3.
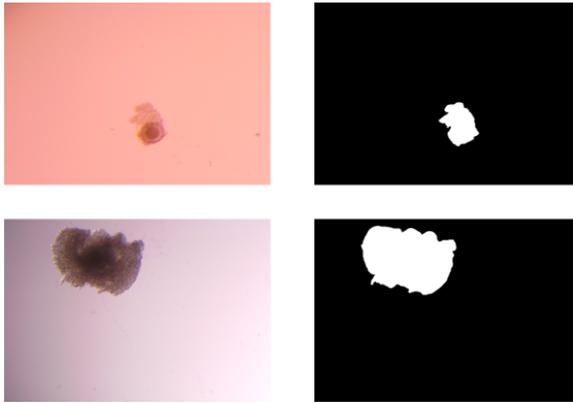
3.1. Datasets

We have used two different datasets: an already existing dataset for pretraining the segmentation model, and a cumulus oocyte dataset which we have created for the task. We describe each of them next.

#### 3.1.1. Pretraining dataset: Melanoma

The dataset was retrieved from the ISIC 2017 Challenge dataset for Skin Lesion Analysis for melanoma detection [18]. It contains 2.000 RGB images manually segmented by medical specialists and forming binary masks for each image (Figure 3). The images and the masks are translated to greyscale and rescaled to 192x240 pixels, before being fit to the CNN, to match its input dimensions.

3.1.2. Cumulus Oocyte Complexes (COCs) dataset

We have created a dataset of bovine cumulus oocyte complexes. It contains images from 100 oocytes. The COCs were incubated for 22 hours, at 38.5 °C , in 5% CO_2 in humidified air [19]. They cultured in tissue culture medium (TCM)-199, supplemented with epidermal growth factor (EGF) and gentamicin, while each oocyte was individually matured in 20μL droplets; briefly, 17 droplets of 20μL medium each were prepared in Petri dishes (60 × 15 mm; Thermo Fisher Scientific, Waltham, MA USA) and covered with 7.5 mL paraffin oil. The microscope used for the pictures was Olympus stereomicroscope, at 56x magnification, using a TOUPCAM UCMOS05100KPA camera and the ImageJ software. The initial size of the images taken were 1944x2592 pixels, and they are all

**Figure 4.** Cumulus Oocyte Complex (COC) dataset sample
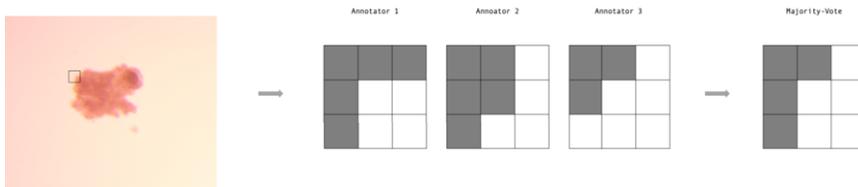
Figure 5. Majority-vote idea.

saved as png. For each of the oocytes we have an immature image (before incubation) and a mature image (after incubation).

We requested three specialists [A1, A2, A3] to segment the images using the ImageJ software, and saving the masks as .png of the same size of the original images. Since the masks provided by each of the three annotators where slightly different we created a consensus segmentation from the three masks. Each pixel of the consensus segmentation was marked as being part of the COC if at least two annotators have marked it as part of the COC in their respective segmentations. For example, if a pixel has 2 positive votes of being part of the COC, and 1 negative, then it is considered as part of the COC. Similarly, we proceed for all of them. The idea is presented in Figure 5 for a random example, where for a 9x9 pixel-square, the majority vote for every pixel is translated to the final output.

For training the model, the previously decided masks were used as ground truth, and from now on, this dataset will be referred to as the majority-vote dataset. The images and the masks were translated to greyscale and to the same ratio as the melanoma dataset, at 192x240, using OpenCV's area interpolation. For the final evaluation, the size of the masks was set back to the initial size (1944x2592), using OpenCV's cubic interpolation. The final results and conclusions remained unaffected by the rescaling process. Other alternatives of combining the information of the experts were considered, such as having probability pixels, instead of deterministic ones, but they were left to be studied in future work.

(a) dice coefficient progress

(b) dice loss progress
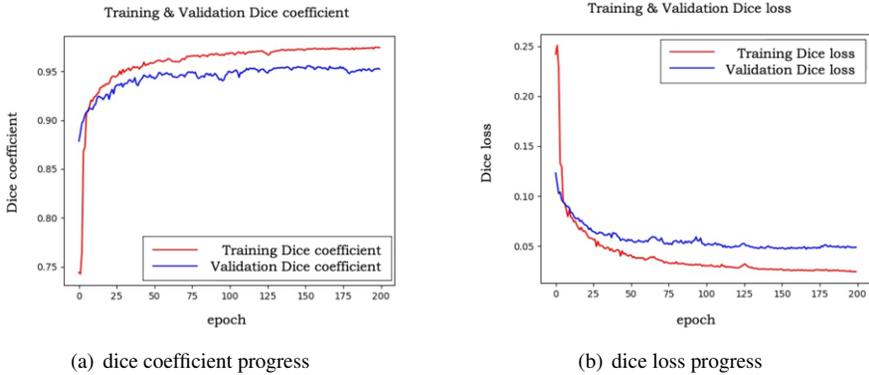
Figure 6. Evolution of mean dice coefficient and mean dice loss.

3.2. Experimental procedure

To evaluate the segmentation models we use 10-fold cross-validation, since the dataset size is limited, and this approach provide safe results. At each fold, 90 oocytes (180 images) were used for training and 10 oocytes (20 images) for validation. The model was trained on minibatches of 32 for 200 epochs. For each fold, we generated the masks for the 20 validation images, resulting in a total of 200 masks after going through all the folds. Then, we compared the masks thus generated, to the masks provided by each of the experts. To evaluate the similarity between two masks we use the dice coefficient. Figure 6 shows the evolution of the mean of dice coefficient and of the mean of the dice loss across all ten folds. The mean dice performance of the majority-vote model is high converging at around 95%.

We are interested in comparing our method against human annotators. To do that we compute the similarity among the annotations of each of the three experts and compare them with the similarity between the each of the human experts and our proposed method. Since dice values do not follow a normal distribution, we use the median of the dice coefficient to evaluate the similarity between any two annotators. This median is expected to be 100% for a perfect similarity and 0% for completely disagreeing annotators.

3.3. Results

The first three rows of Table 1 show how similar are the segmentations of the COC between each pair of human annotators. We see that the numbers are in the 95.15%-95.63% range. In our case, the deep learning model proposed reaches a range between 95.99%-96.48%, higher than the one among the experts. This allows us to consider the results of our model indistinguishable from those of a different human expert. Taking into account the cost of annotating, our method should be considered as a very reasonable alternative to annotation by means of human experts.

To understand the value of each of the components in our pipeline (namely, ROI focusing and transfer learning) we have run an ablation study, removing each of them. The first one concerns the same model, without using the pre-processing stage for detecting the region of interest (ROI), but keeping the transfer learning approach from the

	Annotator 1	Annotator 2	Annotator 3
Annotator 1	-	95.15%	95.49%
Annotator 2	95.15%	-	95.63%
Annotator 3	95.49%	95.63%	-
Our proposal	**96.32%**	**95.99%**	**96.48%**
without ROI	95.95%	95.61%	95.97%
without RT	13.17%	13.55%	14.00%
without TL	no convergence		

Table 1. Comparison of the median of dice coefficients of different models with the ones of human specialists.
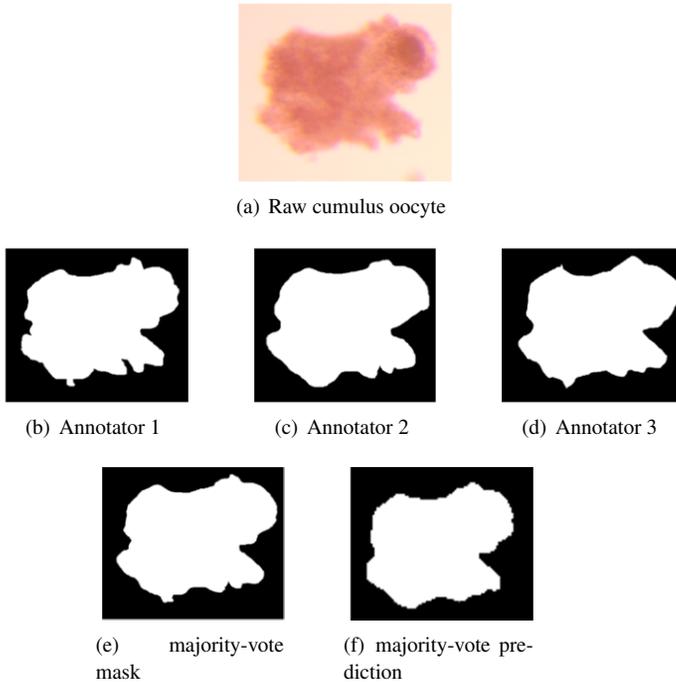
melanoma dataset. The range reached from this configuration (*without ROI* in Table 1) is still very reasonable, being in the same levels of experts' performance (95.61%-95.97%); just a bit lower than the final proposal. On the other hand, when we use the model trained only on the melanoma dataset, without any further retraining (RT) on the COC dataset, the performance of the model only reached values in the range of 13.17%-14.00% (*without RT*). Finally, when we remove transfer learning from our proposal, just starting from random initial weights instead (*without TL* in Table 1), the model failed to converge, concluding that transfer learning is essential for this task, possibly due to our limited access to annotated images.

It is also interesting to provide a visual representation of the region of interest of the segmented images, to better comprehend what is going on the segmented areas (Figure 7). Figure 7(a) shows a cumulus oocyte. In the next row (Figures 7(b), 7(c), 7(d)), show the masks, as were provided by the experts themselves. It is pretty obvious that they are not really coincident, especially around the borders of the oocyte, while it is clear that some of them are more detailed on annotating the perimeter, while others propose a more smooth perimeter. These differences affect the way a model is able to be trained, with a ground truth being controversial. Below, Figure 7(e) shows the mask generated using the majority vote of the three masks above. After that, Figure 7(f), presents the mask generated by using the proposed model. Visually, it is almost identical to the one in Figure 7(e), noting also that the perimeter is smoother than annotator 1's approach, but it tries to keep some important details.

4. Conclusions

In the recent years, there is an increase use of deep learning and image segmentation techniques in Assisted Reproductive Technology field. Some attempts have been made in identifying morphological characteristics from oocyte and embryo bright-field microscopy images, the majority of them in for human species. However, there is limited use in other mammalian species, and no use at all for segmenting bovine cumulus oocyte complexes.

The current research is focusing on segmenting the COCs out of a small-sized dataset. This approach presented a supervised method of detecting the cumulus, using transfer learning of a related domain of melanoma images. The reported dice coefficient of the models proved that the best performing model, using majority-vote annotations for training, is promising, since the scores are identical to the human ones.

(a) Raw cumulus oocyte

(b) Annotator 1   (c) Annotator 2   (d) Annotator 3

(e)   majority-vote
mask

(f) majority-vote pre-
diction

Figure 7. Visual comparison of experts masks and models predictions.

Examining the conditions of the problem in-depth, it became clear that the experts slightly differ when they are annotating or evaluating cumulus oocytes image datasets. Some of them are very detailed and try to be as accurate as possible (Figure 7(b)), without considering the time cost. Others, are less detailed and focus on providing the results faster, leading to smoother, perhaps not so accurate annotations (Figure 7(c), Figure 7(d)). However, trying to find the most beneficial approach, it is rather puzzling to decide and weight more on one of them.

The proposed method of using a majority-vote model, a model that decides if a pixel is part of the cumulus oocyte depending on what the majority of the experts indicates, intents to tackle the issue of partial disagreement among several annotators.

According to the median dice coefficient results, the proposed deep learning model outperforms the human performance, as it is mentioned and presented in Table 1. Noticeably, even with a small-sized dataset and inconsistency among experts of what should be considered as cumulus oocyte part, deep learning algorithms exhibit high and consistent performance, offering more accurate results and a time-saving method.

Acknowledgements

References

[1] Chen L, Russell PT, Larsen WJ. Functional significance of cumulus expansion in the mouse: Roles for the preovulatory synthesis of hyaluronic acid within the cumulus mass. Molecular Reproduction and Development. 1993;34(1):87-93.

[2] Ploutarchou P, Melo P, Day A, Milner C, Williams S. Molecular analysis of the cumulus matrix: Insights from mice with O-glycan-deficient oocytes. Reproduction. 2015 May;149:533-43.

[3] Firuzinia S, Afzali SM, Ghasemian F, Mirroshandel SA. A robust deep learning-based multiclass segmentation method for analyzing human metaphase II oocyte images. Computer Methods and Programs in Biomedicine. 2021 Apr;201:105946.

[4] Targosz A, Przystałka P, Wiaderkiewicz R, Mrugacz G. Semantic segmentation of human oocyte images using deep neural networks. BioMedical Engineering OnLine. 2021 Apr;20(1):40.

[5] Fukunaga N, Sanami S, Kitasaka H, Tsuzuki Y, Watanabe H, Kida Y, et al. Development of an automated two pronuclei detection system on time-lapse embryo images using deep learning techniques. Reproductive Medicine and Biology. 2020;19(3):286-94.

[6] Khan A, Gould S, Salzmann M. Segmentation of developing human embryo in time-lapse microscopy. In: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI); 2016. p. 930-4. ISSN: 1945-8452.

[7] Leahy BD, Jang WD, Yang HY, Struyven R, Wei D, Sun Z, et al. Automated Measurements of Key Morphological Features of Human Embryos for IVF. arXiv:200600067 [cs, q-bio]. 2020 Jul. ArXiv: 2006.00067.

[8] Dirvanauskas D, Maskeliunas R, Raudonis V, Damasevicius R. Embryo development stage prediction algorithm for automated time lapse incubators. Computer Methods and Programs in Biomedicine. 2019 Aug;177:161-74.

[9] Liu Z, Huang B, Cui Y, Xu Y, Zhang B, Zhu L, et al. Multi-Task Deep Learning With Dynamic Programming for Embryo Early Development Stage Classification From Time-Lapse Videos. IEEE Access. 2019;7:122153-63. Conference Name: IEEE Access.

[10] Malmsten J, Zaninovic N, Zhan Q, Toschi M, Rosenwaks Z, Shan J. Automatic prediction of embryo cell stages using artificial intelligence convolutional neural network. Fertility and Sterility. 2018 Sep;110(4):e360. Publisher: Elsevier.

[11] Malmsten J, Zaninovic N, Zhan Q, Rosenwaks Z, Shan J. Automated cell stage predictions in early mouse and human embryos using convolutional neural networks. In: 2019 IEEE EMBS International Conference on Biomedical Health Informatics (BHI); 2019. p. 1-4. ISSN: 2641-3604.

[12] Malmsten J, Zaninovic N, Zhan Q, Rosenwaks Z, Shan J. Automated cell division classification in early mouse and human embryos using convolutional neural networks. Neural Computing and Applications. 2020 Jun.

[13] Lau T, Ng N, Gingold J, Desai N, McAuley J, Lipton ZC. Embryo staging with weakly-supervised region selection and dynamically-decoded predictions. arXiv:190404419 [cs]. 2019 Apr. ArXiv: 1904.04419.

[14] Gingold JA, Ng NH, McAuley J, Lipton Z, Desai N. Predicting embryo morphokinetic annotations from time-lapse videos using convolutional neural networks. Fertility and Sterility. 2018 Sep;110(4):e220. Publisher: Elsevier.

[15] Meseguer M, Herrero J, Tejera A, Hilligsøe KM, Ramsing NB, Remohí J. The use of morphokinetics as a predictor of embryo implantation. Human Reproduction (Oxford, England). 2011 Oct;26(10):2658-71.

[16] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv:150504597 [cs]. 2015 May. ArXiv: 1505.04597.

[17] Dice LR. Measures of the Amount of Ecologic Association Between Species. Ecology. 1945;26(3):297-302.

[18] Gutman D, Codella NCF, Celebi E, Helba B, Marchetti M, Mishra N, et al. Skin Lesion Analysis toward Melanoma Detection: A Challenge at the International Symposium on Biomedical Imaging (ISBI) 2016, hosted by the International Skin Imaging Collaboration (ISIC). arXiv:160501397 [cs]. 2016 May. ArXiv: 1605.01397.

[19] Azari-Dolatabad N, Raes A, Pavani KC, Asaadi A, Angel-Velez D, Van Damme P, et al. Follicular fluid during individual oocyte maturation enhances cumulus expansion and improves embryo development and quality in a dose-specific manner. Theriogenology. 2021 May;166:38-45.