



Improving Deep Learning-Based Algorithm for Ploidy Status Prediction Through Combined U-NET Blastocyst Segmentation and Sequential Time-Lapse Blastocysts Images

Nining Handayani^{1, 2}, Gunawan Bondan Danardono², Arief Boediono^{2, 3, 4}, Budi Wiweko^{5, 6, 7}, Ivan Sini^{2, 3}, Batara Sirait^{2, 3, 8}, Arie A Polim⁹, Irham Suheimi^{2, 3}, Anom Bowolaksono^{10*}

1- Doctoral Program in Biomedical Sciences, Faculty of Medicine, Universitas Indonesia, Jakarta, Indonesia

2- IRSI Research and Training Centre, Jakarta, Indonesia

3- Morula IVF Jakarta Clinic, Jakarta, Indonesia

4- Department of Anatomy, Physiology and Pharmacology, IPB University, Bogor, Indonesia

5- Division of Reproductive Endocrinology and Infertility, Department of Obstetrics and Gynecology, Faculty of Medicine, Universitas Indonesia, Jakarta, Indonesia

6- Yasmin IVF Clinic, Dr Cipto Mangunkusumo General Hospital, Jakarta, Indonesia

7- Human Reproduction, Infertility, and Family Planning Cluster, Indonesia Reproductive Medicine Research and Training Center, Faculty of Medicine, Universitas Indonesia, Jakarta, Indonesia

8- Department of Obstetrics and Gynaecology, Faculty of Medicine, Universitas Kristen Indonesia, Jakarta, Indonesia

9- Department of Obstetrics and Gynecology, School of Medicine and Health Sciences, Atma Jaya Catholic University of Indonesia, Jakarta, Indonesia

10- Cellular and Molecular Mechanisms in Biological System (CEMBIOS) Research Group, Department of Biology, Faculty of Mathematics and Natural Sciences, Universitas Indonesia, Depok, Indonesia

Abstract

Background: Several approaches have been proposed to optimize the construction of an artificial intelligence-based model for assessing ploidy status. These encompass the investigation of algorithms, refining image segmentation techniques, and discerning essential patterns throughout embryonic development. The purpose of the current study was to evaluate the effectiveness of using U-NET architecture for embryo segmentation and time-lapse embryo image sequence extraction, three and ten *hr* before biopsy to improve model accuracy for prediction of embryonic ploidy status.

Methods: A total of 1.020 time-lapse videos of blastocysts with known ploidy status were used to construct a convolutional neural network (CNN)-based model for ploidy detection. Sequential images of each blastocyst were extracted from the time-lapse videos over a period of three and ten *hr* prior to the biopsy, generating 31.642 and 99.324 blastocyst images, respectively. U-NET architecture was applied for blastocyst image segmentation before its implementation in CNN-based model development.

Results: The accuracy of ploidy prediction model without applying the U-NET segmented sequential embryo images was 0.59 and 0.63 over a period of three and ten *hr* before biopsy, respectively. Improved model accuracy of 0.61 and 0.66 was achieved, respectively with the implementation of U-NET architecture for embryo segmentation on the current model. Extracting blastocyst images over a 10 *hr* period yields higher accuracy compared to a three-*hr* extraction period prior to biopsy.

Conclusion: Combined implementation of U-NET architecture for blastocyst image segmentation and the sequential compilation of ten *hr* of time-lapse blastocyst images could yield a CNN-based model with improved accuracy in predicting ploidy status.

Keywords: Artificial intelligence, Image processing, Neural network, Ploidy measurement.

To cite this article: Handayani N, Danardono GB, Boediono A, Wiweko B, Sini I, Sirait B, et al. Improving Deep Learning-Based Algorithm for Ploidy Status Prediction Through Combined U-NET Blastocyst Segmentation and Sequential Time-Lapse Blastocysts Images. *J Reprod Infertil.* 2024;25(2):110-119. <https://doi.org/10.18502/jri.v25i2.16006>.

Introduction

Remarkable efforts have been put forward to discover prominent non-invasive molecular biomarkers for determination of embryo

ploidy status. Non-invasive assessments of embryo ploidy status are desirable to replace the need for performing embryo biopsy as a standard

* Corresponding Author:
Anom Bowolaksono,
Cellular and Molecular
Mechanisms in Biological
System (CEMBIOS)
Research Group,
Department of Biology,
Faculty of Mathematics and
Natural Sciences,
Universitas Indonesia,
Depok, Indonesia
E-mail:
alaksono@sci.ui.ac.id

Received: 18, Dec. 2023

Accepted: 15, May 2024

sample collection procedure for chromosomal analysis in preimplantation genetic testing for aneuploidy (PGT-A). Current attempts at non-invasive preimplantation genetic testing for aneuploidy (niPGT-A) through biomarkers identified in spent embryo culture media have garnered the attention of several IVF experts (1-3). Correspondence in outcomes between the two methods (PGT-A and niPGT-A) has improved over time but its implementation for routine PGT testing in IVF remains under further investigation. Concurrently, embryo morphokinetic parameters (4-8), specific metabolomics (9), proteomics (10), and artificial intelligence-based image analysis (11-14) have exhibited clinical values that could potentially define embryo ploidy without invasive interventions. Among the aforementioned approaches, developing an AI-based model, which could predict IVF outcomes, is at the top of the list as the most imminent niPGT-A approach.

Widespread utilization of embryo images to develop AI-based prediction models has been observed and even commercialized products are currently available (15). In the broad sense of AI, blastocyst images may contain essential information that could not be comprehended by the naked human eye. Through embryo image processing, the availability of such information, which may encompass intelligence on embryo viability (13, 16) or ploidy status classification (17, 18), is explored. Different types of inputs (trained images) have been utilized in the current literature to develop AI models that predict specific outcomes. Huang et al. (17) have developed an AI platform using 10 *hr* of blastocyst expansion images (generating ± 30 sequential images consecutively) extracted from a time-lapse video with the output being embryo ranking for further use. A 2019 study conducted by Tran et al. (19) used raw time-lapse videos (approximately 112 *hr* of culture) as an input to train a deep learning system that can predict implantation potential. However, other research groups have utilized static images captured by a standard light optical microscope (13, 16) or a combination of static images from a light optical microscope and captured time-lapse images to generalize the applicability of the model considering that time-lapse incubator is not widely adopted worldwide (20). In addition to the diverse inputs, there are also variations in the methods used for AI development (1, 11-13). Both machine learning (ML) and deep learning (DL) algorithms were used for image processing. While

image processing in ML sometimes uses additional algorithms such as genetic algorithm (20), a sub-class of DL, convolutional neural network (CNN) is the prominent algorithm that can perform a complex task.

The purpose of the current study was to construct a deep learning-based model for ploidy status prediction of human blastocysts by utilizing U-NET segmentation as well as three-and ten-*hr* sequential time-lapse embryo images before the commencement of blastocyst biopsy. This strategy was performed to confirm whether the two different culture periods contain any useful information that could boost the accuracy of the CNN model, considering embryo development is highly dynamic, particularly when approaching the implantation process.

Methods

Patient population, data collection, and pre-processing images: This was a single-center cohort study. A total of 425 couples who underwent PGT-A were identified in the private online database of Morula IVF Jakarta Clinic, Jakarta, Indonesia between January 2021 and October 2022. This study protocol was reviewed and approved by the Ethical Committee of Universitas Indonesia with approval number of KET-74/UN2.F1/ETIK/PPM.00.02/2022.

The indications of studied subjects were recurrent IVF failure following the transfer of top-quality embryo(s), having a history of recurrent miscarriages, and advanced maternal age. These align with the clinic's policy regarding the recommendation of PGT-A for infertile couples. Infertile couples undergoing PGT-A were excluded from the analysis if their PGT-A sample failed to pass quality control, required re-biopsy, or yielded undetermined results. Ovarian stimulation and embryo culture procedures were conducted as previously described (21). Briefly, the embryo was cultured using a time-lapse incubator (Miri TL; Esco Medical, Denmark) immediately following ICSI or IMSI procedure, under a culture condition of 37°C, 6% CO₂, and 5% O₂. Throughout the study period, either G-TL (Vitrolife, Sweden) or SAGE (Origio, Denmark) was utilized. Blastocyst quality was measured according to the Gardner Grading System by measuring the quality of inner cell mass (ICM), trophectoderm, and the expansion of the blastocyst cavity. Top-quality blastocysts were defined as grades based on blastocoel cavity expansion and AA, AB, and BA according

to the quality of ICM and trophectoderm (22). On day 4 of embryo culture, three pulses of laser (OCTAX Laser Shot™) were shot at the zona pellucida to facilitate herniation of trophectoderm cells. Biopsy procedures were conducted on either day 5 or day 6 depending on the blastocyst expansion. Up to 2-5 trophectoderm cells were collected using a specific pipette (blastomere aspiration pipette; COOK, Ireland). After washing the biopsied embryonic cells in PBS medium supplemented with 1% polyvinylpyrrolidone (Origio, Denmark), samples were then loaded into a 0.2 ml PCR tube (Gen-Follower, China) and sent to the genomic laboratory for ploidy analysis.

Altogether, 1,020 blastocysts were biopsied for PGT-A. Next-generation sequencing (MiSeq Sequencing System; Illumina, USA) was utilized for determining the ploidy status, serving as the ground truth of the dataset. VeriSeq PGS kit was used following the VariSeq PGS Library Prep reference guide (15052877 v04). The PGT-A procedure was conducted following a detailed procedure as previously reported (23). Ploidy analysis was performed using Blue-Fuse software (Illumina, USA), which generated three types of outcomes: euploid (a mixture of euploid and <30% aneuploid cells), aneuploid (mosaicism with more than 80% aneuploid cells), and mosaic (a mixture of euploid and 30-80% aneuploid cells, with low-level mosaicism defined as 30-50% aneuploid cells while the remaining cells were categorized as high-level mosaicism) (24, 25).

Recorded time-lapse videos of 1,020 blastocysts with known ploidy status were retrieved from time-lapse incubators. This study only utilized blastocyst images captured from the TL videos as input for the CNN-based model. The image extraction process involved a combined effort, aligning raw tabular data with time-lapse video files utilizing Python scripting based on their metadata information. Sequential blastocyst images for specific three- and ten-hr periods preceding the biopsy procedure were extracted and summarized in table 1. The decision to extract blastocyst images 10 hr prior to biopsy was based on prior findings, which emphasized the increased importance of data related to blastocyst formation compared to earlier pre-implantation stages (26, 27).

Additionally, an attempt was made to explore whether blastocyst expansion patterns observed 3 hr prior to biopsy could be sufficient and effectively utilized in developing a predictive model. Extracted images with an obscured focus of the

Table 1. Image dataset utilized in developing CNN prediction models

Known ploidy status	TL video	Total images extracted	
		Three hr	Ten hr
Euploid blastocysts	181	5.659	17.772
Aneuploid blastocysts	390	12.094	37.915
Mosaic blastocysts	449	13.889	43.637

Note: from 181 TL videos of euploid blastocysts, 5,659 images were captured within 3 hr prior to biopsy and 17,772 images obtained within 10 hr before biopsy. For the 390 TL videos of aneuploid blastocysts, a total of 12,094 images were taken 3 hr preceding the biopsy, which increased to 37,915 images within the 10-hr timeframe prior to biopsy. Similarly, in the case of 449 TL videos of mosaic blastocysts, the total number of images extracted was higher within 10 hr (43,637 images) compared to the 3-hr timeframe before the biopsy

blastocysts and those with embryos misaligned from the capture of the TL’s internal camera were excluded. The AI environment was established on a Windows operating system, using an Intel CPU and NVIDIA GPU. Python was the programming language for script management and the TensorFlow library was utilized for CNN classification (28) and to build a U-NET image segmentation model. Tabular data was handled using Pandas and NumPy libraries, and partial data preparation was conducted in Microsoft Excel. Image augmentation was performed using the OpenCV library, as depicted in figure 1.

U-NET image segmentation: A fully convolutional

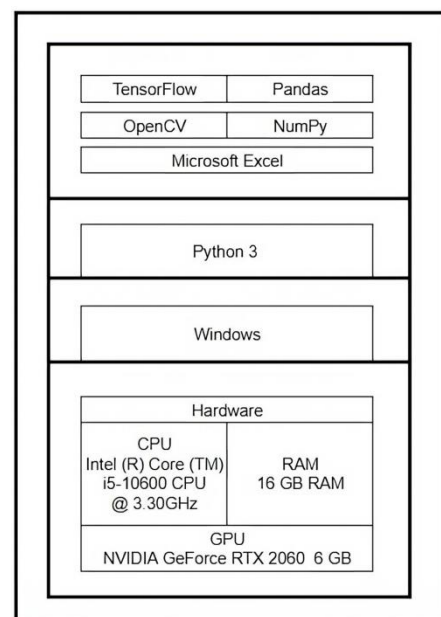


Figure 1. Architecture diagram of environment used for developing the model

neural network architecture, U-NET, was used for blastocyst image segmentation in the present study as it could easily be updated and trained using a limited dataset. The architecture was inspired by Ronneberger et al. (29), consisting of an encoder and a decoder. However, a few modifications were made to the original structure to tailor it to the unique demands of our image dataset. The encoder area processes the input images to learn and identify structures of the whole blastocyst through convolution, dropout, and max pooling methods. Briefly, each pixel on the raw images was assigned to the groups that belong to a specific part within the image. The decoding area mapped the position of blastocysts through convolution, up-sampling, and skip connections. Specifically, the positions were determined by concatenating or joining the encoder parts with the decoder part in an end-to-end fashion.

Conceptually, the U-NET model is built using convolution, max pooling, dropout, up-sampling, and concatenate layers with each layer complementing the encoder and decoder part. U-NET model is named after its architectural shape (Figure 2). The convolution layer serves as an encoder function responsible for image feature extraction, through which a filter is applied to create a feature map of the input image. The size of the filter and feature map can be specified based on the following sequence of layers. The max pooling layer is responsible for selecting the maximum value from the prominent feature map for every pool or group, thereby enhancing the sharpness of features. On the other hand, the dropout layer temporarily reduces the number of features within a node, mitigating the risk of overfitting or underfitting the model. The application of the dropout layer will result in slight changes during each iteration of model training. Up-sampling layer adjusts the layer dimension to an appropriate node size. However, up-sampling cannot recover any lost information during the process. Lastly, to combine and merge two different nodes into one single node, a concatenate layer is utilized. To visualize, the concatenate layer acts as a bridge that combines two different nodes into one. Eventually, targeted blastocysts could be masked as an output for CNN model training (Figure 3).

Training and validation of the CNN-based model: Our research utilized supervised learning throughout the model-building process to avoid any misfits of wrong segmentation or the possibility of

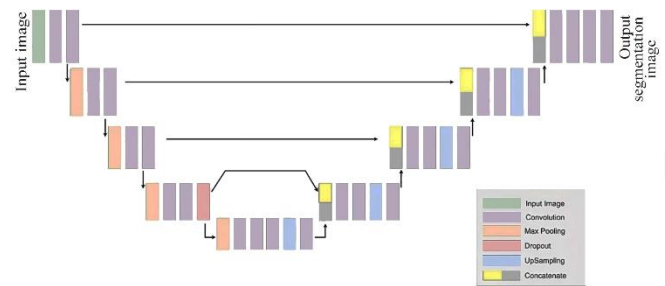


Figure 2. Illustration of U-NET architecture applied in the current study

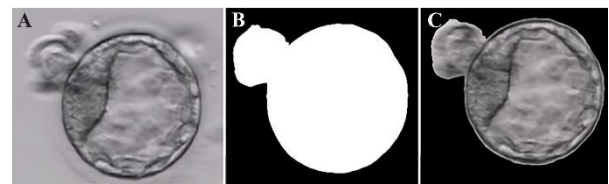


Figure 3. Blastocyst's image extracted from TL video, A) raw image, B) image masking from U-NET model, C) segmented blastocysts image utilizing U-Net model

data cluster misplacements and to reduce the likelihood of a systematic error. The non-segmented and segmented images, captured at three- and ten-hour intervals, respectively were tested individually with 80-20 data split to overview the capability of the deep learning approach in predicting ploidy status based on morphological uniqueness and differentiating features of the embryos, which will be elaborated with a pre-trained CNN model. The ratio was determined to minimize the risk of overestimating measurement error (30).

The CNN training environment was set to a similar condition with its pre-trained model as the differentiator between each training sequence. The three pre-trained CNN models include ResNet (31), InceptionV3 (32), and EfficientNet (33) which are appropriately selected due to their wide applications in the deep learning field (34-36). Each pre-trained model holds various combinations of so-called nodes or layers, and certain layers are designed with features that facilitate efficient model construction. For instance, batch normalization standardizes input layers with a mathematical approach based on the activation condition. Moreover, each pre-trained model was built uniquely and hence possessed different input conditions or image sizes. Despite the relatively small size of an embryo, significant physical information is encapsulated within its image state. Image size or pixels thus indicate meaningful details

Table 2. The pre-trained model input layer utilized for prediction model development

Model	(Batch normalization/ total layer)-%	Input image size
EfficientNetB6	(134/752) 17.82%	528
ResNet50V2	(49/192) 25.52%	224
InceptionV3	(94/313) 30.03%	299

Note: The unique layers or nodes of each pre-trained model possess distinct characteristics, and each layer harbors attributes essential for constructing models efficiently. The batch normalization layer, which optimizes the output classification process, ensures that each model maintains a unique image input size.

that correspond to high-performance computational resources.

Furthermore, the model performance is influenced by a combination of image size and feature extraction methods or the pre-trained model selections. Higher image size and layers, however, do not guarantee a more robust model performance. On the contrary, processing larger images with more pixels would require a greater amount of computing power compared to processing smaller images (37). Table 2 shows several pre-trained models with their special input conditions.

Matrix evaluation: Matrix evaluation has become an important part of the model training algorithm as it impacts the model performance in conducting certain prediction tasks. CNN could automatically learn from its previous training iteration and calibrate its layers to an appropriate condition. However, such automation is not dynamic enough to be refined because the natural process of CNN is somewhat hidden and challenging to disclose. Consolidation of several matrix evaluations, therefore, becomes the standard for an AI-based classifier or predictor.

Our approach to obtaining the most robust AI-based model is through the adoption of two matrix evaluations, namely Accuracy and Loss. Individual pre-trained models were assessed based on the accuracy and loss matrix. Accuracy represents the ratio of true positive and true negative values to the total number of classification cases, signifying the initial overall performance of the model. Accuracy has been defined as a standard evaluation metric due to its simplicity and ability to encompass all classification outcomes whether true or false predictions. Loss matrix calculates the confidence level of a model in creating a prediction. A low loss matrix indicates a model with a high confidence value in performing the classification

and vice versa. Specifically, the loss function plays a crucial role in evaluating whether a model requires an update in its capability to predict the targeted outcomes.

Results

Extraction of sequential TL blastocyst images over a period of three and ten *hr* prior to the biopsy procedure yielded around 31 and 97 unique images per TL video, respectively. From a pool of 181 euploid blastocysts TL videos, 5,659 images were captured within the 3 *hr* preceding biopsy, and 17,772 images were obtained within a 10-*hr* period before biopsy. Of 390 aneuploid blastocysts TL videos, a total of 12,094 images were gathered from the 3-*hr* period prior to the biopsy, which increased to 37,915 images within the 10-*hr* timeframe preceding biopsy. Among 449 TL videos of mosaic blastocysts, the total number of extracted images within the 3 and 10-*hr* window was 13,889 and 43,637 images, respectively (Table 1). Slight disparities between the sequence of embryo images have culminated in a more complex model assignment for classification. Consequently, using multiple time point images leads to a more accurate result compared to a single time point image classification.

In the current study, every pre-trained model employed possessed distinct layers, each characterized by its unique features as depicted in table 2. The image input sizes for EfficientNetB6, ResNet50V2, and InceptionV3 were 528, 224, and 299, respectively.

A comparison of the two-time image extraction and classification pathway between U-NET segmented and non-segmented embryo images is illustrated in table 3. Integration of U-NET image segmentation significantly boosted model accuracy in both time points. With ResNet50V2 algorithm, accuracy elevated from 0.59 to 0.61, while with the InceptionV3 algorithm, it surged from 0.63 to 0.66. The highest accuracy of 0.66 was attained when employing the ten-*hr* image series alongside U-NET image segmentation for model prediction development using the InceptionV3 algorithm. Interestingly, our model exhibited a unique trend of slightly higher accuracy when employing non-segmented InceptionV3 and EfficientNetB6 algorithms with three *hr* of data.

Discussion

Key findings: This study presented a subset of deep learning algorithms known as convolutional

Table 3. Comparison of model predictions with and without U-Net image segmentation using three-and ten-*hr* image series prior to biopsy

Model	Accuracy		Loss	
	Without segmentation	With U-NET segmentation	Without segmentation	With U-NET segmentation
Image series of three <i>hr</i>				
EfficientNetB6	0.6055	0.5634	0.8456	0.9276
ResNet50V2	0.5927	0.6181	1.3509	1.265
InceptionV3	0.5823	0.5751	0.9820	0.8924
Image series of ten <i>hr</i>				
EfficientNetB6	0.6152	0.6484	1.2727	1.314
ResNet50V2	0.6435	0.6562	1.3974	1.3026
InceptionV3	0.6357	0.6699	1.456	1.4022

Note: in both the three-*hr* and ten-*hr* image series, incorporating U-NET image segmentation enhanced model accuracy. For the ResNet50V2 algorithm, accuracy improved from 0.59 to 0.61, while for the InceptionV3 algorithm, it increased from 0.63 to 0.66. The highest accuracy of 0.66 was achieved when utilizing the ten-*hr* image series with U-NET image segmentation applied for model prediction development.

neural network (CNN) to predict blastocyst ploidy status. The utilization of U-NET architecture for image segmentation resulted in a model with higher accuracy compared to using raw images without U-NET segmentation. U-NET image segmentation was proposed to enhance the capability of model classification. U-NET segmentation enables an automated process of isolating embryos from the culture dish and any unnecessary image information, allowing the prediction model to focus on prominent embryo images. The present study also serves to strengthen image recognition-based artificial intelligence in the field of IVF. In this study, an AI-based image classification was conducted through a deep learning procedure.

This study demonstrated two case study comparisons using datasets of different durations: three *hr* and ten *hr*. Increased model accuracy was observed when sequential images from a ten-*hr* embryo culture period (prior to biopsy) were extracted. From an AI perspective, these results proved that a higher amount of training data indeed coincides with improved model performance. In addition, dynamic development of blastocyst, particularly during the expansion phase, may contain meaningful information, thus leading to enhanced performance of the AI model as previously suggested (17). Our model demonstrated a trend of slightly better accuracy when utilizing non-segmented InceptionV3 and EfficientNetB6 models with three *hr* of datasets. However, it is challenging to explicitly elucidate this trend due to the uti-

lization of deep learning methodology, frequently referred to as a "black box" in research.

Artificial neural network (ANN) nodes, mimicking how the human brain works, have the ability to effectively learn specific patterns in the given image. As the ANN nodes are organized hierarchically, each node calculates the weighted sum of the given input by applying a specific activation function to the sum component, which acts as a receiver of weighted input. As a result, ANN can produce the final model that can differentiate the targeted outcome (38). Multilayer perceptron (MLP) is the most common type of ANN architecture and is also a popular foundation of CNN architecture. In general, each neural layer in MLP comprises an input layer, one or several hidden layers, and an output layer. CNN is similar to MLP with regard to using calculated weights in each node and receiving several different inputs as a sum to classify the outcome. Nonetheless, CNN contains multiple MLPs with a high number of neural layers and nodes. In addition, CNN algorithm employs convolution mathematical operations which become an essential part of computation in CNN architecture. Briefly, CNN architecture comprises several extraction phases and fully connected layers that can map specific patterns to classify the outcomes (39). While CNN demonstrates remarkable ability in image classification, the main limitation of CNN is its fully automated differentiation process, which cannot be understood by human logic. This characteristic is often

referred to as a "black box". Hence, many IVF experts have argued if the classification is trustworthy.

Comparison with previous studies: Notably, several studies that utilize embryo images for ploidy prediction are reported in the current literature. A 2020 study conducted by Chavez-Badiola et al. (11) is known to be the first that used static images for ploidy status classification. A total of 751 embryo images with known ploidy status were used to construct a ploidy prediction algorithm called the ERICA model through deep learning neural network. The model attained an accuracy of 0.70 in model validation and testing. In 2021, two studies reported the use of raw time-lapse videos as input for ploidy prediction model development. Lee et al. (12) utilized sequential images of embryos captured from time-lapse video and used a deep learning model, Inflated 3D ConvNet (I3D) which obtained high accuracy of 0.74. In contrast, Huang et al. (40) chose to combine the time-lapse embryo videos with the clinical characteristics of the studied participant into the model and achieved an accuracy of 0.8. An interesting study conducted by Huang et al. (17) showed that the average time of blastocyst expansion could be a better predictor for blastocyst ploidy classification. Using U-NET architecture for semantic segmentation, an AI-based approach called AI-qSEA1.0 expansion assay was implemented to rank the quality of blastocysts for clinical use. The researchers then retrospectively analyzed the outcomes of the respective cycles and observed that euploid blastocysts that resulted in live births had a higher expansion rate than those of euploid blastocysts that did not result in live births ($p=0.007$). As AI-based studies varied highly in terms of the inputs, algorithms or classifiers used, and outcomes, the results cannot be compared (41). Each constructed model is distinct, primarily due to the utilization of different datasets and research designs. In clinical practice, it is expected that the availability of a non-invasive AI-based algorithm could serve as an alternate method for embryo selection in cases in which PTG-A is unaffordable.

The strength of the present study lies in the demonstration of U-NET implementation for blastocyst segmentation which was proven to enhance model training for image classification with the deep learning method. Nonetheless, the limitation of this study is that the blastocyst segmentation

does not differentiate the distinct parts of the blastocyst such as the inner cell mass, trophoctoderm area, blastocoel cavity, or zona pellucida thickness. This could serve as a solution to improve prediction accuracy. Additionally, the accuracy of the obtained model is not sufficient for being used as a non-invasive approach for predicting blastocyst ploidy status in clinical settings. Also, an imbalance was observed in the training step results for both cases. This noteworthy discovery highlights the presence of bias in the image classification model for predicting embryo ploidy status.

Conclusion

This study demonstrated that extracting TL blastocyst images over a ten-*hr* period and implementing image segmentation prior to utilizing embryo images in a CNN-based model could enhance the accuracy of the developed model for predicting ploidy status.

Acknowledgement

The authors express gratitude toward the staff of Morula IVF Jakarta Clinic for providing and allowing the authors to use the time-lapse video for developing a ploidy status prediction model.

Funding: The authors received a PUTI Grant from Universitas Indonesia with grant number NKB-1293/UN2.RST/HKP.05.00/2022.

Conflict of Interest

The authors have no conflicts of interest to declare.

References

- Huang L, Bogale B, Tang Y, Lu S, Xie XS, Racowsky C. Noninvasive preimplantation genetic testing for aneuploidy in spent medium may be more reliable than trophoctoderm biopsy. *Proc Natl Acad Sci USA*. 2019;116(28):14105-12.
- Rubio C, Navarro-Sánchez L, García-Pascual CM, Ocali O, Cimadomo D, Venier W, et al. Multicenter prospective study of concordance between embryonic cell-free DNA and trophoctoderm biopsies from 1301 human blastocysts. *Am J Obstet Gynecol*. 2020;223(5):751.e1-751.e13.
- Feichtinger M, Vaccari E, Carli L, Wallner E, Mädler U, Figl K, et al. Non-invasive preimplantation genetic screening using array comparative genomic hybridization on spent culture media: a proof-of-concept pilot study. *Reprod Biomed Online*. 2017;34(6):583-9.
- Fishel S, Campbell A, Montgomery S, Smith R,

- Nice L, Duffy S, et al. Time-lapse imaging algorithms rank human preimplantation embryos according to the probability of live birth. *Reprod Biomed Online*. 2018;37(3):304-13.
5. Campbell A, Fishel S, Bowman N, Duffy S, Sedler M, Thornton S. Retrospective analysis of outcomes after IVF using an aneuploidy risk model derived from time-lapse imaging without PGS. *Reprod Biomed Online*. 2013;27(2):140-6.
 6. Reignier A, Lammers J, Barriere P, Freour T. Can time-lapse parameters predict embryo ploidy? A systematic review. *Reprod Biomed Online*. 2018;36(4):380-7.
 7. Pennetta F, Lagalla C, Borini A. Embryo morphokinetic characteristics and euploidy. *Curr Opin Obstet Gynecol*. 2018;30(3):185-96.
 8. Kramer YG, Kofinas JD, Melzer K, Noyes N, McCaffrey C, Buldo-Licciardi J, et al. Assessing morphokinetic parameters via time lapse microscopy (TLM) to predict euploidy: are aneuploidy risk classification models universal? *J Assist Reprod Genet*. 2014;31(9):1231-42.
 9. Liang B, Gao Y, Xu J, Song Y, Xuan L, Shi T, et al. Raman profiling of embryo culture medium to identify aneuploid and euploid embryos. *Fertil Steril*. 2019;111(4):753-62.e1.
 10. Bori L, Dominguez F, Fernandez EI, Del Gallego R, Alegre L, Hickman C, et al. An artificial intelligence model based on the proteomic profile of euploid embryos and blastocyst morphology: a preliminary study. *Reprod Biomed Online*. 2021;42(2):340-50.
 11. Chavez-Badiola A, Flores-Saiffe-Farías A, Mendizabal-Ruiz G, Drakeley AJ, Cohen J. Embryo ranking intelligent classification algorithm (ERICA): artificial intelligence clinical assistant predicting embryo ploidy and implantation. *Reprod Biomed Online*. 2020;41(4):585-93.
 12. Lee CI, Su YR, Chen CH, Chang TA, Kuo EES, Zheng WL, et al. End-to-end deep learning for recognition of ploidy status using time-lapse videos. *J Assist Reprod Genet*. 2021;38(7):1655-63.
 13. Diakiw SM, Hall JMM, VerMilyea MD, Amin J, Aizpurua J, Giardini L, et al. Development of an artificial intelligence model for predicting the likelihood of human embryo euploidy based on blastocyst images from multiple imaging systems during IVF. *Hum Reprod*. 2022;37(8):1746-59.
 14. Huang TT, Huang DH, Ahn HJ, Arnett C, Huang CT. Early blastocyst expansion in euploid and aneuploid human embryos: evidence for a non-invasive and quantitative marker for embryo selection. *Reprod Biomed Online*. 2019;39(1):27-39.
 15. Louis CM, Handayani N, Aprilliana T, Polim AA, Boediono A, Sini I. Genetic algorithm-assisted machine learning for clinical pregnancy prediction in in vitro fertilization. *AJOG Glob Rep*. 2023;3(1):100133.
 16. VerMilyea M, Hall JMM, Diakiw SM, Johnston A, Nguyen T, Perugini D, et al. Development of an artificial intelligence-based assessment model for prediction of embryo viability using static images captured by optical light microscopy during IVF. *Hum Reprod*. 2020;35(4):770-84.
 17. Huang TTF, Kosasa T, Walker B, Arnett C, Huang CTF, Yin C, et al. Deep learning neural network analysis of human blastocyst expansion from time-lapse image files. *Reprod Biomed Online*. 2021;42(6):1075-85.
 18. Danardon GB, Handayani N, Louis CM, Polim AA, Sirait B, Periastringrum G, et al. Embryo ploidy status classification through computer-assisted morphology assessment. *AJOG Glob Rep*. 2023;3(3):100209.
 19. Tran D, Cooke S, Illingworth PJ, Gardner DK. Deep learning as a predictive tool for fetal heart pregnancy following time-lapse incubation and blastocyst transfer. *Hum Reprod*. 2019;34(6):1011-8.
 20. Louis CM, Handayani N, Aprilliana T, Polim AA, Boediono A, Sini I. Genetic algorithm-assisted machine learning for clinical pregnancy prediction in in vitro fertilization. *AJOG Glob Rep*. 2022;3(1):100133.
 21. Boediono A, Handayani N, Sari HN, Yusup N, Indrasari W, Polim AA, et al. Morphokinetics of embryos after IMSI versus ICSI in couples with sub-optimal sperm quality: a time-lapse study. *Andrologia*. 2021;53(4):e14002.
 22. Gardner DK, Lane M, Stevens J, Schlenker T, Schoolcraft WB. Blastocyst score affects implantation and pregnancy outcome: Towards a single blastocyst transfer. *Fertil Steril*. 2000;73(6):1155-8.
 23. Polim AA, Handayani N, Nurputra DK, Lubis AM, Sirait B, Jakobus D, et al. Birth of spinal muscular atrophy unaffected baby from genetically at-risk parents following a pre-implantation genetic screening: a case report. *Int J Reprod Biomed*. 2022;20(9):779-86.
 24. García-Pascual CM, Navarro-Sánchez L, Navarro R, Martínez L, Jiménez J, Rodrigo L, et al. Optimized NGS approach for detection of aneuploidies and mosaicism in PGT-A and imbalances in PGT-SR. *Genes (Basel)*. 2020;11(7):724.
 25. Leigh D, Cram DS, Rechitsky S, Handyside A,

- Wells D, Munne S, et al. PGDIS position statement on the transfer of mosaic embryos 2021. *Reprod Biomed Online*. 2022;45(1):19-25.
26. Mumusoglu S, Ozbek IY, Sokmensuer LK, Polat M, Bozdog G, Papanikolaou E, et al. Duration of blastulation may be associated with ongoing pregnancy rate in single euploid blastocyst transfer cycles. *Reprod Biomed Online*. 2017;35(6):633-9.
 27. Mumusoglu S, Yarali I, Bozdog G, Ozdemir P, Polat M, Sokmensuer LK, et al. Time-lapse morphokinetic assessment has low to moderate ability to predict euploidy when patient- and ovarian stimulation-related factors are taken into account with the use of clustered data analysis. *Fertil Steril*. 2017;107(2):413-21.e4.
 28. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. TensorFlow: a system for large-scale machine learning. *Proceedings of the 12th USENIX symposium on operating systems design and implementation (OSDI)*. Savannah, GA, USA, 2016. 265 p.
 29. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In *proceedings of the 18th international conference on medical image computing and computer-assisted intervention—MICCAI*. Munich, Germany, 5–9 October 2015; Springer: Cham, Switzerland, 2015. 234 p.
 30. Handayani N, Louis CM, Erwin A, Aprilliana T, Polim AA, Sirait B, et al. Machine learning approach to predict clinical pregnancy potential in women undergoing IVF program. *Fertil Reprod*. 2022;04(02):77-87.
 31. K. He, X. Zhang, S. Ren and J. Sun. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*. Las Vegas, NV, USA, 2016. 770 p.
 32. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*. Las Vegas, NV, USA, 2016. 2818 p.
 33. Tan M, Le QV. Efficientnet: rethinking model scaling for convolutional neural networks. *Proceeding of the 36th international conference on machine learning (ICML)*. Rovira I Virgilly University, Tarragona, Spain, 2019. 10691 p.
 34. Thirumalaraju P, Kanakasabapathy MK, Bormann CL, Gupta R, Pooniwala R, Kandula H, et al. Evaluation of deep convolutional neural networks in classifying human embryo images based on their morphological quality. *Heliyon*. 2021;7(2):e06298.
 35. Chen TJ, Zheng WL, Liu CH, Huang I, Lai HH, Liu M. Using deep learning with large dataset of microscope images to develop an automated embryo grading system. *Fertil Reprod*. 2019;1(01):51-6.
 36. Danardon GB, Erwin A, Purnama J, Handayani N, Polim AA, Sini I, et al. A Homogeneous ensemble of robust pre-defined neural network enables automated annotation of human embryo morphokinetics. *J Reprod Infertil*. 2022;23(4):250-6.
 37. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. TensorFlow: a system for large-scale machine learning. *Proceedings of the 12th USENIX symposium on operating systems design and implementation (OSDI)*. Savannah, GA, USA, 2016. 265 p.
 38. 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;37(10):2359-76.
 39. Khan A, Sohail A, Zahoora U, Qureshi AS. A survey of the recent architectures of deep convolutional neural networks. *Artif Intell Rev*. 2020; 53:5455-516.
 40. Huang B, Tan W, Li Z, Jin L. An artificial intelligence model (euploid prediction algorithm) can predict embryo ploidy status based on time-lapse data. *Reprod Biol Endocrinol*. 2021;19(1):185.
 41. Kragh MF, Karstoft H. Embryo selection with artificial intelligence: how to evaluate and compare methods? *J Assist Reprod Genet*. 2021;38(7):1675-89.

Supplementary table 1. Baseline and clinical characteristics of the participants in the present study

Baseline and clinical characteristics	Value (n=425)
Female age (years)	36 (32,39)
Infertility duration (years)	6 (3,9)
Body mass index (kg/m^2)	22.86 (20.89,25.78)
Type of infertility [n (%)]	
Primary	254 (59.8%)
Secondary	171 (40.2%)
AMH (ng/ml)	2.38 (1.21,4.33)
AFC	10 (7,15)
Basal FSH (mIU/ml)	6.46 (5.6,7.94)
Basal LH (mIU/ml)	5.39 (4.2,6.95)
Basal Estradiol (pg/ml)	34.25 (26,47.9)
Basal progesterone (ng/ml)	0.17 (0.09,0.27)
Starting dose of gonadotropin used (IU)	300 (225,300)
Stimulation duration (days)	9 (8,9)
Total gonadotropin dose (IU)	2100 (1500,2400)
Estradiol levels on trigger day (pg/ml)	2818 (1814,4490)
Progesterone levels on trigger day (ng/ml)	0.67 (0.4,1.03)
Endometrial thickness (mm)	10 (9,12)

Data was presented as median (q1, q3) or proportion (n (%))