

OPEN

Predicting pregnancy test results after embryo transfer by image feature extraction and analysis using machine learning

Alejandro Chavez-Badiola^{1*}, Adolfo Flores-Saiffe Farias¹, Gerardo Mendizabal-Ruiz², Rodolfo Garcia-Sanchez¹, Andrew J. Drakeley³ & Juan Paulo Garcia-Sandoval⁴

Assessing the viability of a blastocyst is still empirical and non-reproducible nowadays. We developed an algorithm based on artificial vision and machine learning (and other classifiers) that predicts pregnancy using the beta human chorionic gonadotropin (b-hCG) test from both the morphology of an embryo and the age of the patients. We employed two high-quality databases with known pregnancy outcomes ($n = 221$). We created a system consisting of different classifiers that is feed with novel morphometric features extracted from the digital micrographs, along with other non-morphometric data to predict pregnancy. It was evaluated using five different classifiers: probabilistic bayesian, Support Vector Machines (SVM), deep neural network, decision tree, and Random Forest (RF), using a k-fold cross validation to assess the model's generalization capabilities. In the database A, the SVM classifier achieved an F1 score of 0.74, and AUC of 0.77. In the database B the RF classifier obtained a F1 score of 0.71, and AUC of 0.75. Our results suggest that the system is able to predict a positive pregnancy test from a single digital image, offering a novel approach with the advantages of using a small database, being highly adaptable to different laboratory settings, and easy integration into clinical practice.

Since the inception of IVE, morphology has been the gold standard for embryo selection. Despite the existence of clear and detailed guidelines for morphology-based classifications, blastocyst morphology assessment remains an artisanal technique, which lacks objectivity and reproducibility, limiting its overall predictive value. Furthermore, current scoring systems for human blastocysts focus on only three parameters (*i.e.*, degree of expansion, quality of inner cell mass, and quality of trophoctoderm), while other features that may have an impact on implantation potential are mostly disregarded^{1,2}.

The challenges inherent in image-based diagnosis and decision-making are not unique to reproductive medicine. Efforts directed at improving accuracy and standardization of image analysis through development of computer-aided tools have recently gained attention in other medical fields³⁻⁵, including dermatology and oncology, where deep learning techniques and architectures are in current use for image analysis and to assist diagnosis^{6,7}.

The field of reproductive medicine has also ventured into the world of artificial intelligence (AI)⁸ but, as yet with somewhat limited clinical impact. The quantitative evaluation of embryo characteristics with digital imaging systems has become an active research area^{9,10}, and efforts have been made to improve objectivity during the embryo selection process by combining multiple morphological parameters and morphokinetic analysis in mathematical models, aiming to generate predictive capabilities for development potential and implantation. Such models, however, are often limited to initial development stages of the embryo, dependent on currently existing classifications, or require sophisticated time-lapse microscopy systems¹¹⁻¹⁴.

In this study, we present a computer-aided classification system designed to identify new and objective morphological parameters at the blastocyst stage of *in-vitro* development. By using an independent set of transferred

¹New Hope Fertility Center Mexico, Research and Development, Guadalajara, PC, 44630, Mexico. ²Universidad de Guadalajara, Departamento de Ciencias Computacionales, Guadalajara, PC, 44430, Mexico. ³Hewitt Centre for Reproductive Medicine, Clinical Director, Liverpool, PC, L8 7SS, United Kingdom. ⁴Universidad de Guadalajara, Departamento de Ingeniería Química, Guadalajara, PC, 44430, Mexico. *email: drchavez-badiola@nhfc.mx

Feature	dbA	dbB
Objectives (20X, 40X)	(90, 44)	4, 83)
Donated oocytes (%)	16.92%*	12.64%
Mean oocyte age (STD)	34.11 (6.09)	34.93 (4.96)
Embryos expanding	62.69%	78.17%
Embryos hatching	32.83%	17.23%
Embryos hatched	4.48%	4.60%
b-hCG ≥ 20 mUI/mL (%)	52.98%	56.32%
Proportion of fresh oocytes	13.43%	6.89%

Table 1. Databases description. *4 unknown.

	dbA	dbB	ALL
FPR	0.23	0.61	0.47
FNR	0.27	0.22	0.32

Table 2. False positive and negative rates.

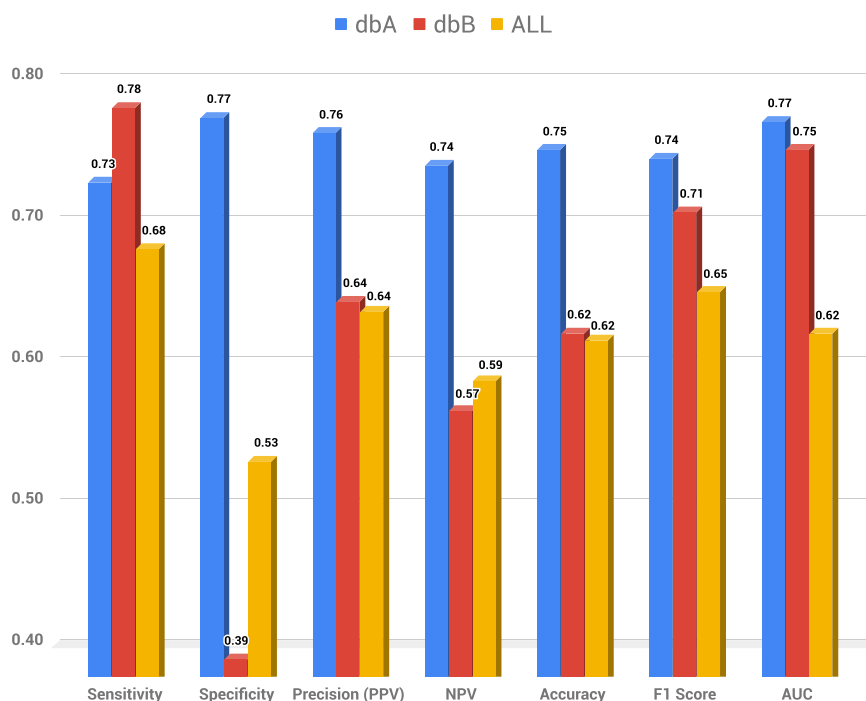


Figure 1. Comparative chart of the best model for each database. The dbA and ALL databases obtained the best F1 score using the Support Vector Machine model, and the dbB database using the Random Forest model. AUC - Area Under the receiver operating characteristic Curve.

embryos with known outcomes, we validated the ability of this system to predict the outcome defined as beta-human chorionic gonadotropin (b-hCG) results.

Results

A total of 221 transferred blastocysts were included in this study: 134 from dbA; and 87 from dbB. Overall positive pregnancy rate per embryo transfer was 54.1% (95% CI 47.3–60.8%). The mean age at oocyte retrieval was 34.4 years. Table 1 shows the differences and similarities between the two data-sets.

Based on the F1 score, the SVM was the model that obtained the best performance in dbA and ALL databases, while Random Forest performed better in the dbB data. Figure 1 shows sensitivity, specificity, precision, accuracy, F1 score, and AUC for the best model for the three data-sets. False-positive rate (FPR) and False-negative rate (FNR) were also computed for all three data-sets (See Table 2).

To ensure our results are not a consequence of over-fitting, we randomized the pregnancy test results overall samples and tested the models. For the dbA database, the SVM obtained an F1 of 0.52; the dbB database obtained an F1 of 0.52 on the RF model, and the ALL database obtained an F1 of 0.57 on the SVM model.

Discussion

The results in this study suggest software prototype AIR E, can successfully extract features from individual blastocyst micrographs to feed its embedded algorithm and then predict b-hCG test results.

For sensitivity, the dbB performed slightly better than dbA (See Fig. 1). However, in specificity, precision, and accuracy, dbA overcame dbB, because of the high FPR of the later (See Fig. 1 and Table 2). The main differences between these two data-sets are the objective used in the microscopy and the proportion of expanding/hatching blastocysts, which suggests that these variables could be influencing the specificity. Still, we have not excluded the possibility of other confounding factors that might be affecting the outcome.

Using all the data-set, we obtained a sensitivity of 0.68 when the algorithm labels a blastocyst will result in a positive b-hCG. This result is close to the success rate expected from transferring euploid blastocysts following a pre-implantation genetic test for aneuploidies (PGT-A)¹⁵. However, we emphasize that the present work is only a pilot study.

Image-based diagnoses and subsequent decision-making processes are challenging in that unless based on objective and standardized criteria, they are likely to jeopardize prognostic accuracy and efficiency of decision making. Current blastocyst selection is in just such a predicament since most accepted classifications are based on subjective assessments during an artisanal process performed by embryologists. Furthermore, and perhaps in the interest of simplicity, current non-morphokinetic classifications are based on characteristics that require no measurements. They are therefore subjective, disregarding variables or characteristics that cannot be consistently identified by the naked eye, but which could potentially be reflective of development potential.

By contrast, computer-aided image processing tools can identify, in an automated, objective, and replicable fashion, key image characteristics beyond what is recognizable by the “naked eye”. Thus, these technologies could help transform current blastocyst classification from subjective and morphology-based, to an automated, objective, and standardized image processing software. In reproductive medicine, efforts have been directed towards implementing computational tools and artificial intelligence software for the blastocyst selection process, but until recently most of these efforts have either focused on predicting blastocyst formation⁸ or have been designed to fit current embryo classifications^{10,16,17}, moving a step further into automation, but failing to link embryo classification to viability prediction.

Our current software prototype, on the other hand, takes a different approach by disregarding existing classifications and linking independent and new variables to prognosis. By bypassing “standard” morphology, the prognosis determined by the algorithm could allow an embryo ranking system and a more straightforward decision-making process during embryo selection for transfer.

In the same line as our study, and to associate morphological characteristics extracted from micrographs to prognosis, Tran *et al.* (2019) trained an algorithm using 8836 embryos to predict fetal heart pregnancy using a time-lapse incubator¹⁴. However and despite the high AUC obtained (0.93), these results should be taken with caution, since they used a highly unbalanced data-set (92% of negative) that could be deceptive of the model performance, especially when the AUC is the only reported measure and this test has been described as sub-optimal for such data-sets^{18,19}. Besides, the authors included embryos with “failed or abnormal fertilization and grossly abnormal morphology” into the negative group, which could also mislead the reported measure because they did not reach the blastocyst stage and therefore the outcome was known before the embryo transfer even without the need of an algorithm.

Another cornerstone to Tran’s, Rocha’s, and Khosravi’s approach is that input information came from micrographs obtained from the EmbryoScope^{10,14,17}, allowing the investigators to standardize variables related to the use of different imaging protocols and settings (*e.g.*, camera type, objectives, microscope type, etc.). This advantage, however, seems to be tempered by a lack of flexibility, for example, the inability to classify hatching embryos, and even more critical, a limitation in the form of a requirement for expensive time-lapse equipment.

Once again, we decided against a tempting standardized approach but instead trained our algorithm for image extraction and processing after identifying and automatically standardizing for imaging variables. In this way, we aimed to generate a computational tool more natural to implement in clinical practice, with equipment currently available in most IVF laboratories.

Approaches to predict blastocyst quality, like the ones discussed above, rely on the need for large sample size due to their bottom-up approach. AIR E, on the other hand, achieves high accuracy using a relatively small sample size due to the software’s ability first to extract features from the images and then to feed these to the classifier (top-down approach). This top-down approach allows AIR E to learn from each microscope and thus to make it highly flexible and adaptable to multiple clinics and conditions.

To the best of our knowledge, this is the first AI classifier trained and evaluated on static micrographs of blastocyst-stage embryos of single transferred embryos that predicts b-hCG.

We acknowledge classification algorithms thrive on numbers, and therefore, the authors are currently pursuing a study with a larger sample size with the final goal of developing a blastocyst ranking system based on prognosis. We believe the implementation of this program could impact on standardization of embryologists decisions, assist towards IVF laboratory automation and improved efficiency, assist with research protocols where embryo grading is the core of the study, and more important, to reduce time to pregnancy and live birth by supporting the selection of the best embryo available for transfer each cycle. Still, we believe a study design like the one we present, is not strong enough to attempt reaching such conclusions and a prospective study, as the subject of future work, would validate the clinical application of the proposed model.

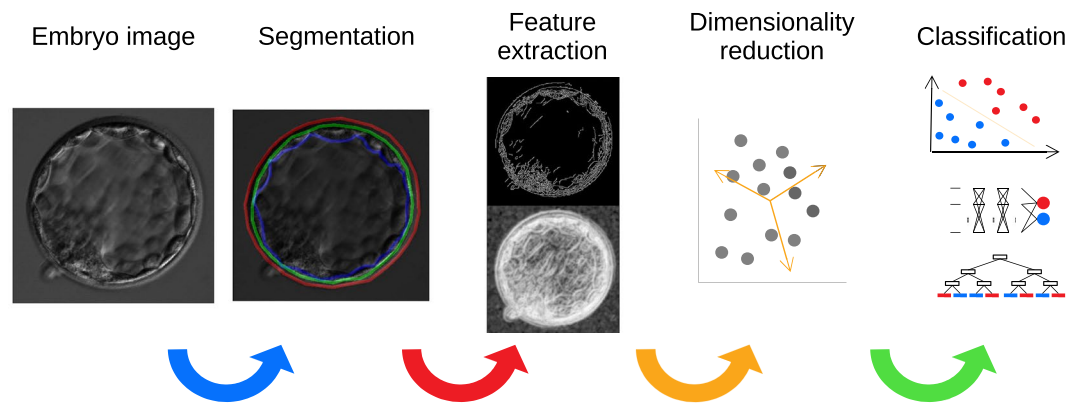


Figure 2. Overview of the AIR E pipeline. Once the embryo image is taken, it is passed through Segmentation, Feature Extraction, Feature Selection, and Classification.

Methods

We performed a morphometric analysis of micrographs of blastocysts transferred as single embryos between 2015 and 2019 at two fertility centers in Mexico (dbA and dB). The databases were tested both independently and combined (data-set named ALL). A total of 221 blastocysts were photographed on day 5–6 after insemination using either of two inverted microscopes: Olympus IX71 (dbA); or Olympus IX73 (dB), both with Hoffmann modulation contrast, a digital camera and a Hamilton Thorne ZILOS-tki $\frac{1}{2}$ Laser camera. The complete data-set consisted of images taken at 20x (total magnification 200x), and 40x (total magnification 400x). Embryologists did not adjust light features and calibration in a standard fashion. The later means that the observed features are developed from varying picture settings akin to a large variation similar to using different equipment. Each blastocyst was photographed one time only, with a focus on the zona pellucida and trophectoderm.

Table 2 describes relevant metadata of the databases, including the objectives used, the oocyte source, the mean age of the oocyte, the embryo stage, the proportion of fresh samples, and b-hCG ≥ 20 .

All transfers were single embryo transfers of blastocysts (180 in diameter as measured before transfer), performed following previously published protocols²⁰. In brief, fresh transfers were performed five days following oocyte retrieval with luteal support given as 400 mg vaginal progesterone (Geslutin, Asofarma, USA), every 12 hours until the pregnancy test. For frozen embryo transfer, endometrial preparation was started with 4 mg of oral estradiol (Primogyn, Schering AG, Colombia) from day 3 of the menstrual period and supplemented with 400 mg vaginal progesterone (Geslutin, Asofarma, USA) every 12 hours starting on day 14–16 of the cycle; both continued until the pregnancy test. Blastocysts were thawed 2–4 hours before embryo transfer on day +6 from the beginning of luteal phase support. Seven days following embryo transfer, a quantitative b-hCG assay was performed. Beta hCG levels ≥ 20 mUI/mL were interpreted as positive.

Image analysis and statistical assessment. An image processing algorithm available in the AIR E prototype system was used to analyze all 221 blastocysts included in the study (see Fig. 2). Using AIR E, two embryologists manually marked the boundaries (delimited) of the zona pellucida and the trophectoderm, as it is observed in the Fig. 2 (Segmentation), allowing the algorithm to extract and compute a total of 24 features from each image by performing pattern variation analysis of pixel intensities, and other relevant morphological properties (e.g., areas and perimeters). To homogenize images, AIR E transformed all the features from pixels to micrometers according to the calibration data of each microscope. In a step-wise fashion, AIR E extracts the following features: perimeter, area, pixel information (obtained by applying an entropy filter to the embryo image), the standard deviation of the pixel information, and the amount of boarders found within the image (obtained through the canny algorithm). These metrics were obtained independently for the group of pixels labeled as zona pellucida, trophectoderm and inner area (blastocoel and inner cell mass). The feature extraction step is followed by principal components analysis (PCA) on all of the features as a pre-processing step, obtaining 14 components which represent 99% of the variability in the data-set. Patients' age was included in AIR E as the only feature not based on image analysis.

AIR E performed training of five supervised ML models to classify samples with b-hCG ≥ 20 mUI/mL. The selected ML models are representative of the different types of classifiers: (1) Naive Bayes, (2) v - Support Vector Machine (v-SVR model for minimization of the errors, and RBF kernel with $g = 0.14$), (3) deep Neural Network (three layers of 30 units each and one with 10 units, ReLu activation function, Adam solver, and L2-Regularization of 0.0001), (4) Decision Tree (unrestricted), and (5) Random Forest (100 trees). To assess the generalization capability of the classification models and avoid over-fitting, we used stratified 10-fold cross-validation as suggested in previous studies²¹, and the reported measures are the average of the results of each fold. To assess the results, we computed the sensitivity, specificity, precision, accuracy, F1, and Area Under the receiver operating characteristic curve (AUC).

Ethical considerations. All patients signed appropriate consents for treatment according to internal protocols and the use of the images of the blastocysts for research purposes. The data has been anonymised upon collection. None of the authors accessed the patients information during the data analysis. The data has restricted

access. Since this study was a retrospective study for image processing validation with no additional intervention, IRB (CONBIOETICA 09-CEI-00120170131) approval was waived (RA 2018-01) by the New Hope Fertility Center Mexico's Ethics Committee. All methods were performed in accordance with the relevant guidelines and regulations.

All the data employed for this study was anonymized, and the authors did not have access to any information that could reveal the patient's identities during the study.

The data that support the findings of this study are available on request from the corresponding author ACB. The data are not publicly available due to them containing information that could compromise participant consent. The authors did not have access to any information that could reveal the patient's identities during the study.

Received: 30 August 2019; Accepted: 19 February 2020;

Published online: 10 March 2020

References

- Sciorio, R., Thong, J. K. & Pickering, S. J. Comparison of the development of human embryos cultured in either an EmbryoScope or benchtop incubator. *J. Assist. Reproduction Genet.* **35**, 515–522, <https://doi.org/10.1007/s10815-017-1100-6> (2018).
- Yoshida, I. H. *et al.* Can trophoctoderm morphology act as a predictor for euploidy? *JBRA Assist. Reproduction* **22**, 113–115, <https://doi.org/10.5935/1518-0557.20180036> (2018).
- Gandomkar, Z., Brennan, P. C. & Mello-Thoms, C. MuDeRN: Multi-category classification of breast histopathological image using deep residual networks. *Artif. Intell. Medicine* **88**, 14–24, <https://doi.org/10.1016/j.artmed.2018.04.005> (2018).
- Rebouças Filho, P. P. *et al.* Analysis of human tissue densities: A new approach to extract features from medical images. *Pattern Recognit. Lett.* **94**, 211–218, <https://doi.org/10.1016/j.patrec.2017.02.005> (2017).
- Tang, Q., Liu, Y. & Liu, H. Medical image classification via multiscale representation learning. *Artif. Intell. Medicine* **79**, 71–78, <https://doi.org/10.1016/j.artmed.2017.06.009> (2017).
- Guo, X. *et al.* From spoken narratives to domain knowledge: Mining linguistic data for medical image understanding. *Artif. Intell. Medicine* **62**, 79–90, <https://doi.org/10.1016/j.artmed.2014.08.001> (2014).
- Hu, Z. *et al.* Deep learning for image-based cancer detection and diagnosis - A survey. *Pattern Recognit.* **83**, 134–149, <https://doi.org/10.1016/j.patcog.2018.05.014> (2018).
- Goodman, L. R., Goldberg, J., Falcone, T., Austin, C. & Desai, N. Does the addition of time-lapse morphokinetics in the selection of embryos for transfer improve pregnancy rates? A randomized controlled trial. *Fertility Steril.* **105**, 275–285.e10, <https://doi.org/10.1016/j.fertnstert.2015.10.013> (2016).
- Lagalla, C. *et al.* A quantitative approach to blastocyst quality evaluation: morphometric analysis and related IVF outcomes. *J. Assist. Reproduction Genet.* **32**, 705–712, <https://doi.org/10.1007/s10815-015-0469-3> (2015).
- Rocha, J. C. *et al.* Automatized image processing of bovine blastocysts produced *in vitro* for quantitative variable determination. *Sci. Data* **4**, 170192, <https://doi.org/10.1038/sdata.2017.192> (2017).
- Mio, Y. Morphological Analysis of Human Embryonic Development Using Time-Lapse Cinematography. *J. Mammalian Ova Res.* **23**, 27–35, <https://doi.org/10.1274/jmor.23.27> (2006).
- Storr, A. *et al.* Morphokinetic parameters using time-lapse technology and day 5 embryo quality: a prospective cohort study. *J. Assist. Reproduction Genet.* **32**, 1151–1160, <https://doi.org/10.1007/s10815-015-0534-y> (2015).
- Majumdar, G., Majumdar, A., Verma, I. C. & Upadhyaya, K. C. Relationship Between Morphology, Euploidy and Implantation Potential of Cleavage and Blastocyst Stage Embryos. *J. human reproductive sciences* **10**, 49–57, <https://doi.org/10.4103/0974-1208.204013> (2017).
- Tran, D., Cooke, S., Illingworth, P. J. & Gardner, D. K. Deep learning as a predictive tool for fetal heart pregnancy following time-lapse incubation and blastocyst transfer. *Hum. Reproduction* **34**, 1011–1018, <https://doi.org/10.1093/humrep/dez064> (2019).
- López-Rioja, M. *et al.* Estudio genético preimplantación para aneuploidias: resultados de la transición entre diferentes tecnologías. *Ginecol Obstet Mex* **86**, 96–107, <https://doi.org/10.24245/gom.v86i2.1634> (2018).
- Rocha, J. C. *et al.* A Method Based on Artificial Intelligence To Fully Automate The Evaluation of Bovine Blastocyst Images. *Sci. Reports* **7**, 7659, <https://doi.org/10.1038/s41598-017-08104-9> (2017).
- Khosravi, P. *et al.* Deep learning enables robust assessment and selection of human blastocysts after *in vitro* fertilization. *npj Digit. Medicine* **2**, 21, <https://doi.org/10.1038/s41746-019-0096-y> (2019).
- Lobo, J. M., Jiménez-Valverde, A. & Real, R. AUC: a misleading measure of the performance of predictive distribution models. *Glob. Ecol. Biogeogr.* **17**, 145–151, <https://doi.org/10.1111/j.1466-8238.2007.00358.x> (2008).
- Saito, T. & Rehmsmeier, M. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLoS One* **10**, e0118432, <https://doi.org/10.1371/journal.pone.0118432> (2015).
- Zhang, J. J. *et al.* Minimal stimulation IVF vs conventional IVF: a randomized controlled trial. *Am. J. Obstet. Gynecol.* **214**, 96.e1–96.e8, <https://doi.org/10.1016/j.ajog.2015.08.009> (2016).
- Vabalas, A., Gowen, E., Poliakoff, E. & Casson, A. J. Machine learning algorithm validation with a limited sample size. *PLoS One* **14**, e0224365, <https://doi.org/10.1371/journal.pone.0224365> (2019).

Acknowledgements

The authors want to thank Mina Alikani Ph.D., for her input, comments, and encouragement during the preparation of this paper. We also want to thank the embryologists' teams at New Hope Guadalajara and Mexico City.

Author contributions

A.C.B. Conceived the main idea and experiments, A.F.-S.F., R.G.S. and G.M.R. conducted the experiments and analysed the results, A.J.D. and J.P.G.S were advisers during the whole process. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.C.-B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020