

# Automatic characterization of human embryos at day 4 post-insemination from time-lapse imaging using supervised contrastive learning and inductive transfer learning techniques

Elena Payá<sup>a,b,\*</sup>, Lorena Bori<sup>b</sup>, Adrián Colomer<sup>a</sup>, Marcos Meseguer<sup>b</sup>, Valery Naranjo<sup>a</sup>

<sup>a</sup> Instituto de Investigación e Innovación en Bioingeniería, Universitat Politècnica de València, Valencia, 46022, Spain

<sup>b</sup> IVI-RMA Valencia, Spain



## ARTICLE INFO

### Article history:

Received 4 January 2022

Revised 3 May 2022

Accepted 15 May 2022

### Keywords:

Supervised contrastive learning

Inductive transfer learning

Viability assessment

Quality assessment

Embryo grading

Convolutional neural networks

## ABSTRACT

**Background:** Embryo morphology is a predictive marker for implantation success and ultimately live births. Viability evaluation and quality grading are commonly used to select the embryo with the highest implantation potential. However, the traditional method of manual embryo assessment is time-consuming and highly susceptible to inter- and intra-observer variability. Automation of this process results in more objective and accurate predictions.

**Method:** In this paper, we propose a novel methodology based on deep learning to automatically evaluate the morphological appearance of human embryos from time-lapse imaging. A supervised contrastive learning framework is implemented to predict embryo viability at day 4 and day 5, and an inductive transfer approach is applied to classify embryo quality at both times.

**Results:** Results showed that both methods outperformed conventional approaches and improved state-of-the-art embryology results for an independent test set. The viability result achieved an accuracy of 0.8103 and 0.9330 and the quality results reached values of 0.7500 and 0.8001 for day 4 and day 5, respectively. Furthermore, qualitative results kept consistency with the clinical interpretation.

**Conclusions:** The proposed methods are up to date with the artificial intelligence literature and have been proven to be promising. Furthermore, our findings represent a breakthrough in the field of embryology in that they study the possibilities of embryo selection at day 4. Moreover, the grad-CAMs findings are directly in line with embryologists' decisions. Finally, our results demonstrated excellent potential for the inclusion of the models in clinical practice.

© 2022 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

## 1. Introduction

Fertility impairment affects approximately 80 million people worldwide. Although infertility rates vary widely, ranging from less than 5% to more than 30%, it is estimated that one in ten couples suffers from primary or secondary infertility problems [1]. To address this problem, Assisted Reproduction Techniques (ART) were developed. A typical procedure begins with the retrieval of multiple oocytes from the ovaries of the patient and subsequent in vitro fertilization (IVF) by intracytoplasmic sperm injection (ICSI). Then,

the zygotes are cultured in incubators; optimal culture conditions and correct embryo selection methods are essential for a successful treatment.

In vitro embryo evaluation has improved considerably over the last 20 years, focusing on the analysis of morphological and morphokinetic characteristics. Its ultimate goal is to predict embryo development and the chances of successful implantation and pregnancy. However, only 30% of embryo transfers result in clinical pregnancy [2]. To increase the chance of clinical pregnancy, multiple embryo transfer (MET) has been used. However, MET can result in multiple gestation and the associated maternal and neonatal risks [3], leading to significant pressure to move towards single embryo transfer (SET). There is a clear need for a rapid, non-invasive, and accurate method to assist in the evaluation of em-

\* Corresponding author at: Instituto de Investigación e Innovación en Bioingeniería, Universitat Politècnica de València, Valencia 46022, Spain.

E-mail address: [elpabos@i3b.upv.es](mailto:elpabos@i3b.upv.es) (E. Payá).

bryos, as until now there have been no reliable methods to assess embryo viability and quality. The introduction of time-lapse incubators allowed continuous monitoring of the embryo, responding to the limitations of conventional evaluation by facilitating the study of embryo development without disturbing the culture conditions.

Currently, embryos are manually assessed by an expert embryologist who studies morphology at different stages of development and classifies them into predetermined classes. Morphokinetics and morphology have been shown to be related to implantation success and ultimately live births [4]. However, the traditional method of manual annotation is time-consuming and highly susceptible to inter- and intra-observer variability, making it subjective and inaccurate [5,6]. Studies with multiple highly trained embryologists demonstrated large disagreements even with simplified grading methods. This can be caused by a range of factors, including the exact timing of the observation or the discretization of a dynamic evolution.

On the one hand, embryo transfer is conventionally performed at the cleavage stage (days 2–3 after fertilization), or at the blastocyst stage (days 5–7 after fertilization). At the latter stage, the evaluation of expanded blastocysts is more accurate due to the differentiation of their structures. In addition, attempts to predict blastocyst formation and quality have focused mainly on day 3 assessment, but the results are not particularly accurate due to the changes that the embryo can experience. On the other hand, day 4 assessment is not routinely performed by the embryologist, although it is an alternative day of selection and transfer. Some studies show equal implantation and pregnancy rates for embryo transfer at the morula stage (day 4 after fertilization) [7–9]. Taking this into consideration, an accurate evaluation on day 4 post insemination, just before the increased complexity of the culture conditions demanded in the transition from day 4 to day 5, could allow for early decision making and save time in the clinic.

In terms of the methods used, older approaches have recently been replaced by more efficient new techniques. Deep learning algorithms, specifically, convolutional neural networks (CNNs), have been used to address various medical imaging problems. Thus, they have become the technique of choice in computer vision and are the most successful type of models for image analysis. Recently, deep CNNs architectures such as ResNet [10] have dramatically increased the progress rate of deep learning methods in image classification. The following section briefly describes the evolution of the state-of-the-art in the field of embryology with the introduction of computer vision and presents the novelty and contributions of the work.

The rest of the paper is structured as follows. Section 3 describes the databases used during the study. Section 4 outlines the underlying methodologies, which are composed of two strategies: supervised contrastive and inductive transfer learning. Section 5 explains the ablation experiments performed during the validation stage. Section 6 shows qualitative and quantitative results achieved in the test prediction. Section 7 presents a discussion about the proposed framework and contributions to the field. Finally, Section 8 summarizes the main conclusions.

## 2. Related work

The emergence of time-lapse incubators has led to a considerable increase in the development of experimental methods to analyze the implantation potential and live birth rate (LBR) of IVF embryos from images or videos, particularly in the creation of embryo selection algorithms (ESAs). Multiple ESAs with different parameters, optimal times, and endpoints have been published [11–14]. However, most of these proposals have not achieved a positive predictive value higher than 45% in the selection of good-quality

embryos with relatively poor clinical outcomes. Moreover, the results obtained are different; e.g., Meseguer et al. [11] reports that t5 has significance with implantation while Cruz et al. [12] does not. These papers recognize that the data used come from fertility clinics with different annotation protocols and the authors claim that a consensus is needed to achieve an efficient ESA. In global terms, the generalization of manual morphological and morphokinetic annotations is very difficult due to the high variability. In this context, some works were proposed to solve these problems and to be able to automate the annotations by emulating the skill of a trained specialist in embryo evaluation.

Regarding morphokinetic analysis, the determination of the moment of cell division led to the development of algorithms capable of automatically counting the number of blastomeres. Different methods grouped in segmentation-based [15–17] and classification-based approaches [18–25] can be found in the literature. Traditional classification methods performed manual feature extraction from image shapes and textures [18–20], whereas more recent works developed methods based on deep learning techniques where the prior extraction of characteristics is automatically performed [21–23]. Dirvanauskas et al. [26] introduced a novel AI technique that used generative adversarial network to generate one-, two-, and four-cell stage images. In addition, some methods used consecutive frames, avoiding the individual tracking of embryo cells by dynamic programming to enforce monotonicity of predictions [24,25].

In contrast to morphokinetic analysis, morphological assessment refers to static appearance. Some studies attempted to classify blastocyst using specific characteristics manually obtained such as inner cell mass (ICM) area, trophoctoderm (TE) area, zona pellucida thickness, or expanded blastocyst diameter [27,28]. Despite the reasonable accuracies obtained, these methods required advanced embryological knowledge and several processing steps. Consequently, they are not scalable to large datasets or to clinical practice.

Furthermore, static analysis of morphology has proven to be completely relevant in embryo assessment and many studies have focused on automating this part by making use of CNNs. They solve tasks related to the quality [29–33], the stage reached [34–36], or viability of the embryo [37,38]. Others segmented the embryo structures at different stages using convolutional encoder-decoder architectures for oocytes [39,40] or blastocysts [38]. In addition, there is a need to add external information in order to predict the implantation potential or the pregnancy likelihood since it can be affected by other variables such as the age of the patient, the condition of the uterus or the chromosome abnormalities [30,41].

Khosravi et al. [30] developed a framework (STORK) for blastocyst quality analysis by training a neural network based on the Google Inception architecture [42] using approximately 12,000 time-lapse images. Typical pre-trained deep architectures have been widely used, but with the aim of classifying embryos according to whether they are at the blastocyst stage [31,32,35–37], highlighting some differences. Wang et al. [32] used multiple images with different focal depths so that the network had more information about embryos and Bormann et al. [37] employed a genetic algorithm to weight the outputs of the network during training. In a novel way, Rad et al. [38] added a Compact-Contextualize-Calibrate block that helped guide the feature extraction process and performed segmentation-based classification techniques to predict implantation. A different approach developed by Kragh et al. [33] to predict the morphological quality of blastocyst structures included temporal information by training a CNN with images from 90 to 115 h post insemination (hpi) to feed a recurrent neural network (RNN) that connected features from adjacent frames.

## 2.1. Contribution of this work

Motivated by the possibilities offered by categorization at day 4 post insemination [43] and by the existing gap in the study of embryos at this stage, AI-based methods are developed for embryo study based on images exclusively. In particular, this paper proposes an innovative framework based on supervised contrastive learning and inductive transfer learning for the prediction of embryo viability and quality at 90 hpi. To the best of the author's knowledge, this is the first study in the embryology field which presents an approach based on these novel AI-techniques for automatic embryo characterization. In addition, we further studied possible day 4 characterization that describes and distinguishes between embryos, thus advancing automation and decreasing culture time without disturbing development and implantation success rates.

Several papers have explored the shortcomings of cross-entropy loss, such as the lack of robustness to noisy labels [44,45] and the possibility of poor margins, leading to reduced generalization performance. The recent emergence of the seminar work on contrastive learning has led to significant advances in self-supervised learning approaches [46,47]. The common idea behind contrastive learning is to encourage putting an anchor and a positive sample representation together in the embedding space and separating the anchor from the negative samples.

Inspired by the high performance reported by previous studies based on contrastive learning [48,49] and by the inefficiency of the cross-entropy loss (CEL) function in embryo image analysis, in this work, we propose a supervised contrastive loss for solving viability classification tasks. The contrastive loss outperforms cross-entropy on fully supervised learning tasks [50] and has been shown to be more stable to hyperparameter tuning, optimizers, and data augmentation. An encoder convolutional network followed by a projection head is trained with supervised contrastive loss (SCL). Research of optimal projection heads has not been carried out yet [50] for this loss function, therefore, we also study the performance of different approaches.

In parallel, interest in transfer learning techniques has grown in recent years. The fundamental motivation is to apply knowledge learned previously to solve new problems faster and/or improving the performance and the capacity for generalization of predictive models. In particular, it aims to extract knowledge from a source task to infer it in a target task. Transfer learning strategies can be classified into various branches depending on the nature of the data and the problem to be solved. Whether the source and target domains are the same and the tasks are different but related, inductive transfer learning is used. In contrast, if the source and target domains are different but the task is the same, transductive transfer learning is preferred [51]. In this work, an inductive transfer learning (ITL) method is required, since both the source and target domains are embryo images, but the source and target tasks are viability and quality embryo estimation, respectively.

Numerous papers adopt an ITL strategy to solve medical problems [52–56]. Caruana et al. [52] used multi-task learning and proposed an ITL scheme for pneumonia risk prediction. Silver et al. [53] implemented a task rehearsal method as an approach to lifelong learning that used representation of previously learned tasks as a source of inductive bias. Zhou et al. [55] proposed an inductive method to improve the performance of ocular multi-disease identification.

In summary, the main contributions of this work are:

- New loss function, SCL, for embryo viability classification in charge of maximizing the separation between samples of different classes and minimizing the distance among examples of the

**Table 1**

IVI database description. Number of patients with their respective cycles.

	Number of patients	Number of cycles
<b>Viability dataset</b>	1289	3014
<b>Quality dataset</b>	559	830

same classes leading to increase the performance of the classification task.

- Research of different projection head modules for the supervised contrastive framework.
- Novel inductive transfer learning technique on the embryology field for quality embryo classification.
- First automatic system to analyze 90 hpi images for automatic embryo characterization.
- No requirement of a prior curation step (i.e. manual embryo selection) in order to integrate it into clinical practice. System uses raw data with no intervention required by an embryologist.

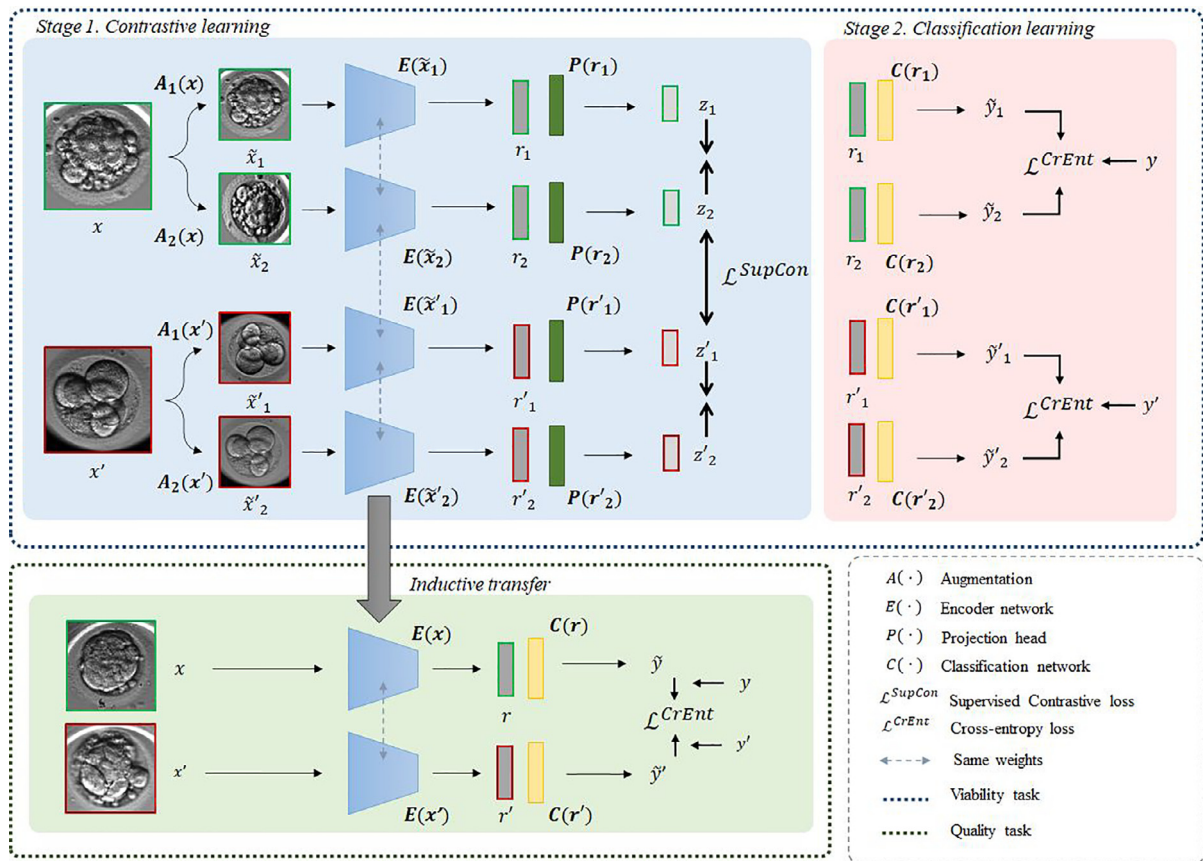
## 3. Material

This research was a single-center retrospective study carried out at IVI Valencia. We included recipients who underwent ICSI cycles for three consecutive years. Embryos were cultured in the EmbryoScope time-lapse system up to the blastocyst stage. This equipment provides images with a resolution of  $500 \times 500$  pixels which were taken automatically every 10–20 min and in up to 7–11 focal planes. Fertilization was evaluated at 16–19 h after ICSI and confirmed by the presence of two pronuclei and two polar bodies. Later, the blastocyst was assessed by applying a hierarchic classification procedure based on the standard “Asociación para el Estudio de la Biología de la Reproducción” (ASEBIR) morphologic grading. The embryo images and other patient data collected in this study are not publicly available due to reasonable ethics and privacy concerns.

A private database was created from the cycles described above. It contains 3014 embryo cycles from 1289 different patients for the development and evaluation of the viability models and 830 embryo cycles from 559 different patients for the quality models (see Table 1). Quality assessment was performed only on viable embryos; therefore the dataset is more limited. Both the viability and quality datasets were balanced between their two classes. The viability dataset contained the viable and non-viable labels, while the quality dataset included the high and low-quality labels. The latter was based on the ASEBIR classification, with high quality corresponding to classes A and B and low quality to class C. Classes A and B generate discrepancy among embryologists, so we could not entirely rely on the labels.

From each embryo, two inputs for independent models were constructed consisting of frames at 90 (day 4) and 115 hpi (day 5) using embryo detection and segmentation from the detector and cropper modules (see Appendix A). Note that we also obtained images at 115 hpi to validate and compare our models concerning the state-of-the-art. From the available focal planes, three were selected, corresponding to the center and the adjacent focals. The final image size, i.e. the input size of the model, was  $224 \times 224 \times 3$  (one grayscale channel per focus).

In this work, data partitioning to develop and evaluate the predictive models was carried out as follows: 85% for the training stage and 15% for the blind test. From the training set, we randomly split the data again into training and validation subsets according to 90% and 10%. The validation set was used to optimize the model hyper-parameters and monitor the fitting of the net-



**Fig. 1.** Overview of the framework proposed to classify embryo viability and quality. Data augmentation module, encoder network, and projection head (blue) are implemented for the first stage of contrastive learning. The loss function acts by attracting and repelling samples from the same class and different classes, respectively. Then, the trained encoder extracts latent features to feed a neural network to classify images according to the viability task (red). The inductive transfer learning technique is applied to infer knowledge from a primary task and solve the quality embryo prediction task (green).

work. This partitioning pipeline was performed for both viability and quality tasks.

Experiments were conducted on the NVIDIA DGX A100 system. NVIDIA DGX A100 is the universal system for all workloads, providing unprecedented compute density, performance, and flexibility in 5 petaFLOPs. All the scenarios were implemented in TensorFlow 2.4.0 on Python 3.6.

## 4. Methods

In our embryo evaluation scenario, we advocated for a supervised contrastive strategy to separate viability classes and get a rich encoder to improve the quality task. Specifically, we carried out a two-step scenario as it is shown at the top of Fig. 1 (Section 4.1). Firstly, an encoder and a projection head were trained with supervised contrastive loss to solve the viability task. Secondly, a classification network was trained to classify viability classes. Once we obtained this encoder, we applied a transfer learning approach for accomplishing the quality task (Section 4.2). We used the trained viability models as a source knowledge to transfer information of embryo properties to improve quality models (see bottom of Fig. 1). In the following sections, we detail the methodology and the learning steps.

### 4.1. Supervised contrastive learning for viability embryo characterization

Making use of the two-stage method, we aimed to obtain a rich feature space capable of separating viable and non-viable em-

bryo classes. In this section, we describe the supervised contrastive framework and loss function implemented.

#### 4.1.1. Learning framework

The implemented learning framework is based on Khosla et al. structure [50], which is similar to that used for self-supervised contrastive learning [47,48]. The overview scheme is shown at the top of Fig. 1.

During the encoding stage (see blue Fig. 1 – Stage 1), we optimized the supervised contrastive loss to obtain more accurate inter and intra-class boundaries. The embedding network consists of three major components:

- Data augmentation module,  $A(\cdot)$ , that randomly transforms the given data,  $\tilde{x} = A(x)$ , resulting in two correlated views of the same sample, denoted by  $\tilde{x}_1$  and  $\tilde{x}_2$ . It helps to increase the amount of relevant data, thus enabling invariance and generalization. Note that data augmentation was conducted only during training. Random flipping, rotation, and contrast were applied to each image with a probability of 0.2.
- Encoder network,  $E(\cdot)$ , which extracts  $\tilde{x}$  to a vector representation,  $r = E(\tilde{x}) \in \mathbb{R}^{D_E}$ . Augmented samples are separately input to the same encoder, resulting in a pair of representation vectors. Our framework allows various choices of the base network architecture. Supervised contrastive loss (SCL) has shown more efficiency for complex encoder architectures.
- Projection head,  $P(\cdot)$  which maps the representations  $r$  to a lower dimensionality space  $z = P(r) \in \mathbb{R}^{D_P}$ , where SCL was applied. Although latent space projections increase performance, the research of the optimal has not been conducted

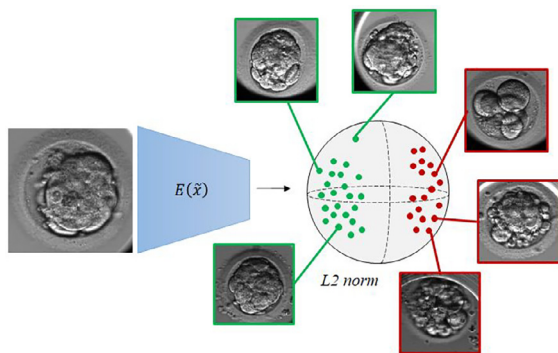


Fig. 2. Representation of normalization function in 2-dimensional space.

yet. Motivated by the latter, we studied different approaches: (1) A multilayer perceptron (MLP) with one hidden layer,  $z = W^2\sigma(W^1r)$  where  $\sigma$  is a LeakyRelu non-linear activation. (2) A clustering layer for dimensionality reduction where the number of clusters is the projection size. (3) A Random Fourier Projection (RFP) based on Rahimi et al. approach [57], which implements a mapping from input space to an output space, which approximates shift-invariant kernels. The parameters were sampled from a Gaussian distribution.

During the classifier training (see red Fig. 1 – Stage 2) and for the final model, we discarded the projection head  $P(\cdot)$  and used the trained encoder network  $E(\cdot)$  and representation  $r$  for resolving the embryo characterization for the viability task. The classifier  $C(\cdot)$  was a fully connected layer with a normalization to a probability distribution over predicted classes and it was trained with cross-entropy loss (CEL).

#### 4.1.2. Supervised contrastive loss

Given the proposed framework, we used a contrastive loss adapted to the supervised domain which aims to maximize the agreement between different positive pairs (samples that belong to the same class) in the latent space while minimizing the agreement between negative pairs (samples that are from different classes). Before applying the loss function, we normalized the output of the projection network to lie on the unitary hypersphere (Fig. 2), which enables the use of an inner product to measure distances in the projection space. An  $l_2$  regularization layer outputs the normalized vector:

$$z = \frac{\tilde{z}}{\|\tilde{z}\|_2} = \frac{\tilde{z}}{\sqrt{\sum_i \tilde{z}_i^2 + \varepsilon}}$$

For a set of  $I = \{X, Y\}$  independent training samples, where  $y_i$  refers to the ground truth of the  $i$ -sample,  $x_i$ , being  $i = 1, 2, \dots, N$  the number of image, the corresponding batch for training consists of positive pairs and negative pairs. Then, the loss function is defined as:

$$\mathcal{L}^{\text{SC}} = \sum_{i \in I} -\frac{1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(z_i \cdot z_p / \tau)}{\sum_{a \in A(i)} \exp(z_i \cdot z_a / \tau)}$$

where  $z_i = P(E(\tilde{x}_i)) \in \mathbb{R}^{D_p}$ , the symbol  $\cdot$  denotes the inner product  $u \cdot v = \mu^T v / \|u\| \|v\|$  between  $l_2$  normalized vectors  $z_i$  and  $z_p$ ,  $\tau \in \mathbb{R}^+$  is a scalar temperature hyperparameter, and  $P(i) \equiv \{p \in A(i) : y_p = y_i\}$  is the set of all positives samples in the batch where  $i \neq p$ .

The function encourages the encoder to give closely aligned representations to entries from the same class, resulting in a more robust clustering of the representation space. By using many positive pairs and many negative pairs, we are able to better model both intra- and inter-class variability. Optimal temperature can

Table 2  
Classification results reached during the validation step for the different base encoder networks.

	ResNet50	VGG16
<b>Accuracy</b>	<b>0.8277</b>	0.8103
<b>Precision</b>	<b>0.8349</b>	0.8145
<b>Sensitivity</b>	<b>0.8161</b>	0.8036
<b>Specificity</b>	<b>0.8393</b>	0.8169
<b>F1-Score</b>	<b>0.8254</b>	0.8090

outperform cross-entropy experiments. Low values improve representation learning since they strongly weight on harder negative pairs, which have been shown to increase classification accuracy [58].

#### 4.2. Inductive transfer learning for quality embryo prediction

One of the main challenges of deep learning solutions is the limited availability of large labeled datasets. The inductive transfer learning (ITL) paradigm allows the transference of knowledge from a larger source dataset to improve the predictive performance of a smaller target dataset. In addition, it helps to solve a task of higher difficulty by providing information from a related dataset. In this work, the source and target tasks are the estimation of embryo viability and embryo quality, respectively. Making use of the inductive transfer method, we aimed to better separate high- and low-quality embryo classes.

Given a source domain  $D_S$  and a learning task  $T_S$ , a target domain  $D_T$  and a learning task  $T_T$ , ITL aims to help improve the learning of the target predictive function  $f_T(\cdot)$  in  $D_T$  using the knowledge in  $D_S$  and  $T_S$ , where  $T_S \neq T_T$  [51]. A task is defined as  $T = \{y, P(Y|X)\}$ ; therefore, the condition  $T_S \neq T_T$  implies that either the label spaces between the domains are different, i.e.,  $y_S \neq y_T$ , or the conditional probability distribution between domains is different, i.e.,  $P(Y_S|X_S) \neq P(Y_T|X_T)$ .

The approach used is parameter-transfer [51]. The intuitive idea behind this is the assumption that the models for related tasks share some parameters or prior distributions of the hyperparameters. Thus, by discovering the shared parameters, knowledge is transferred across tasks. In this case, we used the model trained with the SCL method as the source knowledge (see green Fig. 1).

## 5. Ablation experiments

#### 5.1. Supervised contrastive learning settings

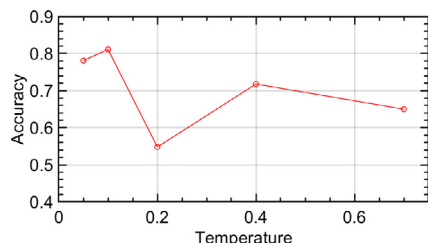
In this section, we report the validation performance of the contrastive learning-based approach for the different proposed scenarios. The target task is the prediction of viability with embryo imaging at day 4 post-insemination.

According to the state-of-the-art of contrastive learning and embryo evaluation, the well-known ResNet50 and VGG16 networks pre-trained with ImageNet were applied as backbone encoder. To address an objective comparison of the proposed backbones, we kept constant the projection heads and the temperature. For the projection head, we used the MLP with one hidden layer of 128 neurons non-linearly activated by the ReLU function. For the temperature, the value of 0.1 suggested by Khosla et al. [59] as optimal, was used. In Table 2, we contrast the validation results achieved by the different encoders trained in a binary-class scenario (viable/non-viable classes). The comparison was handled through different figures of merit, such as accuracy, precision, sensitivity, specificity, and F1-score. Notably, the base encoder network reporting the best performance, ResNet50, was selected for the following experiments.

**Table 3**

Classification results reached during the validation step for the different projection head modules.

	(1) MLP	(2) K-means	(3) RFP
<b>Accuracy</b>	0.8277	0.8018	<b>0.8485</b>
<b>Precision</b>	<b>0.8349</b>	0.7787	0.7609
<b>Sensitivity</b>	0.8161	0.8444	<b>0.8974</b>
<b>Specificity</b>	<b>0.8393</b>	0.7589	0.8167
<b>F1-Score</b>	<b>0.8254</b>	0.8102	0.8235

**Fig. 3.** Accuracy value for different values of the temperature parameter.

In this section, we also report the validation performance using the projection head modules described in Section 4.1.1. Specifically, we compare a MLP with one hidden layer of 128 neurons non-linearly activated by the ReLU function, a clustering layer, and an RFP (see Table 3). All projection head modules mapped the representations to an embedding vector in a 128-dimensional space. Note that for all the settings, pre-trained ResNet50 was used as backbone network and only the projection head module was changed. Regarding the specific parameter of the contrastive learning-based strategy, the temperature selected was 0.1.

Finally, we studied the effect of the temperature on loss function because of the important role it plays in the supervised contrastive approach. The optimal value of this parameter depends on the task and the dataset. We found that lower temperatures benefited the learning of our representation spaces as they give more weight to harder negatives (see Fig. 3).

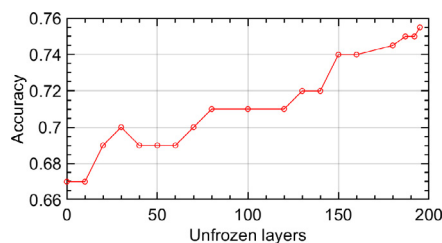
**Training details** Training hyperparameters were obtained from empirical evaluations with a wide range of setting and optimal configuration were presented. Early Stopping was applied to prevent overfitting. For the first stage, all the models were trained during 200 epochs using a learning rate of 0.00001 with a batch size of 16. Adam optimizer was used to minimize the SCL function. For the second stage, models were trained during 20 epochs using a cosine decay rate initialized with a learning rate of 0.00001. Adam was applied to minimize CEL function.

## 5.2. Inductive transfer learning configuration

In this section, we report the validation performance of the inductive transfer learning approach for the different scenarios. The target task is the binary classification of embryo quality with images at day 4 post-insemination. We used the model trained earlier with SCL as a base, thus transferring the knowledge from the viability dataset to improve the results with the quality dataset. In this work, the viability task is considered simpler and, in addition, we had a larger database, which made it a good option as a source task.

To achieve the highest performance, we optimized empirically the layer from which the freezing strategy is applied. We compare the results following the unfreezing of model layers (see Fig. 4).

**Training details** The optimal hyperparameters combination was achieved by training models during 20 epochs using a cosine decay rate initialized with a learning rate of 0.00001. Adam was applied to minimize CEL function.

**Fig. 4.** Improvement representation with the unfrozen layers in ITL framework.

## 6. Results

### 6.1. Viability prediction

In this section, we report the testing performance of the contrastive learning-based approach in comparison to the conventional classification methods for embryo viability estimation. Note that, as previously mentioned, although the objective is to predict viability at 90 hpi, results were also obtained for images at 115 hpi to validate and compare our approach concerning the state-of-the-art. Predictions were performed using the architectures with the best performance during the validation stage, i.e., the ResNet50 architecture with MLP as a projection head and a temperature of 0.1.

Table 4 presents the quantitative results achieved by the proposed learning strategy and the conventional cross-entropy loss approach during the blind test, and Fig. 5 shows the receiver operating characteristic (ROC) curves and area under curve (AUC) values. Additionally, to provide a more comprehensive interpretation, we illustrate a 2D map from the test predictions corresponding to the latent space (see Fig. 6) by means of t-distributed Stochastic Neighbor Embedding (t-SNE) [60].

Gradient-weighted Class Activation Mapping (Grad-CAM) [61] were computed to highlight the regions of interest in which the proposed method paid attention to predict the viability of the test samples. The reported activation maps allow a better understanding of the automatic extractor of features, i.e. the CNN, by remarking the most relevant information. This technique provides helpful information about patterns and shapes in the relevant embryonic structures and their association with the predicted class. Fig. 7 shows some images with their corresponding Grad-CAMs for correctly classified embryos and Fig. 8 for misclassified embryos.

### 6.2. Quality prediction

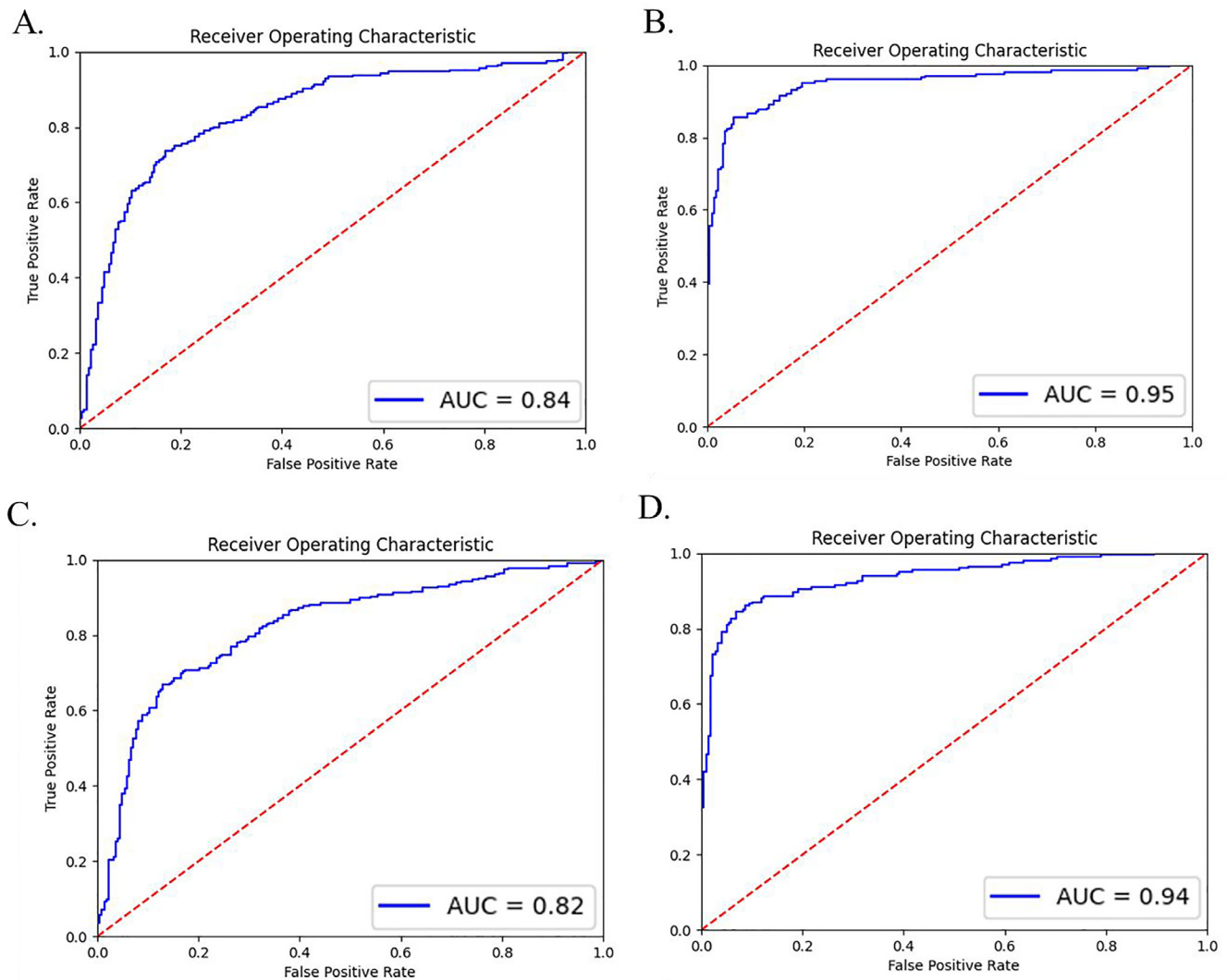
We report the testing performance of the inductive transfer learning approach for quality embryo evaluation. Table 5 summarizes the quantitative results and Fig. 10 shows ROC curves and AUC values. In order to validate our method, we trained the same deep neural network but without knowledge transfer. In the same way as the viability task, predictions were performed using the best architectures, i.e., by unfreezing all layers and retraining them. In addition, we also show 2D maps of the latent space generated by t-SNE to support our quantitative findings (see Fig. 9).

## 7. Discussion

In this section, we refer to the main contributions detailed throughout the paper and we review the obtained results. Concerning the embryology research field, it is important to note that this work is the first that conducted day 4 embryo characterization through artificial intelligence methods, according to the literature. In addition, we perform two different tasks: viability and quality prediction. One task precedes the next, whereby once it has been determined which embryos are viable, the quality of the embryos

**Table 4**  
Test results reached for viability prediction by the different learning strategies.

	Day 4 (90 hpi)		Day 5 (115 hpi)	
	Conventional categorical cross-entropy	Proposed supervised contrastive	Conventional categorical cross-entropy	Proposed supervised contrastive
<b>Accuracy</b>	0.7405	<b>0.8103</b>	0.8835	<b>0.9330</b>
<b>Precision</b>	0.7768	<b>0.8170</b>	0.8045	<b>0.9364</b>
<b>Sensitivity</b>	0.7250	<b>0.8062</b>	<b>0.9725</b>	0.9364
<b>Specificity</b>	0.7585	<b>0.8145</b>	0.8130	<b>0.9293</b>
<b>F1-Score</b>	0.7500	<b>0.8115</b>	0.8806	<b>0.9364</b>
<b>AUC</b>	0.8200	<b>0.8400</b>	0.9400	<b>0.9500</b>



**Fig. 5.** ROC Curve and AUC values for A, Day-4 viability proposed model; B, Day-5 viability proposed model; C, Day-4 viability conventional model; and D, Day-5 viability conventional model.

**Table 5**  
Test results reached for quality prediction by the different learning strategies.

	Day 4 (90 hpi)		Day 5 (115 hpi)	
	Conventional from scratch	Proposed inductive transfer	Conventional from scratch	Proposed inductive transfer
<b>Accuracy</b>	0.6810	<b>0.7500</b>	0.7980	<b>0.8001</b>
<b>Precision</b>	0.5345	<b>0.7931</b>	0.6739	<b>0.7347</b>
<b>Sensitivity</b>	<b>0.7561</b>	0.7302	0.8611	<b>0.9231</b>
<b>Specificity</b>	0.6400	<b>0.7736</b>	0.7619	<b>0.7937</b>
<b>F1-Score</b>	0.6263	<b>0.7603</b>	0.7561	<b>0.8182</b>
<b>AUC</b>	0.7200	<b>0.7700</b>	0.8200	<b>0.8700</b>

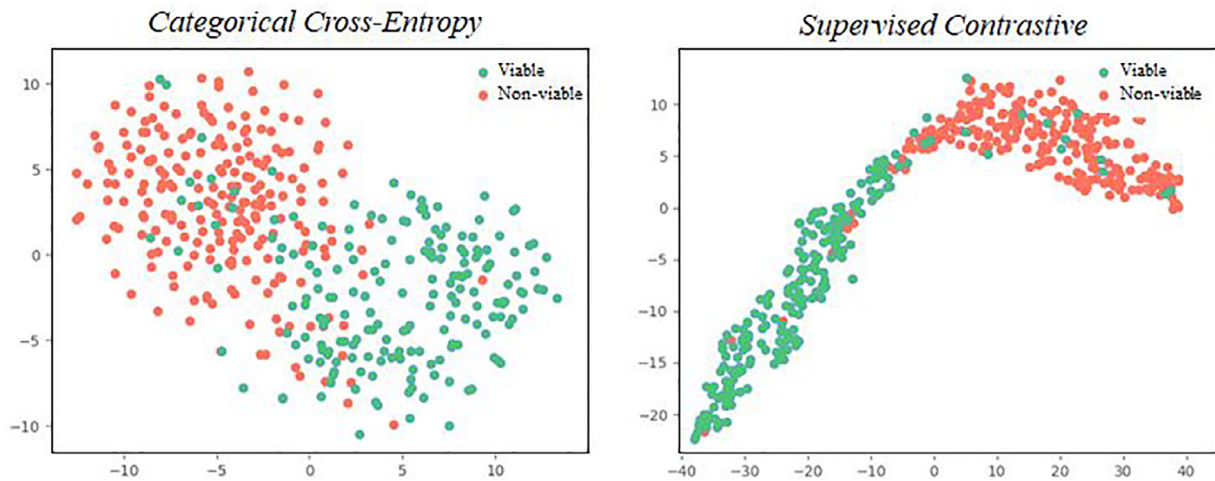


Fig. 6. Representation of the latent space from the prediction of the data set. Red dots correspond to non-viable embryos and green dots correspond to viable embryos.

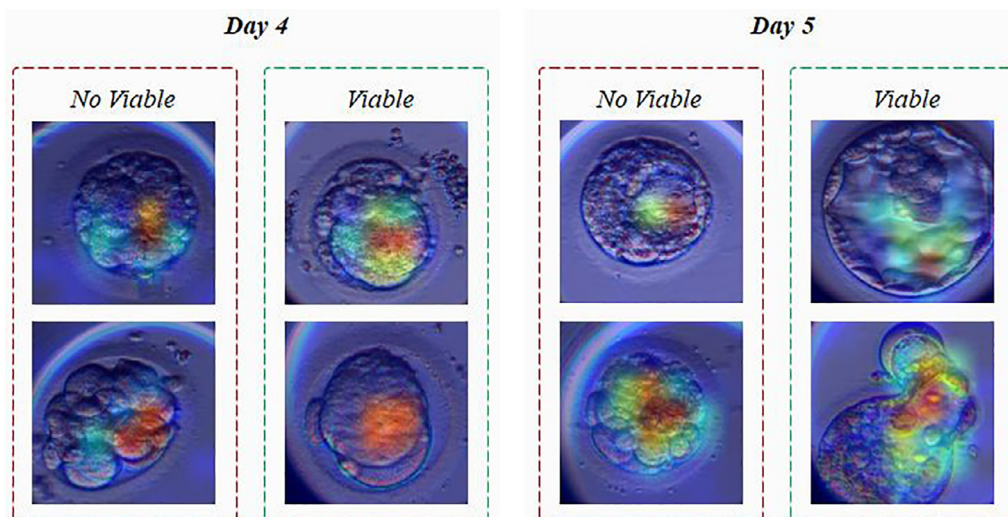


Fig. 7. Grad-CAMs. Heat maps extracted from the selected architecture corresponding to well-classified images.

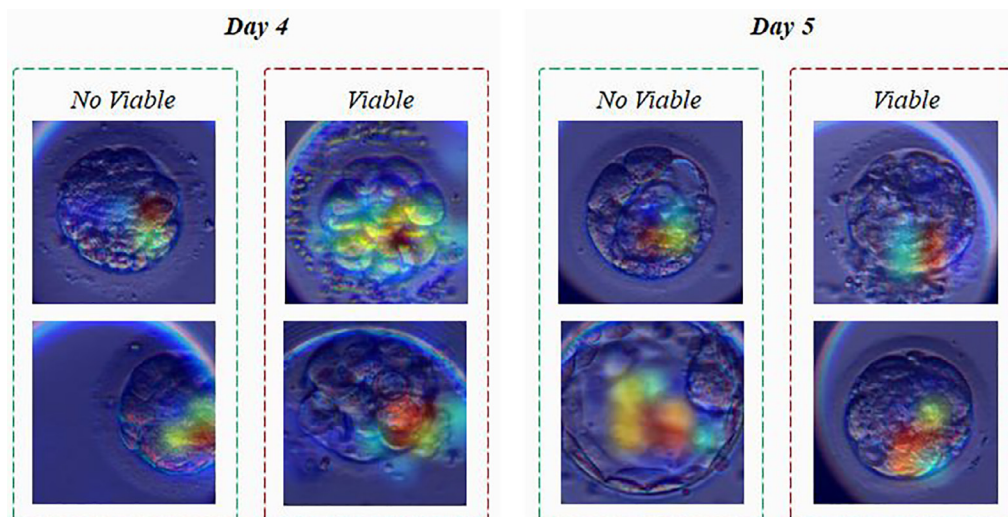
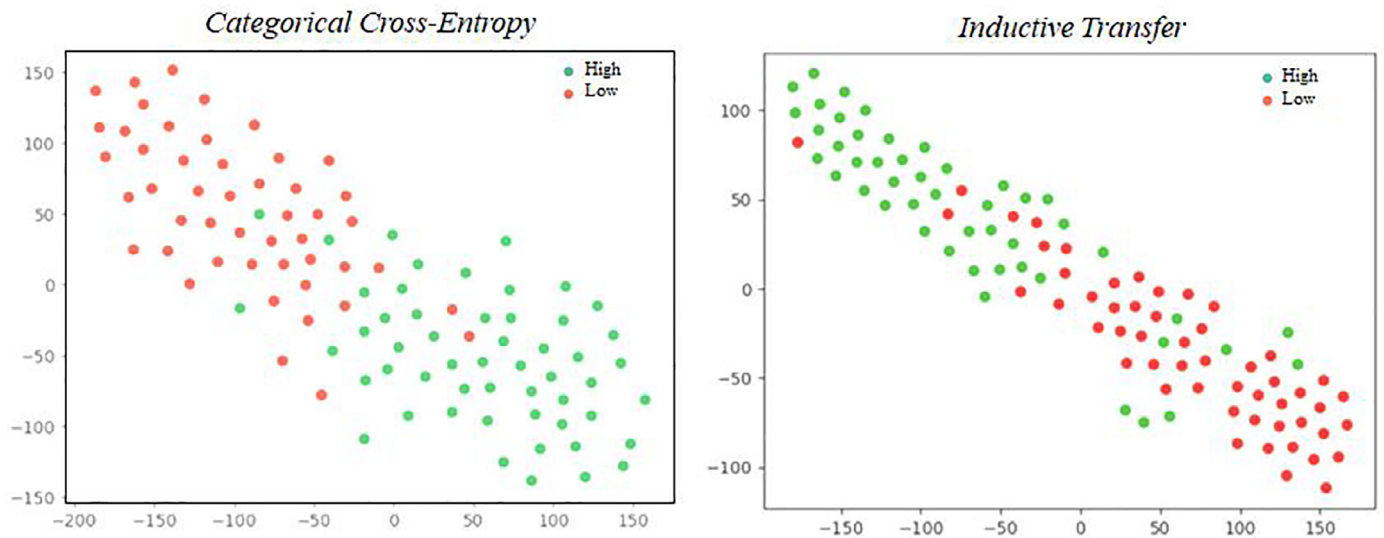
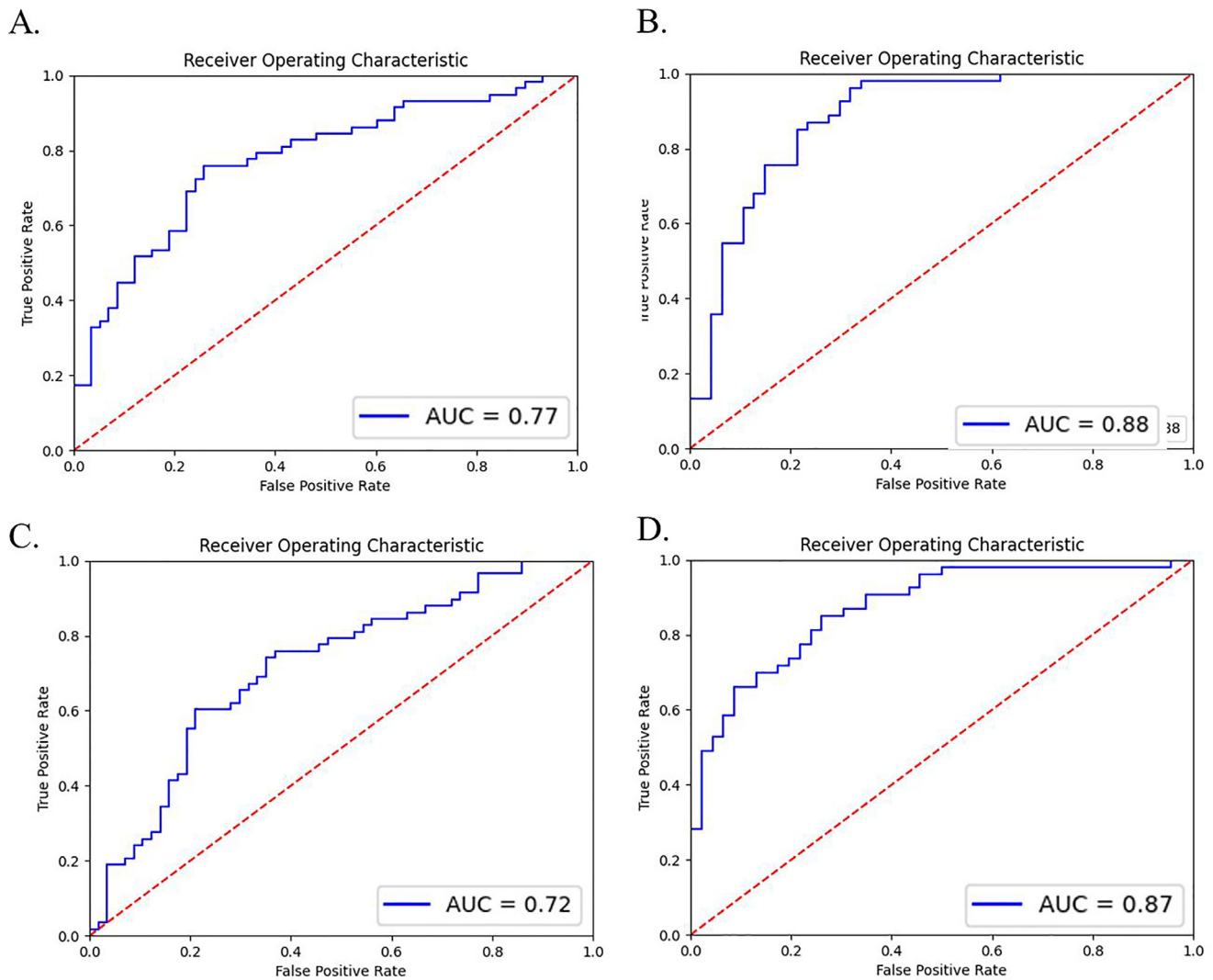


Fig. 8. Grad-CAMs. Heat maps obtained with the selected architecture on misclassified images.





**Fig. 9.** Representation of the latent space from the prediction of the data set. Red dots correspond to low-quality embryos and green dots correspond to high-quality embryos.



**Fig. 10.** ROC Curve and AUC values of quality models. A, Day-4 proposed model; B, Day-5 proposed model; C, Day-4 conventional model; and D, Day-5 conventional model.

is evaluated. A high-quality viable embryo is more likely to implant and, ultimately, lead to a successful live birth.

### 7.1. About the ablation experiments

By means of the ablation experiments, we determined the best configuration for each method used and for each task. In this way, different models were constructed depending on the possibilities of our databases.

On the one hand, one of the main novelties addressed in this paper is the proposed framework which used supervised contrastive learning for embryo characterization. We contrasted different architectures for the proposed method to find the best experimental setting for our problem. More specifically, we corroborated that ResNet50 enables the extraction of a richer latent space of features improving the traditional VGG16 architecture. We observed that the use of residual blocks provided better results. In addition, we studied different projection heads and we observed that a random Fourier feature projection reports better accuracy, sensitivity, and F1-score. MLP with one hidden layer obtained better results for precision and specificity. Finally, we confirmed that the optimal value for the temperature parameter was 0.1, which is consistent with the value found by Khosla et al. [59].

On the other hand, we decided to use the fine-tuning technique by inferring viability knowledge extracted from the supervised contrastive model for quality prediction, considering the limited amount of available samples for this task. Specifically, we observed that the more trainable layers, the better the performance, but with decreasing improvement at the end.

### 7.2. About the prediction results

Firstly, as observed in the 2D projections in Fig. 6, supervised contrastive loss results in a richer feature space, with better-defined clusters boundaries and more compactness of the samples belonging to a given class. These findings are supported by the quantitative results in Table 4, which show that the models trained with the novel loss function report higher performance. In addition, the conventional loss function has a higher discrepancy between the metrics obtained, evidencing imbalanced classifications. Regarding the inductive transfer method proposed using the viability as source task to infer knowledge in the quality task, our method outperforms the conventional training of a deep CNN according to Table 5 and, as shown in Fig. 9, results in a more compact latent space.

During the embryo evaluation, the specialist analyzes the stage reached, viability, and quality of the embryo. These are studied on day 5 post-insemination when the blastocyst stage is reached. For this purpose, the morphology of the structures involved is assessed, specifically the ICM and the TE. According to ASEBIR, a compact ICM with a size of 1900–3800  $\mu\text{m}^2$  and a homogeneous, cohesive, multi-cellular TE correspond to high-quality embryos and, ultimately, higher implantation rates. To be more precise, TE morphology is the main factor in giving an overall grade, which is consistent with our findings of day 5 grad-CAMs. As shown in the right-side image of Fig. 7, the resulting model for day 5 pays more attention to central texture, probably due to the use of three focal frames that enable the central area covered with TE cells to be seen. In addition, day 4 grad-CAMs (left-side image of Fig. 7) show higher importance to the texture of the morula or cell edges depending on viability or non-viability in agreement with previous understanding. Thus, the qualitative results reported keep consistency with the clinical interpretation. Regarding the misclassified embryos (Fig. 8), many of these contain fragments or are not centered in the image, which appears to be a source of error. Moreover, model seems to have difficulty differentiating the degree

of compaction when the edges of the cells are smoothed. Besides, although the day-5 image shown in the first column below was incorrectly labeled as a non-viable embryo, our model classified it as viable, thus demonstrating its ability to overcome the subjectivity associated with embryo evaluation and labeling.

In this paper, we contrast the proposed model with other embryo evaluation methods recently published in Khosravi et al. [30], Thirumalaraju et al. [31], Wang et al. [32], Kragh et al. [33], Kanakasabapathy et al. [35], Thirumalaraju et al. [36], Bormann et al. [37], Rad et al. [38]. A direct comparison is challenging not only because of the diversity of the data sources but also because of the variation of the acquisition process and annotation methodologies among fertility clinics. Moreover, each study aimed to solve different tasks, with some determining whether the embryo reached the blastocyst stage; this is comparable to our viability task. Bormann et al. [37] and Thirumalaraju et al. [31] obtained an accuracy of 90.97% with images at 113 hpi, and Thirumalaraju et al. [36] and Manoj et al. [35] obtained an accuracy of approximately 71.42% and 71.87% with images at 70 hpi, respectively. In addition, Wang et al. [32], who performed a classification similar to our viability task, reached an accuracy rate of 91.74% with images at 116 hpi. These values are in the range of those achieved in the present work, with ours being slightly higher (see Table 4). Other studies performed similar tasks to our quality assessment. Kragh et al. [33] predicted the quality of the TE and ICM independently. They obtained accuracy values of 58.6% and 61.1% for 115 hpi images and an AUC value of 0.646, for which the proposed methodology outperforms them (see Table 5).

Through the accomplishment of this project, a new embryo selection tool based on artificial intelligence will be available to support the embryologist's decision during IVF treatment. The resulting algorithm will eliminate inter-observer variability with an objective and reliable method of selecting the embryo most likely to be implanted, because of the relation with the viability and high-quality classes. This differs substantially from the usual clinical practice since up to now embryologists have invested a large number of hours in the analysis and evaluation of embryonic development. Additionally, accurate assessment on day 4 will save culture time and increase clinic efficiency, allowing a decision to be made with a sufficient amount of time. Regarding the translational capacity of the experimental project into the field of clinical embryology, the implementation of the corresponding findings should necessarily constitute a benefit for the human-assisted reproduction area.

Although the present study obtained a high predictive accuracy, it still has limitations. AI algorithms should be trained on an extensive database, so future works with more data and a prospective approach should be carried out. Moreover, this research was a single-center retrospective; therefore, our future direction is to expand the sample size by collecting data from multiple IVF clinics to improve performance and allow the generalization of deep learning models. Additionally, the existing algorithm has been developed only from images of one of the existing time-lapse incubators, the EmbryoScope. Nevertheless, several time-lapse incubators are available in the market, with different image quality, definition, magnification, focal planes, or frequency of images; therefore, the algorithm would be redesigned to be applied in multiple systems. Finally, a crowdsourcing approach will improve generalization and will let us include a separation between A and B quality grades.

## 8. Conclusion

In this paper, we have proposed several artificial intelligence solutions, a supervised contrastive framework to perform viability prediction, and an inductive transfer framework for quality estimation. The proposed model introduces supervised contrastive loss

for embryo analysis and studies different projection head networks in pursuit of more accurate predictions. Furthermore, we transfer the viability-trained encoder to resolve the quality prediction task for which we have a shorter data set. This innovative approach adds substantial improvements to the final prediction.

In summary, the results suggest that the proposed method is promising in general and specifically in the field of embryology. By removing the need for manual pre-processing of embryos, the system becomes fully automated and, consequently, ready for inclusion in the clinic as a tool to assist embryologists.

### Funding

This work has been partially funded by Agencia Valenciana de la Innovación (AVI) (2002-VLC-011-MM).

The work of Elena Pay Bosch has been supported by the Spanish Government (DIN2018-009911) and the work of Valery Naranjo Ornedo by the Generalitat Valenciana (AEST/2021/054).

We gratefully acknowledge the support from the Generalitat Valenciana with the donation of the DGX A100 used for this work, action co-financed by the European Union through the Operational Program of the European Regional Development Fund of the Comunitat Valenciana 2014–2020 (IDIFEDER/2020/030).

### Declaration of Competing Interest

Authors declare that they have no conflict of interest.

### Appendix A. Embryo detection and cropping modules

In order to reduce computational complexity while optimizing the CNN filters, we cropped out images centered on the embryo. Previously, we detected the existence of it in the well (see Fig. A.11).

For the cropping operation, we trained a CNN that returned bounding box values and selected a size of 280 pixels for 90 hpi and 300 for 115 hpi due to blastocyst expansion. This process removed a considerable part of the image that did not contain rel-

evant information. Finally, the cropped image is resized to  $224 \times 224$  pixels.

### References

- [1] WHO | Current practices and controversies in assisted reproduction, 2014.
- [2] C. De Geyter, C. Calhaz-Jorge, M.S. Kupka, C. Wyns, E. Mocanu, T. Motrenko, G. Scaravelli, J. Smeenk, S. Vidakovic, V. Goossens, T. E. I. monitoring Consortium (EIM) for the European society of human reproduction, E. (ESHRE), ART in Europe, 2014: results generated from European registries by ESHRE, Hum. Reprod. 33 (9) (2018) 1586–1601.
- [3] E.R. Norwitz, V. Edusa, J.S. Park, Maternal physiology and complications of multiple pregnancy, in: Seminars in Perinatology, vol. 29, Elsevier, 2005, pp. 338–348.
- [4] A. Ferraretti, V. Goossens, J. de Mouzon, S. Bhattacharya, J. Castilla, V. Korsak, M. Kupka, K. Nygren, A. Nyboe Andersen, T. E. I. monitoring (EIM), f. t. E. S. o. H. R. Consortium, E. (ESHRE), assisted reproductive technology in Europe, 2008: results generated from European registers by ESHRE, Hum. Reprod. 27 (9) (2012) 2571–2584.
- [5] A. Storr, C.A. Venetis, S. Cooke, S. Kilani, W. Ledger, Inter-observer and intra-observer agreement between embryologists during selection of a single day 5 embryo for transfer: a multicenter study, Hum. Reprod. 32 (2) (2017) 307–314.
- [6] E. Adolphsson, A.N. Andershed, Morphology vs. morphokinetics: a retrospective comparison of inter-observer and intra-observer agreement between embryologists on blastocysts with known implantation outcome, JBRA Assist. Reprod. 22 (3) (2018) 228.
- [7] D. Feil, R.C. Henshaw, M. Lane, Day 4 embryo selection is equal to day 5 using a new embryo scoring system validated in single embryo transfers, Hum. Reprod. 23 (7) (2008) 1505–1510.
- [8] J.C. Skorupski, D.E. Stein, U. Acholonu, H. Field, M. Keltz, Successful pregnancy rates achieved with day 4 embryo transfers, Fertil. Steril. 87 (4) (2007) 788–791.
- [9] V. Holschbach, J. Weigert, J.E. Dietrich, S. Roesner, M. Montag, T. Strowitzki, B. Toth, Pregnancy rates of day 4 and day 5 embryos after culture in an integrated time-lapse incubator, Reprod. Biol. Endocrinol. 15 (1) (2017) 1–8.
- [10] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
- [11] M. Meseguer, J. Herrero, A. Tejera, K.M. Hilligsøe, N.B. Ramsing, J. Remohí, The use of morphokinetics as a predictor of embryo implantation, Hum. Reprod. 26 (10) (2011) 2658–2671.
- [12] M. Cruz, N. Garrido, J. Herrero, I. Pérez-Cano, M. Muñoz, M. Meseguer, Timing of cell division in human cleavage-stage embryos is linked with blastocyst formation and quality, Reprod. Biomed. Online 25 (4) (2012) 371–381.
- [13] S. Chamayou, P. Patrizio, G. Storaci, V. Tomaselli, C. Alecci, C. Ragolia, C. Crescenzo, A. Guglielmino, The use of morphokinetic parameters to select all embryos with full capacity to implant, J. Assist. Reprod. Genet. 30 (5) (2013) 703–710.
- [14] M. Dal Canto, G. Cotichio, M.M. Renzini, E. De Ponti, P.V. Novara, F. Brambilasca, R. Comi, R. Fadini, Cleavage kinetics analysis of human embryos predicts development to blastocyst and implantation, Reprod. Biomed. Online 25 (5) (2012) 474–480.
- [15] A. Khan, S. Gould, M. Salzmann, Segmentation of developing human embryo in time-lapse microscopy, in: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), IEEE, 2016, pp. 930–934.
- [16] R.M. Rad, P. Saeedi, J. Au, J. Havelock, A hybrid approach for multiple blastomeres identification in early human embryo images, Comput. Biol. Med. 101 (2018) 100–111.
- [17] L. Lockhart, P. Saeedi, J. Au, J. Havelock, Human embryo cell centroid localization and counting in time-lapse sequences, in: 2020 25th International Conference on Pattern Recognition (ICPR), IEEE, 2021, pp. 8306–8311.
- [18] F. Mousavi, Y. Wang, P. Lorenzen, J. Oakley, D. Russakoff, S. Gould, A unified graphical models framework for automated mitosis detection in human embryos, IEEE Trans. Med. Imaging 33 (7) (2014) 1551–1562.
- [19] A. Khan, S. Gould, M. Salzmann, Automated monitoring of human embryonic cells up to the 5-cell stage in time-lapse microscopy images, in: 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), IEEE, 2015, pp. 389–393.
- [20] C. Strouthopoulos, G. Anifandis, An automated blastomere identification method for the evaluation of day 2 embryos during IVF/ICSI treatments, Comput. Methods Programs Biomed. 156 (2018) 53–59.
- [21] J. Gingold, N. Ng, J. McAuley, Z. Lipton, N. Desai, Predicting embryo morphokinetic annotations from time-lapse videos using convolutional neural networks, Fertil. Steril. 110 (4) (2018) e220.
- [22] R.M. Rad, P. Saeedi, J. Au, J. Havelock, Blastomere cell counting and centroid localization in microscopic images of human embryo, in: 2018 IEEE 20th International Workshop on Multimedia Signal Processing (MMSp), IEEE, 2018, pp. 1–6.
- [23] D. Dirvanauskas, R. Maskeliunas, V. Raudonis, R. Damasevicius, Embryo development stage prediction algorithm for automated time lapse incubators, Comput. Methods Programs Biomed. 177 (2019) 161–174.
- [24] J. Silva-Rodríguez, A. Colomer, M. Meseguer, V. Naranjo, Predicting the success of blastocyst implantation from morphokinetic parameters estimated through

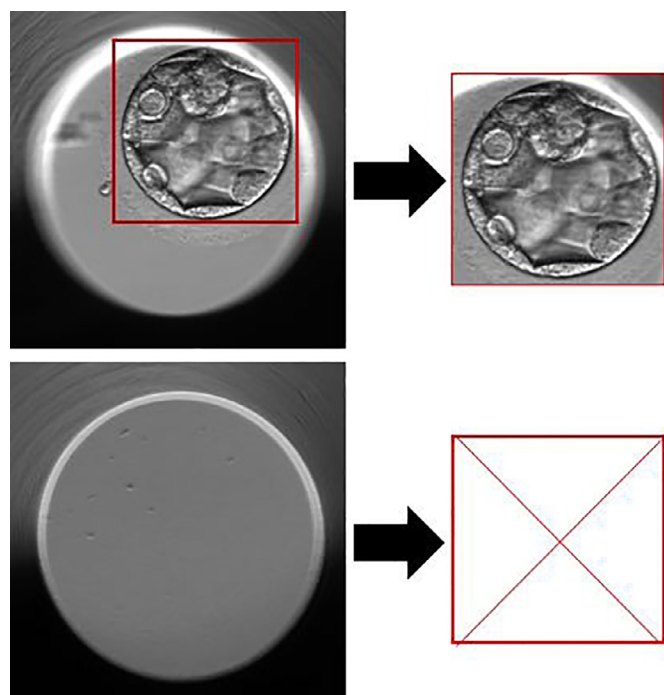


Fig. A1. Detection and cropping of time-lapse images.

- cnn and sum of absolute differences, in: 2019 27th European Signal Processing Conference (EUSIPCO), IEEE, 2019, pp. 1–5.
- [25] N.H. Ng, J.J. McAuley, J. Gingold, N. Desai, Z.C. Lipton, Predicting embryo morphokinetics in videos with late fusion nets & dynamic decoders, ICLR (Workshop), 2018.
- [26] D. Dirvanauskas, R. Maskeliūnas, V. Raudonis, R. Damaševičius, R. Scherer, Hemigen: human embryo image generator based on generative adversarial networks, *Sensors* 19 (16) (2019) 3578.
- [27] L. Bori, E. Payá, L. Alegre, T.A. Viloría, J.A. Remohi, V. Naranjo, M. Meseguer, Novel and conventional embryo parameters as input data for artificial neural networks: an artificial intelligence model applied for prediction of the implantation potential, *Fertil. Steril.* 114 (6) (2020) 1232–1241.
- [28] E.S. Filho, J. Noble, M. Poli, T. Griffiths, G. Emerson, D. Wells, A method for semi-automatic grading of human blastocyst microscope images, *Hum. Reprod.* 27 (9) (2012) 2641–2648.
- [29] R.S.H. Wicaksono, A.A. Septiandri, A. Jamal, Human embryo classification using self-supervised learning, in: 2021 2nd International Conference on Artificial Intelligence and Data Sciences (AIDAS), IEEE, 2021, pp. 1–5.
- [30] P. Khosravi, E. Kazemi, Q. Zhan, J.E. Malmsten, M. Toschi, P. Zisimopoulos, A. Sigaras, S. Lavery, L.A.D. Cooper, C. Hickman, M. Meseguer, Z. Rosenwaks, O. Elemento, N. Zaninovic, I. Hajirasouliha, Deep learning enables robust assessment and selection of human blastocysts after in vitro fertilization, *npj Digit. Med.* 2 (21) (2019) 1–9.
- [31] P. Thirumalaraju, M.K. Kanakasabapathy, C.L. Bormann, R. Gupta, R. Pooniwala, H. Kandula, I. Souter, I. Dimitriadis, H. Shafiee, Evaluation of deep convolutional neural networks in classifying human embryo images based on their morphological quality, arXiv (2020), arXiv:2005.10912
- [32] S. Wang, C. Zhou, D. Zhang, L. Chen, H. Sun, A deep learning framework design for automatic blastocyst evaluation with multifocal images, *IEEE Access* 9 (2021) 18927–18934.
- [33] M.F. Kragh, J. Rimestad, J. Berntsen, H. Karstoft, Automatic grading of human blastocysts from time-lapse imaging, *Comput. Biol. Med.* 115 (2019) 103494.
- [34] L. Lockhart, P. Saeedi, J. Au, J. Havelock, Automating embryo development stage detection in time-lapse imaging with synergic loss and temporal learning, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2021, pp. 540–549.
- [35] M.K. Kanakasabapathy, P. Thirumalaraju, C.L. Bormann, R. Gupta, R. Pooniwala, H. Kandula, I. Souter, I. Dimitriadis, H. Shafiee, Deep learning mediated single time-point image-based prediction of embryo developmental outcome at the cleavage stage, arXiv (2020), arXiv:2006.08346
- [36] P. Thirumalaraju, J. Hsu, C. Bormann, M. Kanakasabapathy, I. Souter, I. Dimitriadis, K. Dickinson, R. Pooniwala, R. Gupta, V. Yogesh, et al., Deep learning-enabled blastocyst prediction system for cleavage stage embryo selection, *Fertil. Steril.* 111 (4) (2019) e29.
- [37] C.L. Bormann, M.K. Kanakasabapathy, P. Thirumalaraju, R. Gupta, R. Pooniwala, H. Kandula, et al., Performance of a deep learning based neural network in the selection of human blastocysts for implantation, *ELife* 9 (2020) 1–14.
- [38] R.M. Rad, P. Saeedi, J. Au, J. Havelock, Predicting human embryos' implantation outcome from a single blastocyst image, in: Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. 2019, 2019, pp. 920–924.
- [39] A. Targosz, P. Przystałka, R. Wiaderkiewicz, G. Mrugacz, Semantic segmentation of human oocyte images using deep neural networks, *Biomed. Eng. Online* 20 (1) (2021) 1–26.
- [40] S. Firuzinia, S.M. Afzali, F. Ghasemian, S.A. Mirroshandel, A robust deep learning-based multiclass segmentation method for analyzing human metaphase ii oocyte images, *Comput. Methods Programs Biomed.* 201 (2021) 105946.
- [41] M. Nagaya, N. Ukita, Embryo grading with unreliable labels due to chromosome abnormalities by regularized pu learning with ranking, *IEEE Trans. Med. Imaging* 41 (2) (2021) 320–331.
- [42] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.
- [43] Criterios ASEBIR de Valoración Morfológica de Oocitos, Embriones Tempranos y Blastocistos Humanos, 2015.
- [44] S. Sukhbaatar, J. Bruna, M. Paluri, L. Bourdev, R. Fergus, Training Convolutional Networks with Noisy Labels, arXiv (2014).
- [45] Z. Zhang, M.R. Sabuncu, Generalized Cross Entropy Loss for Training Deep Neural Networks with Noisy Labels, arXiv (2018).
- [46] O.J. Hénaff, A. Srinivas, J. De Fauw, A. Razavi, C. Doersch, S.M.A. Eslami, A.v.d. Oord, Data-Efficient Image Recognition with Contrastive Predictive Coding, arXiv (2019).
- [47] Y. Tian, D. Krishnan, P. Isola, Contrastive Multiview Coding, arXiv (2019).
- [48] T. Chen, S. Kornblith, M. Norouzi, G. Hinton, A Simple Framework for Contrastive Learning of Visual Representations, arXiv (2020).
- [49] K. He, H. Fan, Y. Wu, S. Xie, R. Girshick, Momentum Contrast for Unsupervised Visual Representation Learning, arXiv (2019).
- [50] P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, D. Krishnan, Supervised Contrastive Learning, arXiv (2020).
- [51] S.J. Pan, Q. Yang, A survey on transfer learning, *IEEE Trans. Knowl. Data Eng.* 22 (10) (2009) 1345–1359.
- [52] R. Caruana, Multitask learning, *Mach. Learn.* 28 (1) (1997) 41–75.
- [53] D.L. Silver, R.E. Mercer, The task rehearsal method of life-long learning: overcoming impoverished data, in: Advances in Artificial Intelligence, Springer, Berlin, Germany, 2002, pp. 90–101.
- [54] S. Zhang, F. Sun, N. Wang, C. Zhang, Q. Yu, M. Zhang, P. Babyn, H. Zhong, Computer-aided diagnosis (CAD) of pulmonary nodule of thoracic CT image using transfer learning, *J. Digit. Imaging* 32 (6) (2019) 995–1007.
- [55] Y. Zhou, B. Wang, L. Huang, S. Cui, L. Shao, A Benchmark for Studying Diabetic Retinopathy: Segmentation, Grading, and Transferability, arXiv (2020).
- [56] R. del Amor, L. Launet, A. Colomer, A. Moscardó, A. Mosquera-Zamudio, C. Monteagudo, V. Naranjo, An attention-based weakly supervised framework for spitzoid melanocytic lesion diagnosis in WSI, arXiv (2021).
- [57] A. Rahimi, B. Recht, et al., Random features for large-scale kernel machines, in: NIPS, vol. 3, Citeseer, 2007, p. 5.
- [58] F. Schroff, D. Kalenichenko, J. Philbin, Facenet: a unified embedding for face recognition and clustering, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 815–823.
- [59] P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, D. Krishnan, Supervised contrastive learning, arXiv preprint arXiv:2004.11362(2020).
- [60] L. Van der Maaten, G. Hinton, Visualizing data using t-SNE, *J. Mach. Learn. Res.* 9 (Nov) (2008) 2579–2605.
- [61] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-cam: visual explanations from deep networks via gradient-based localization, in: 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 618–626.