Embryo Evaluation Based on ResNet with AdaptiveGA-optimized Hyperparameters

Wenju Zhou[1], Xiaofei Han[1], Yuan Xu[2], Rongfei Chen[1], Zhenbo Zhang[2*]

[1] Shanghai Key Laboratory of Power Station Automation Technology, Shanghai University, China
[2] Department of Obstetrics and Gynecology, Shanghai General Hospital, Shanghai Jiaotong University, China
zhouwenju@shu.edu.cn, hanxiaofei@shu.edu.cn, xuyuan0218@163.com, rongfei_chen@163.com,
zhangzhenbozzb@aliyun.com

Abstract

In vitro fertilization (IVF) embryo evaluation based on morphology is an effective method to improve the success rate of transplantation. Although convolutional neural networks (CNNs) have made great achievements in many image classifications, there are still great challenges in accurately classifying embryos due to the insufficient samples, interference of exfoliated cells, and inappropriate hyperparameter configuration in the classification network. In this paper, a residual neural network optimized by the adaptive genetic algorithm is proposed to evaluate embryos. Firstly, a novel algorithm for extracting the region of interest (ROI) is embedded in the preprocessing part of the model to eliminate exfoliated cells close to the embryo. Secondly, several kinds of specific transformation methods are established to expand the dataset based on the symmetry of embryos. In addition, an adaptive genetic algorithm is adopted to search for optimal hyperparameters. Experiments on the data set provided by Shanghai General Hospital show that the algorithm has an excellent performance in embryo evaluation. The accuracy of our model is 86.4%, the recall is 88.4%, and the AUC is 0.93. Our results indicated that the proposed model can effectively improve the classification performance of ResNet, and thus achieve the clinic requirements of embryo evaluation.

Keywords: Neural network, Embryo evaluation, Genetic algorithm, Image processing

1 Introduction

For infertile couples, in vitro fertilization (IVF) is one of the effective treatments to help them born a baby [1-2]. In many cases, the embryo in vitro fertilization will be transferred on day-2 or day-3. Before that, the embryologist will select the embryos to transfer or freeze based on the morphology evaluation criteria performed by the IVF laboratory [3-4]. Recently, some researchers found that the embryo morphology in day-2 and day-3 should be seriously considered [5]. However, the available grading systems rely on the visual information obtained by the embryologist are susceptible to differences among observers (to some extent, differences within the observer). Inevitably, this uncertainty will influence the decision about which embryo to be selected for transplantation, and directly affect the result of IVF. To improve the success rate of embryo transfer and reduce the risk of surgery, it is necessary to utilize the Machine Learning (ML) method to continually evaluate the quality of embryos before transplantation or freezing.

A lot of literatures focus on automatic embryo evaluation. Some researchers use digital image processing technology to automatically detect and classify embryos. However, traditional digital image processing cannot evaluate embryos comprehensively, and it is usually regarded as a semi-automatic auxiliary technology that helping embryologists to find the morphological characteristics of embryos [6]. Many solutions adopt deep learning methods, such as adopting the Deep Neural Network (DNN) to grade embryo quality. Although the development and application of deep learning have improved the performance of embryo evaluation [7-8], evaluating the IVF embryos effectively is still a challenge due to various reasons. Firstly, insufficient sample, as we all know, training neural networks requires a large number of samples [9], but these samples are hard to prepare because it requires experienced embryologists to do manual annotation. And the useable dataset in this field is scarce. Secondly, the degree of cell division, embryo symmetry, and cell size of the embryo are the essential morphology criteria to evaluate the quality of an embryo. In many cases, there will be a lot of exfoliated cells that have nothing to do with the embryo quality in the sample images, but the exfoliated cells would have a great influence on neural networks to focus on the feature of the embryo. Some typical embryo images with and without exfoliated cells are indicated in Figure 1. Finally, for most neural networks, the weights are always be learned and updated, but the hyperparameters in the network will not be learned in the process of training. They are fixed at the initialization of the model. Different hyperparameter combinations will have a greater impact on classification performance, and there are many kinds of the combination when faced with multiple hyperparameters. A lot of literature also introduces different hyperparameter optimization algorithms, such as Bayesian optimization, heuristic algorithm, random search optimization algorithm, and many other algorithms.

In this paper, a residual neural network model optimized by an adaptive Genetic Algorithm (GA) was proposed for embryo evaluation. At the same time, to make sure the model can extract and learn the features in the embryo image effectively, A novel ROI extract algorithm is adopted to extract the embryo region without the exfoliated cells. In this way, the feature learning ability of the ResNet network could be fully utilized. To expand the training dataset, some data augmentation technology such as flip, rotation in different

*Corresponding Author: Zhenbo Zhang; E-mail: zhangzhenbozzb@aliyun.com
DOI: 10.53106/160792642022052303011

angles, and blur were taken to generate enough samples. ResNet50 is used as a classification network [10]. The residual network is more suitable than other networks when the samples are not very enough, because it uses a "short cut" to skip one or more layers, which can be constructed deeper network. We also adopt the adaptive Genetic Algorithm for global optimization of the hyperparameters in the residual network. The update of each generation of individuals in the GA ensures that the hyperparameter combination is the current optimal value. This will always improve the classification performance of the residual network. We validated our algorithm model on the dataset provided by Shanghai General Hospital. And achieved the latest performance level (ie AUC of 0.93).

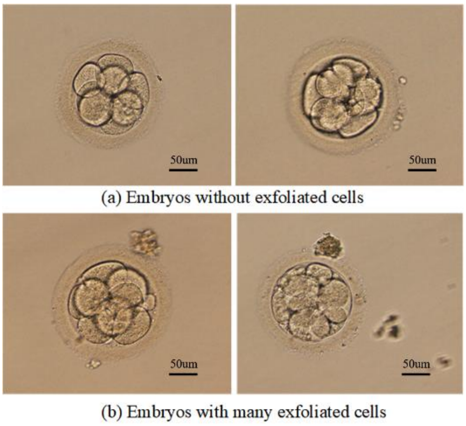


Figure 1. Embryos with and without exfoliated cells

2 Related Work

2.1 ResNet and Its Variants

The depth of the neural network is constantly increasing through simple stacking, which causes the gradient to become infinitely small when it reaches the foremost network layer in backpropagation. This is the reason that the performance of the neural network is saturated or even drops rapidly when there are too many layers. Before ResNet, many methods attempted to solve the problem of vanishing gradients [11-12]. For example, L. Shao, proposed an auxiliary loss in a middle layer as extra supervision [11], but none seemed to tackle the problem once and for all. The proposal of the deep residual network was the breakthrough in the field of computer vision and deep learning. ResNet solves the problem of the difficulty of training deep CNN models. In 2014, VGG only had 19 layers, and in 2015 ResNet reaches 152 layers. The residual network can be effectively applied in image classification, target detection, and semantic segmentation, and the robustness of the residual network has been proved by various visual recognition tasks and non-visual tasks of designing speech and language. As ResNet has received more and more attention among researchers, some people have proposed variants of ResNet, as shown in Figure 2. Xie,s. proposed a variant of ResNet that is named ResNeXt with the following building block. In this variant, the outputs of different paths are combined by adding them together, while the different paths have the same topology [13].

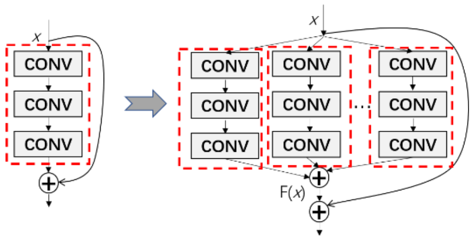


Figure 2. ResNet variant structure

2.2 Embryo Quality Evaluation

Researches have shown that the day-5 or day-6 embryos could also be considered for transplantation and freezing. However, the embryos on day-2 and day-3 are most suitable when there are several equivalent high-quality embryos available [5], After selecting high-quality embryos, sub-optimal embryos can be left for frozen storage. Effective evaluation of embryos is one of the key factors that affect IVF technology.

One way to evaluate the embryo quality is by analyzing the morphology feature, like cleavage stage, embryo size, and embryo symmetry. This method is more convenient and concise, However, due to many factors that need to be considered, it is often impossible to have high accuracy. And because of the subjective differences of different embryologists, it is difficult to guarantee the consistency of the classification results. Therefore, it has been proposed to use computer technology to assist embryologists in completing embryo grading. For example, Sujata N Patil et al. proposed an enhanced template matching technology attempt to automatically detect and classify cells in embryos [4]. However, the method cannot effectively find all cell edges accurately. This type of algorithm is generally regarded as the preprocessing stage of the embryo classification algorithm. In recent years, the development of Deep Learning has been successfully used in many fields. Some researchers utilized Machine Learning to evaluate the embryo quality. For example, Morales proposed an approach to select the good embryos based on Bayesian Classifier and shown preliminary classification outcome [14], but they ignored the advantage of the traditional method in extracting the embryo features. Pegah Khosravi implemented an AI approach based on deep neural networks (DNNs) to select the highest quality embryos [15]. However, this method requires a large training data to support it. After obtaining the best parameter set on the ImageNet database, they conducted 50,000 iterations of training on 50,392 images of 10,148 embryos. Tsung-Jui Chen. applied Convolutional Neural Network (CNN) on embryo images, using ResNet50 architecture to fine-tune the amount of ImageNet parameters [16]. This method also requires a lot of training data to support.

2.3 Hyperparameters Optimization

The research and application of machine learning in the past decades has solved many problems in the field of academic and application. However, the design and training of neural networks are critical and complex. Automated hyper-parameter optimization (HPO) has become a popular topic.

Hyperparameters are systematically categorized into structure related and training related. Optimizer is one of the hyperparameters related to model training. The most adopted optimizer is stochastic gradient descent (SGD) with momentum, AdaGrad, RMSprop, and Adam are also the alternative optimizers. In addition to the optimizer, there also many hyperparameters that are critical to the training of the model, such as, batch size and learning rate. The most typical hyperparameters related to the model structure are the number of hidden layers and the width of neural networks.

The common search algorithm for hyperparameter optimization includes Grid search, Random search, Bayesian Optimization, Genetic Algorithm, and meta-learning. David Gonzalez-Cuautle proposed a botnet detection model based on grid search to optimize hyperparameters [17]. Theopilus Bayu Sasongko compared the optimization of grid search parameter with the GA and results showed that the GA accuracy is better than grid search [18]. Rafael G. Mantovani used the random search to adjust the hyper-parameters of SVMs. M H M Tarik optimizing the ANN hyperparameters using Bayesian optimization and the result shown better than random search [19]. Maryam Parsa proposed a Bayesian-based hyperparameter optimization approach for spiking neuromorphic systems [20]. The genetic algorithm is a heuristic algorithm to optimize hyperparameters. Ji-Hoon Han used the GA to optimize hyperparameters and obtain the proper verification time and accuracy [21]. The GA is also used to optimize the hyperparameters of DNN and Long Short-Term Memory network (LSTM) [22-23]. meta-learning is also an efficient method for hyperparameter optimization. Khac-Hoai Nam Bui used meta-learning to tune hyperparameter and improved the automatic learning process and reduced time-consuming tasks [24]. Matthias Feurer based on meta-learning to initialize Bayesian hyperparameter optimization [25].

3 Method

The proposed embryo evaluation residual model optimized by genetic algorithm falls into three aspects: (1) An ROI extraction algorithm designed by us is embedded in the preprocessing part of the model. (2) Several kinds of targeted transformation methods are established to expand the dataset. (3) In the training process, we adopt the adaptive Genetic Algorithm to optimize the hyperparameters of the model. The architecture of this model is shown in Figure 3. Now we delve into the details.

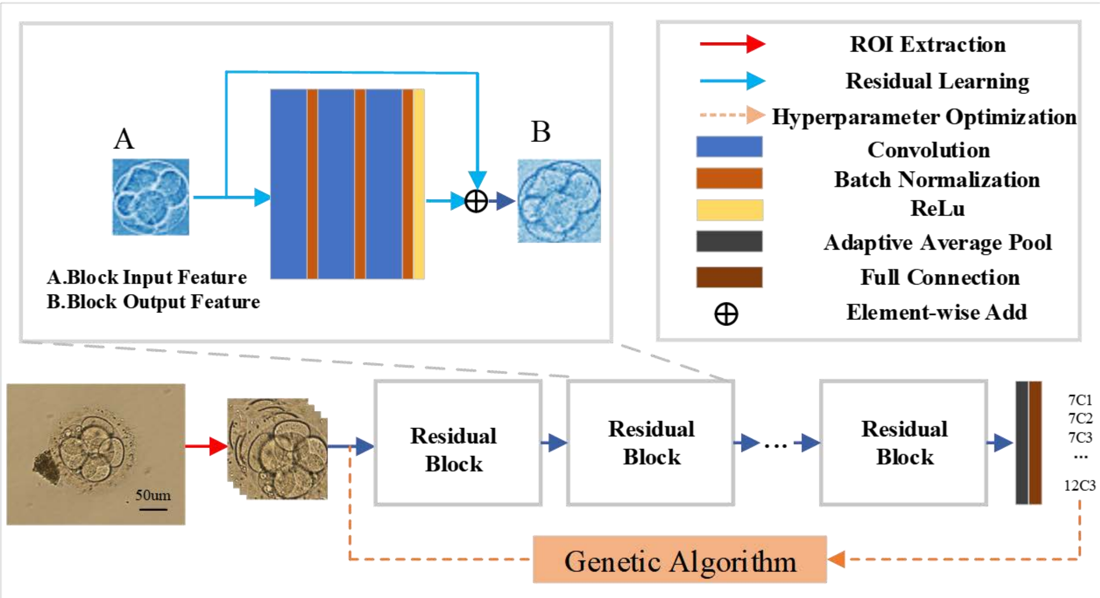


Figure 3. Architecture of the proposed model

3.1 Embryo Extraction

A common algorithm for embryo extraction is template matching, However, this method has poor efficiency, if this method is embedded into the preprocessing part of the neural network, it would cause a great waste of resources and consume a lot of time. Another embryo extraction algorithm is to analyze the connected domains in the image, but when the exfoliated cells and the embryo overlap in space, the two parts are connected when displayed in an image, as shown in Figure 4. They are in the same connected domain. It is difficult to separate them by analyzing the connected domains. Another challenge is to distinguish the minimum contour of the embryo and the inner contour of ZP (Zona Pellucida, a glycoprotein membrane encapsulating the oocyte and early embryo [1]).

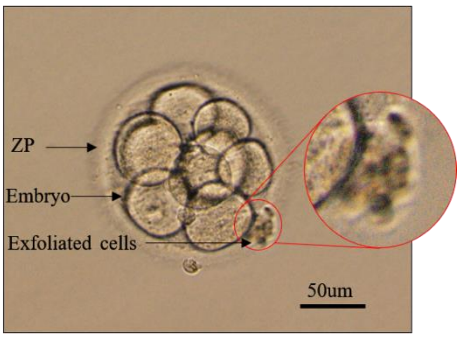


Figure 4. Relative position of exfoliated cells

The proposed ROI extraction algorithm can extract meaningful region in the image, eliminate the interference of exfoliated cells, even the exfoliate cells and embryo are closely connected. At first, an adaptive binarization method is adopted on the original image P to get a binary image Pb. Since the background of the embryo culture dish is relatively single, in the image after the binarization process, both the embryo and the exfoliated cells can be clearly distinguished from the background. Then the next task is to further distinguish the embryo and exfoliated cells and extract the embryo. It can be seen from the above that the difficulty is that we need to overcome is to find and delete the exfoliated cells area. Because the size of embryo culture dishes is the same, and the microscope magnification is fixed before the observation process. Therefore, during the imaging process, the embryo occupies the field of view with a small fluctuation range.

Therefore, we design a separation method to distinguish embryo and exfoliated cells. First, we obtain all contours in the image through the edge detection algorithm. Then we choose the largest contour as the contour of the embryo, but in this step, there is a serious interference from the contour of ZP. As we can see, the contour of the embryo is clearer and irregular compared with ZP, so we design an adaptive edge detection algorithm to get the exact contour of the embryo. Once we get the exact contour of the embryo, we could fit the smallest bounding rectangle of the contour. Finally, we can extract the rectangle by performing a perspective transformation on the four corner points of the smallest bounding rectangle. The steps of our ROI extraction algorithm in Table 1. The processing effect is shown in Figure 5.

Table 1. ROI extraction algorithm

ROI Extraction Algorithm
Step1: Input original image P, and initialize thresholds of channel B, G, R.
Step 2: Call adaptive binarization method to derive binary image P_b.
Step 3: Call adaptive edge detection method on binary image P_b to get the contour of embryo.
Step 4: Fit the smallest bounding rectangle of the contour of embryo.
Step 5: Extract the final image P_f with the four cornor points of the smallest bounding rectangle.

The traditional digital image processing technology is only a preprocess of the dataset. It can not auto classify an embryo sample accurately. so it is only used the traditional image processing technology is insufficient. But we can not ignore it in the task of image classification. The traditional image processing technology is used to determine the target range of the image, enhance the embryonic feature gap at different levels, simplify the information amount of the dataset, etc. provides support for the training in the neural network.

3.2 Hyperparameters in Neural Networks

The hyperparameter optimization problem is a global optimization black box problem. The ultimate goal of a typical optimization method is to find an algorithm model with a minimum loss function and obtain the hyperparameter configuration of the algorithm. The neural network will have many model parameters to set before training, but once these parameters are determined, they cannot be changed during the training process. For different parameters, there will be different ranges of value spaces. Finding out a set of suitable hyperparameters quickly is crucial for model training. Hyperparameter optimization in neural networks is a complex optimization problem, which contains both continuous variables and discrete categorical variables. There are many search algorithms for hyperparameter optimization, such as the genetic algorithm. In our algorithm, we use ResNet50 as our model, and the hyperparameters need to be optimized as follow:

(1) The optimizer in our model is stochastic gradient descent (SGD), so it contains two hyperparameters that need to be optimized: initial learning rate and momentum.

(2) There are two hyperparameters in the learning rate adjustment mechanism, the learning rate decay period (Step) and the learning rate decay multiplication factor (γ).

(3) The number of the epoch.

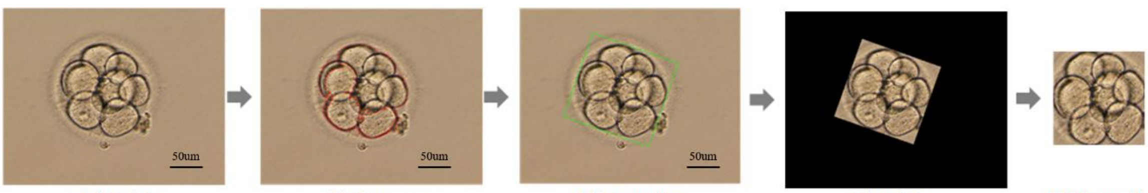


Figure 5. The process of ROI algorithm

These five hyperparameters constitute a five-dimensional hyperparameter vector = [η, m, Step, γ, Epoch]. For the training of this hyperparameter vector, the loss function of the neural network is still used as the optimization objective function, In the above formula, it is the hyperparameter vector that makes the model loss function obtain the minimum value, which represents the parameter selection space. It is the loss function of the model.

3.3 ResNet Optimized by Adaptive Genetic Algorithm

Based on the hyperparameter optimization process mentioned above, we delve into the details. Figure 6 shows the general process of the genetic algorithm optimizing the hyperparameter of residual neural networks [26].

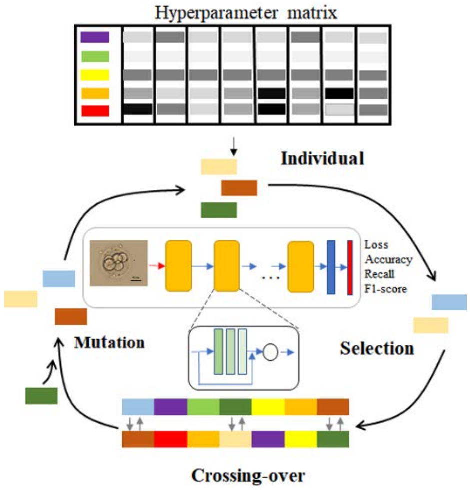


Figure 6. The process of genetic algorithm optimization

Encoding is the primary problem for the genetic algorithm to optimize neural network hyperparameters [27]. That is the process of expressing the hyperparameter combination of the neural network into the search space that the genetic algorithm could handle. Each hyperparameter configuration scheme corresponds to an individual $S = (s_1, s_2, \cdots, s_i, \cdots, s_n)$. n is the length of genes. Gene s_i represents the value of each hyperparameter.

The gene value of each chromosome in the first-generation population is randomly selected in the whole search spaces, that is, each parameter in the hyperparameter vector is randomly selected as an initial value respectively. Through this random method, initial values are set for all individuals in the first-generation population to complete the population initialization. For each element in the population, the value is used as the corresponding parameter to determine the current neural network training rules, and the residual neural network model is trained on the premise of the same dataset and records the optimal loss value and accuracy.

Now we introduce the main components of our improved genetic algorithm. During the design and testing of our algorithm, we implemented a few strategies that aimed at helping the exploration and development of our algorithm.

The operations performed on individuals in the algorithm include selection, crossover, and mutation. We use the optimal individual preservation method to perform selection operations [28]. According to the fitness function value (In our algorithm, the reciprocal of the loss value as the fitness function), the best individuals in the parent population are selected to replace the worst individuals in the offspring population. The number of selected individuals is determined by selection parameters p_s.

Since each hyperparameter has a different value range, only genes at the same position can be exchanged when performing a crossover operation. The details of the crossover operation as follow. Firstly, according to the crossover rate p_c of each individual to select two different individuals from the population as the father and mother. Then, generating an integer C which greater than 0 and less than n (the gene length of an individual) as the cross interval. Then the genes of father and mother will exchange for every interval of C genes, that is $s_1, s_{1+C}, s_{1+2C}, \cdots$.

The mutation operation, that is, according to the crossover rate P_m of each individual to select an individual in the optimal first half of the population, and randomly selecting a parameter value of the individual to reassign the new value generated by the mutation operation.

In the standard genetic algorithm, the crossover and mutation rate of each individual in the population share the same fixed value, which cannot reflect the evolution process of the population. Each individual should have its crossover rate P_c and mutation rate P_m and should automatically change with the fitness of the population. When the fitness of individuals in the population tends to be consistent or local optimal, P_c and P_m should be increased to jump out of the local optimal; and when the fitness of the population is relatively dispersed, P_c and P_m should be reduced to facilitate the survival of good individuals.

$$p_c = \begin{cases} p_{c1} & f' < f_{avg} \\ \dfrac{p_{c2} + (p_{c1} - p_{c2}) \times (f' < f_{avg})}{f_{max} - f_{avg}} & f' \geq f_{avg} \end{cases} \quad \textbf{(1)}$$

$$p_m = \begin{cases} p_{m1} & f' < f_{avg} \\ \dfrac{p_{m2} + (p_{m1} - p_{m2}) \times (f' < f_{avg})}{f_{max} - f_{avg}} & f' \geq f_{avg} \end{cases} \quad \textbf{(2)}$$

Among them: f_{avg} represents the average fitness value of each generation of the group, f_{max} represents the largest fitness value in the group, f' represents the larger fitness value of the two individuals to be crossed, f represents the fitness value of the individual to be mutated [29].

After the new generation of the population is generated, the gene of each individual will be decoded to the value of hyperparameter and sent to the residual neural network for training again until the termination condition is reached, that is, the number of genetic iterations reaches the maximum number of genetic iterations. Finally, the optimal hyperparameter and model will be output.

4 Experiment and Results

4.1 Dataset

The algorithm we proposed was trained and tested on the embryo image dataset provided by Shanghai General Hospital. The dataset consisted of train and test parts. The train dataset includes a total of 442 embryo micro-pictures. These embryos are divided into 20 categories according to the number of cells and their morphological characteristics: {‘5C2’, ‘6C2’, ‘6C3’, ‘7C1’, ‘7C2’, ‘7C3’, ‘7C4’, ‘8C1’, ‘8C2’, ‘8C3’, ‘9C1’, ‘9C2’, ‘9C3’, ‘10C1’, ‘10C2’, ‘10C3’, ‘12C1’, ‘12C2’, ‘12C3’, ‘14C2’}. Take ‘14C2’ as an example. This category means that the current embryo contains 14 cells in total, and the level is 2. Typical embryo images of some categories in the dataset are indicated in Figure 7. The photomicrograph in the fifth column of third row is derived from a cryopreservation embryo. The test dataset includes total of 126 embryo photomicrographs in 20 different categories.

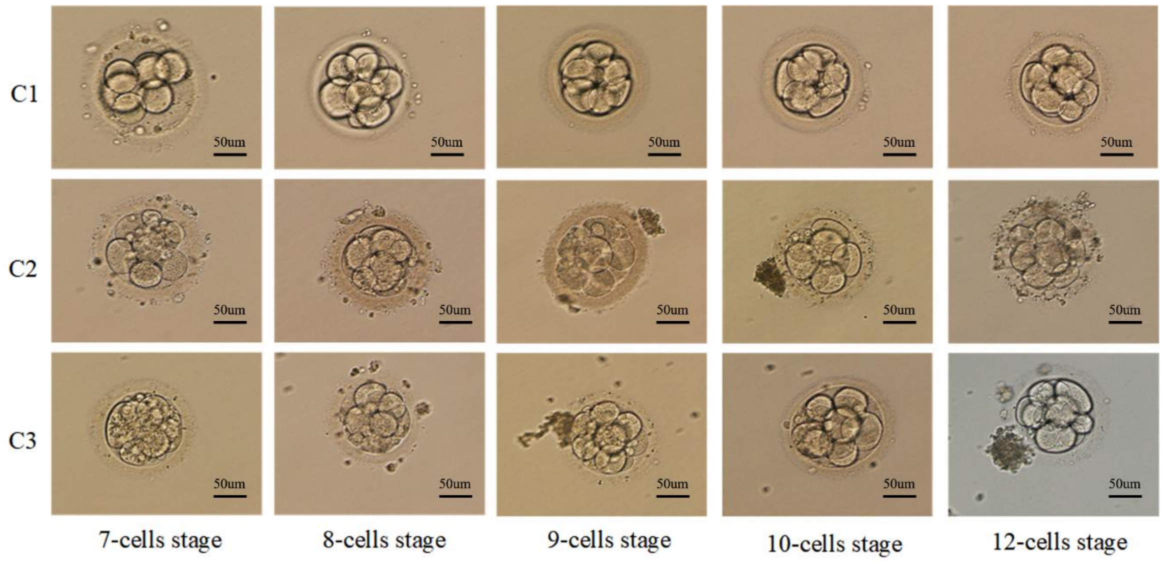


Figure 7. Categories of embryos

4.2 Implementation

To make the residual network extract the morphological features of the embryo image effectively, firstly, we use the ROI extraction method to preprocess all the embryo images, extract the part that only contains the embryo, and expand the dataset in different transform methods. The transform methods include rotation at different angles, horizontal flip, vertical flip, translation in different directions and distances, adding noise in different degrees, and blur operations in different degrees. Finally, we have expanded the train dataset by 13 times, and the images in train dataset have increased to 5746. The preprocessing process of the train dataset is indicated in Figure 8. It should be noticed that the network does not conduct the data augmentation in the test mode.

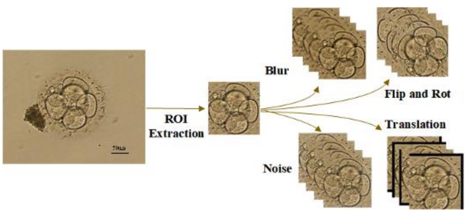


Figure 8. The process of expanding dataset

A total of five parameters in the ResNet50 are optimized using genetic algorithms. Before optimization, the parameter spaces of these five parameters must be determined. The different permutations and combinations between the hyperparameter values will greatly affect the performance of the residual network. Table 2 shows the hyperparameters that need to be optimized.

Table 2. Hyperparameters to be optimized

Symbol	Hyperparameter
η	learning rate
m	momentum in optimizer
step	adjustment step in learning rate scheduler
γ	learning rate decay multiplication factor
epoch	number of iterations

Stochastic gradient descent (SGD) is the most commonly used algorithm for machine learning parameter optimization. During the solution of this algorithm, two parameters need to be set: learning rate and momentum value. The learning rate of the learning rate control algorithm is too fast, which may lead to non-convergence, and too slow will waste time and resources. Momentum accelerates the current optimization by accumulating previous gradients and combining the current gradients. The learning rate adjustment algorithm is Step LR. There are also two hyperparameters in the algorithm that need to be optimized, namely the learning rate decay period (Step) and the learning rate decay multiplication factor (γ). The learning rate decay period controls how many epoch learning rates are to be dynamically decayed, and the learning rate decay factor controls the amplitude of each decay. Epoch controls how many times a total sample needs to be trained.

4.3 Evaluation Metrics

(1) Qualitative Evaluation: We extract the typical feature map after the first residual block convolution in the forward calculation process, to observe the extraction of features such as the contour and shape of the embryo in the sample image by the upper network

(2) Quantitative Evaluation: To quantitatively evaluate the optimization effect of the proposed embryo extraction algorithm and genetic algorithm on ResNet50, we used the accuracy, recall, F1-measure, and area under the receiver operating characteristic curve (AUC) as performance metrics, which are defined as follow:

$$accuracy = \frac{TP + TN}{TP + FN + TN + FP} \tag{3}$$

$$recall = \frac{TP}{TP + FN} \tag{4}$$

$$precision = \frac{TP}{TP + FP} \tag{5}$$

$$F1-measure=\frac{2\times precision\times recall}{precision+recall} \quad (6)$$

$$AUC=\int_0^1 t_{pr}(f_{pr})df_{pr}=P(X1>X0) \quad (7)$$

Where TP, FN, TN, and FP represent true positive, false negative, true negative, and false positive, t_{pr} is the true positive rate, f_{pr} is the false positive rate, and X_0 and X_1 are the confidence scores for a negative and positive instance, respectively. The AUC value describes the probability that a classifier ranks a randomly chosen positive instance higher than a randomly chosen negative one.

The confusion matrix is the basis for drawing the ROC curve to calculate AUC. It separately counts the proportions of the correct and incorrect sample sizes in each category. All the results are displayed in a graph, which is a measure of the accuracy of a classification model. The most basic. the most intuitive method.

4.4 Ablation Studies

In the proposed algorithm, the ROI extraction algorithm and GA for searching hyperparameters are adopted to optimized the model. To understand which part are critical for classification performance, we analyzed results on the test dataset for each of the proposed part. ROI was described in Section 3.1. The optimization of hyperparameters by GA was described in Section 3.2. The base classification model is ResNet50. Each group of results obtained from the models with same parameter settings including SGD optimizer, cross-entropy loss function, initial learning rate, and a maximum number of epochs. The models which adopted GA had the same hyperparameter search space and initial population.

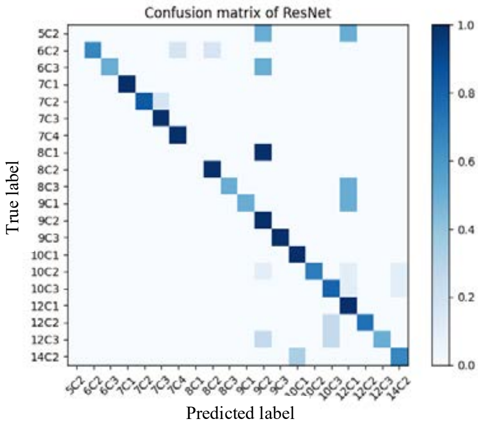
Table 3. Results of ablation experiments

Method	ROI	GA	Precision	Recall	F1	Acc	AUC
1	x	x	0.7351	0.7208	0.6859	0.7301	0.86
2	√	x	0.8353	0.8621	0.8742	0.8416	0.92
3	x	√	0.8364	0.8584	0.8567	0.8529	0.92
4	√	√	**0.8934**	**0.88847**	**0.8768**	**0.8641**	**0.93**

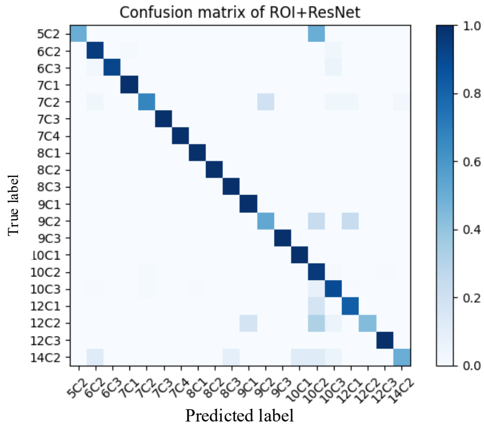
We start by looking at results from the ResNet50 with ROI and without GA (Table 3 rows 1 and 2). The second and third columns of the table are used to mark whether the current experiment uses the ROI extraction algorithm or the GA optimization algorithm. The next five columns represent the evaluation indicators introduced in the previous section. According to experiments 1 and 2, it can be seen that the classification performance of the model after using the ROI preprocessing algorithm has been significantly improved, with precision raised from 0.73 to 0.83, the figures for recall and F1 increased by 0.14 and 0.19 respectively, accuracy increased to 0.84, and the AUC increase is obvious from 0.86 to 0.92. Row 3 in Table 3 includes the results from the ResNet50 without ROI but with the GA to optimize the hyperparameters. It almost has the same pattern with row 2, with the AUC is 0.92 which is same with that in experiment 2. What can be conclude is that both ROI algorithm and the optimization of GA have a critical contribution to the improvement of classification performance. The results from the model adopted both ROI and GA parts are showed on the last row in Table 3. All the figures saw the biggest increase, with the higher AUC, precision, recall, F1, and accuracy than the results of any other experiments. We consider that the ROI part could improve the attention to embryo region of the model, because the exfoliated cells around the embryo did cause inevitable interference to the network. On the other hand, general ResNet has a significant limitation to the searching space of hyperparameters. The proposed Adaptive Genetic Algorithm could keep searching for the optimal solution in each iteration. So we believe that both ROI and Adaptive Genetic Algorithm parts are effective.

To compare the gap between these four methods more intuitively, the confusion matrix of the four comparison experiments is shown in Figure 9. It is apparent that the residual network that uses both the ROI preprocessing algorithm and the genetic algorithm has fewer classification errors. The AUC can reach 0.93 when using the ROI extraction algorithm and genetic algorithm to optimize the residual network classification.

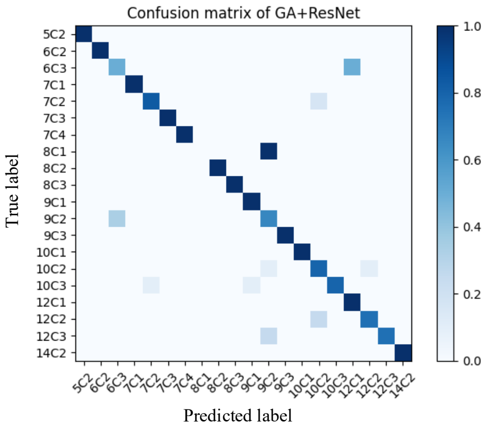
The model optimized by adaptive GA has scored each category during the test, which can effectively separate different categories. Some results of classification are shown in Table 4. The photomicrographs in row 1-2 are level C1 and C2 respectively. Particularly, the photomicrograph in row 1 is derived from a cryopreservation embryo. Photomicrographs in row 3-6 are level C3. In addition, we illustrated the morphological features. The label of embryo in row 3 of Table 4 is 8C3. There are 8cells in the embryo. But this embryo has an asymmetric cells layout and some of the cells are not the standard circles, the level of this embryo is C3. The embryo in row 4 labeled 9C3. There are 9 cells in this embryo. The sizes of the cells are various, and thre are many fragments in the embryo. So the level of the embryo is C3. The cells edge of embryo in row 5 are terrible blurred, there is no doubt that this embryo is level C3. The embryo in the last row has the same situation with row 5, and the level is C3 too. Besides, embryos in row 2 and row 6 suffered the interfere of impurity cells.



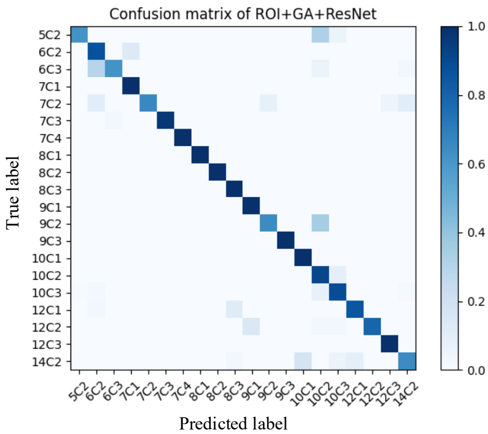
(a) Confusion matrix of ResNet



(b) Confusion matrix of ROI+ResNet



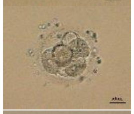
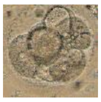
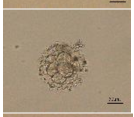
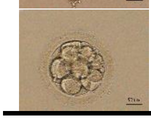
(c) Confusion matrix of GA+ResNet



(d) Confusion matrix of ROI+GA+ResNet

Figure 9. Confusion matrixes

Table 4. Several results of classification

Method	ROI	label	5C2	6C2	6C3	7C1	7C2	7C3	7C4	8C1	8C2	8C3	9C1	9C2	9C3	10C1	10C2	10C3	12C1	12C2	12C3	14C2
		7C1	0.17	0.88	0.5	1.0	0.98	0.1	0.0	0.28	0.97	0.07	0.03	0.64	0.06	0.17	0.6	0.94	0.65	0.58	0.47	0.02
		10C2	0.53	0.06	0.34	0.17	0.67	0.45	0.09	0.1	0.06	0.04	0.08	0.44	0.66	0.43	0.99	0.13	0.34	0.04	0.43	0.87
		8C3	0.20	0.81	0.34	0.15	0.09	0.24	0.17	0.16	0.21	0.96	0.08	0.46	0.43	0.84	0.67	0.56	0.47	0.02	0.37	0.43
		9C3	0.54	0.75	0.35	0.29	0.05	0.73	0.54	0.07	0.06	0.71	0.46	0.63	0.15	0.98	0.48	0.09	0.51	0.19	0.27	0.05
		10C3	0.14	0.05	0.24	0.08	0.06	0.27	0.09	0.17	0.31	0.17	0.31	0.07	0.42	0.18	0.68	1.0	0.75	0.16	0.43	0.06
		12C3	0.34	0.53	0.88	0.39	0.67	0.56	0.06	0.27	0.09	0.53	0.05	0.9	0.22	0.94	0.3	0.94	0.79	0.04	0.96	0.61

4.5 Visualization of Feature Map

The size of the original embryo image in the dataset is very large, and the background area occupies a large proportion of the entire image. If we train our model with original images, it would not only increase the calculation, and more importantly, would affect the final classification results. The ROI extraction algorithm we designed for the dataset can distinguish embryos from exfoliated cells effectively, so improve the accuracy of our model. We also embed the entire data preprocessing algorithm into the preprocessing module of the model, so that the entire training and testing is completer and more concise.

To validate whether the model extracts the feature of the embryo, like the edge and the impurities inside the embryo. We visualized the feature maps obtained by the first residual layer of the trained model in Figure 10. the feature maps of good embryos in the left column show the clear cell edges and symmetrical embryo structure. However, in the feature maps of poor embryos in the right column, the cell edges are blurred and the embryo structure is asymmetric.

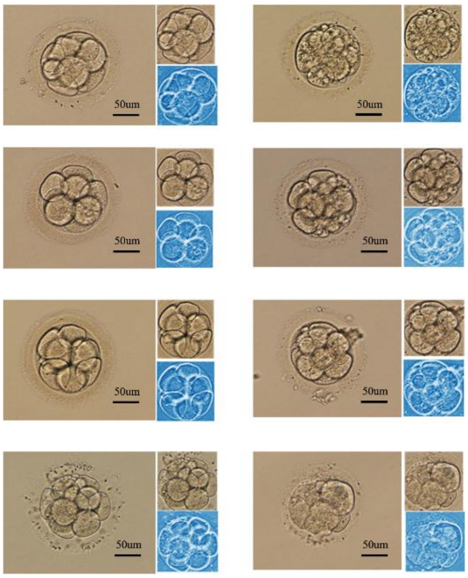


Figure 10. Feature maps of some images

5 Discussion

5.1 Comparing to Other Hyperparameter Optimization Methods

Hyperparameters are a set of parameter data that define the network structure and control the training process. Hyperparameters are generally determined before the network starts training. However, it is difficult to select a set of optimal hyperparameter values at the beginning. Normally, the optimization algorithm is used to select a set of optimal hyperparameters for the model. Hyperparameter optimization algorithms mainly include a meta-heuristic method, random search method, and the grid search method. The meta-heuristic optimization algorithm is an algorithm based on intuitive or empirical construction, including tabu search algorithm, simulated annealing algorithm, genetic algorithm, ant colony optimization algorithm, and so on. Although the random search method and grid search method have low computational cost, they are not enough to support a larger search space. For more complex optimization problems, it is difficult to provide better solutions. They are generally used to optimize the hyperparameters of classification algorithms such as SVM [30].

Grid search has the characteristic of adjusting parameters automatically. It is a basic parameter optimization method. However, grid search is generally suitable for small-scale data sets. Once the data set increases, it is difficult to obtain the optimal results. Bayesian optimization is also often used in hyperparameter search, but as the data size increases, the complexity and running time of Bayesian optimization will also increase. Therefore, when solving the hyperparameter selection problem of neural networks, heuristic optimization algorithms such as genetic algorithms have obtained better results.

We compared the hyperparameter optimization methods with the four popular algorithms: Grid Search, Random Search, Bayesian optimization, and Genetic Algorithm. Table 5 shows the performance in Accuracy, Recall, AUC.

Table 5. Evaluation performance of different hyperparameter optimization methods

Method	ACC	Recall	AUC
Grid Search	0.83	0.81	0.85
Random Search	0.85	0.83	0.88
Bayesian Optimization	0.86	0.85	0.91
Genetic Algorithm	**0.86**	**0.88**	**0.93**

Grid Search and Random Search are limited to search the better hyperparameter configuration. Although the Bayesian optimization and Genetic algorithm both took a long computational time before outputting a satisfactory result, they have similar results.

In our experiment, the time for training our proposed model took about 30 hours with one NVIDIA GTX 1080 GPU. It takes an average of 0.4 seconds per image when the trained model is used for image classification. The speed suggested that our model could meet the routine clinical work requirements. Overall, the computational time is acceptable, and the adaptive genetic algorithm shows better performance.

5.2 Deep Classification Model

In the proposed algorithm, ResNet50 is adopted to be the classifier, However, many other classification models are used to evaluate the embryo. They may show different performance in various classification tasks. To validate the performance of ResNet50 in embryo evaluation, we compared the performance of our model with the SVM classifier and the random forest (RF) classifier [31]. Both of the two classifiers also use a small dataset that has 221 images. Table 6 shows the classification performance of the three models in Acc, F1-measure, and AUC.

Table 6. Classification performance of SVM, RF, and our model

Method	ACC	F1	AUC
SVM	0.75	0.74	0.77
RF	0.62	0.71	0.75
Ours	**0.86**	**0.87**	**0.93**

Considering that the area of embryo only occupies a small part of the image, and most of the area is in the culture medium which does not provide any reference value, but it will interfere the classification of embryo. Our ROI algorithm can extract the embryo region from the image. This should be the reason that our model gets more excellent performance. Some researchers also take ROI algorithm to extract the embryo before classification, like Qiang Cao et al. proposed a threshold method to ignore all other parts of the embryo image and used a 10-layer Deep CNN network to classify embryo images. The accuracy of the network is 78.14%. It can be seen that the accuracy of their model is not higher than our accuracy, the main reason we speculate is that the classification network is different. There are two advantages over them: one is that we use a more complex and efficient classification model with a network structure, the other is that we design and embed an ROI extraction algorithm into the model.

To verify the effectiveness of the genetic algorithm in optimizing the ResNet network, we also use the genetic algorithm to optimize the hyperparameters of the Vgg16 and SEnet50 [32-34], and they are applied in three different ways. The accuracy, recall, and F1 index of the three networks processing the same data set are shown in Table 7 and Table 8. Table 7 shows the scores of the four evaluation indicators of the residual network optimized with and without the genetic algorithm under the premise of not using the ROI extraction algorithm. Table 8 shows the scores of the four evaluation indicators when the ROI extraction algorithm is used. The sum is not the result of the genetic algorithm optimizing the residual network. It can be seen that both the ROI extraction algorithm and the genetic algorithm have improved the classification performance of the network.

Table 7. Classification performance of Vgg16, Senet50 and ResNet50 comparison between with and without GA optimization (Without ROI Algorithm)

Model	Without ROI Algorithm							
	NO GA				GA			
	Acc	Recall	F1	AUC	Acc	Recall	F1	AUC
Vgg16	0.696	0.535	0.608	0.77	0.734	0.586	0.659	0.79
SEnet50	0.732	0.720	0.703	0.80	0.794	0.719	0.655	0.74
ResNet50	0.735	0.720	0.685	**0.86**	0.852	0.858	0.856	**0.92**

Table 8. Classification performance of Vgg16, Senet50 and ResNet50 comparison between with and without GA optimization (With ROI Algorithm)

Model	With ROI Algorithm							
	NO GA				GA			
	Acc	Recall	F1	AUC	Acc	Recall	F1	AUC
Vgg16h	0.696	0.550	0.614	0.76	0.756	0.540	0.661	0.84
SEnet50	0.748	0.616	0.683	0.81	0.803	0.621	0.732	0.89
ResNet50	0.841	0.862	0.874	**0.92**	0.864	0.843	0.876	**0.93**

5.3 Limitations

Admittedly, there are limitations to the current study. First, there are different systems for embryo morphology evaluation, although all consider the same key factors, such as cell number, the uniformity of cells and the size of the cells. We relied on the doctors who have abundant clinical experience in Shanghai General Hospital to classify the embryo micro-pictures for model training and verification. The main morphological characteristics they focused on include: (1) Cell number: At different stages, embryos should have corresponding cell numbers. Through the number of cells determines whether the growth speed is normal. (2) The uniformity of cells: The cells in an embryo should have similar and standard shapes. (3) The size of cells: The size of cells should match the embryo growth stage. Furthermore, the selection of patients is irregular, as we take the embryo micro-pictures from the all patients we had. However, they have different ages and causes of infertility. We were limited by the long cycle of treatment and observational nature of analysis, resulting in heterogeneous diagnostic evaluation, no structured patient's data classification, analysis of cases at different moments of their disease course, and missing data.

6 Conclusion

An embryo evaluation algorithm based on ResNet neural network optimized by the adaptive genetic algorithm is proposed in this paper. In the model, a ROI extraction algorithm is embedded in the preprocessing part of the ResNet model to eliminate the interference of exfoliated cells. An adaptive genetic algorithm is adopted to optimize the hyperparameters in the ResNet50 network to further improve the feature extraction ability of the residual network. In our experiment, the proposed algorithm can effectively evaluate the embryos in different stages, and the final AUC in the test dataset reaches 0.93. The classification performance is significantly better than the existing classifier. Of course, the proposed algorithm may not be able to adapt to all embryo evaluation criteria, but the overall results of classification are reliable. The future work includes research on unsupervised attention learning and embryo evaluation. Also, The relationship between evaluation results and the age distribution or physical conditions of patients will be taken into consideration, which may significantly improve the performance of the proposed algorithm.

Acknowledgements

This work was supported by the Natural Science Foundation of China (No. 61877065, No. 81872111 and No. 81672562), the National Key Research and Development Program of China (No. 2019YFB1405500), the National Key Technology R&D Program of China (No. 2019YFC-1005200 and No. 2019YFC1005201),Shanghai Municipal Science and Technology Committee of Shanghai outstanding academic leaders plan (No. 19XD1423100), and the project of Outstanding Medical Doctor for ZZ, Shanghai Municipal Education Commission- Gaofeng Clinical Medicine (No. 20181714).

References

[1] K. M. Yenkie, U. M. Diwekar, V. Bhalerao, Modeling the Superovulation Stage in In Vitro Fertilization, *IEEE Transactions on Biomedical Engineering*, Vol. 60, No. 11, pp. 3003-3008, November, 2013.

[2] B. Zhang, Y. Cui, M. Wang, J. Li, L. Jin, D. Wu, In Vitro Fertilization (IVF) Cumulative Pregnancy Rate Prediction From Basic Patient Characteristics, *IEEE Access*, Vol. 7, pp. 130460-130467, September, 2019.

[3] A. Giusti, G. Corani, L. Gambardella, C. Magli, L. Gianaroli, Blastomere segmentation and 3D morphology measurements of early embryos from Hoffman Modulation Contrast image stacks, *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, Rotterdam, Netherlands, 2010, pp. 1261-1264.

[4] S. N. Patil, U. Wali, M. K. Swamy, N. Patil Automation Technique for classification of Human In-Vitro Fertilized (IVF) Embryos using Digital Image

Processing Techniques, *International Journal of Technology and Science*, Vol. 3, No. 3, pp. 46-49, 2016.

[5] C. Herbemont, S. Sarandi, J. Boujenah, I. Cedrin-Durnerin, N. Sermondade, A. Vivot, C. Poncelet, M. Grynberg, C. Sifer, Should we consider day-2 and day-3 embryo morphology before day-5 transfer when blastocysts reach a similar good quality? *Reproductive biomedicine online*, Vol. 35, No. 5, pp. 521-528, November, 2017.

[6] P. Saeedi, D. Yee, J. Au, J. Havelock, Automatic Identification of Human Blastocyst Components via Texture, *IEEE Transactions on Biomedical Engineering*, Vol. 64, No. 12, pp. 2968-2978, December, 2017.

[7] A. Singh, J. Au, P. Saeedi, J. Havelock, Automatic Segmentation of Trophectoderm in Microscopic Images of Human Blastocysts, *IEEE Transactions on Biomedical Engineering*, Vol. 62, No. 1, pp. 382-393, January, 2015.

[8] J. Zhang, Y. Xie, Y. Xia, C. Shen, Attention Residual Learning for Skin Lesion Classification, *IEEE Transactions on Medical Imaging*, Vol. 38, No. 9, pp. 2092-2103, September, 2019.

[9] Y. S. Lin, C. F. Lai, C. L. Chuang, X. Ge, H. C. Chao, Collaborative Framework of Accelerating Reinforcement Learning Training with Supervised Learning Based on Edge Computing, *Journal of Internet Technology*, Vol. 22, No. 2, pp. 229-238. March, 2021.

[10] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, *IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 2016, pp. 770-778.

[11] L. Shao, D. Wu, X. Li, Learning Deep and Wide: A Spectral Method for Learning Deep Networks, *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 25, No. 12, pp. 2303-2308, December. 2014.

[12] C. Chang, Deep and Shallow Architecture of Multilayer Neural Networks, *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 26, No. 10, pp. 2477-2486, October, 2015.

[13] S. Xie, R. Girshick, P. Dollár, Z. Tu, K. He, Aggregated residual transformations for deep neural networks, *IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 2017, pp, 5987-5995.

[14] D. A. Morales, E. Bengoetxea, P. Larrañaga, Selection of human embryos for transfer by Bayesian classifiers, *Computers in Biology and Medicine*, Vol. 38, No. 11-12, pp. 1177-1186, November, 2008.

[15] 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 digital medicine*, Vol. 2, No. 1, pp. 1-9, April, 2019

[16] T. J. Chen, W. L. Zheng, C. H. Liu, I. Huang, H. H. Lai, M. Liu, Using Deep Learning with Large Dataset of Microscope Images to Develop an Automated Embryo Grading System, *Fertility & Reproduction*, Vol. 1, No. 1, pp. 51-56, March, 2019.

[17] D. Gonzalez-Cuautle, U. Y. Corral-Salinas, G. Sanchez-Perez, H. Perez-Meana, K. Toscano-Medina, A. Hernandez-Suarez, An Efficient Botnet Detection Methodology using Hyper-parameter Optimization Trough Grid-Search Techniques, *7th International Workshop on Biometrics and Forensics*, Cancun, Mexico, 2019, pp. 1-6.

[18] T. B. Sasongko, O. Arifin, H. A. Fatta, Optimization of Hyper Parameter Bandwidth on Naïve Bayes Kernel Density Estimation for the Breast Cancer Classification, *2019 International Conference on Information and Communications Technology*, Yogyakarta, Indonesia, 2019, pp. 226-231.

[19] M. H. M. Tarik, M. Omar, M. F. Abdullah, R. Ibrahim, Optimization of neural network hyperparameters for gas turbine modelling using Bayesian optimization, *5th IET International Conference on Clean Energy and Technology*, Kuala Lumpur, Malaysia, 2018, pp. 1-5.

[20] M. Parsa, J. P. Mitchell, C. D. Schuman, R. M. Patton, T. E. Potok, K. Roy, Bayesian-based Hyperparameter Optimization for Spiking Neuromorphic Systems, *IEEE International Conference on Big Data*, Los Angeles, CA, USA, 2019, pp. 4472-4478.

[21] J. H. Han, D. J. Choi, S. U. Park, S. K. Hong, Hyperparameter Optimization for Multi-Layer Data Input Using Genetic Algorithm, *IEEE 7th International Conference on Industrial Engineering and Applications*, Bangkok, Thailand, 2020, pp. 701-704.

[22] A. Shrestha, A. Mahmood, Optimizing Deep Neural Network Architecture with Enhanced Genetic Algorithm, *18th IEEE International Conference On Machine Learning And Applications*, Boca Raton, FL, USA, 2019, pp. 1365-1370.

[23] N. Gorgolis, I. Hatzilygeroudis, Z. Istenes, L. G. Gyenne, Hyperparameter Optimization of LSTM Network Models through Genetic Algorithm, *10th International Conference on Information, Intelligence, Systems and Applications*, Patras, Greece, 2019, pp. 1-4.

[24] K. N. Bui, H. Yi, Optimal Hyperparameter Tuning using Meta-Learning for Big Traffic Datasets, *IEEE International Conference on Big Data and Smart Computing*, Busan, Korea, 2020, pp. 48-54.

[25] M. Feurer, J. Springenberg, F. Hutter, Initializing bayesian hyperparameter optimization via meta-learning, *Proceedings of the AAAI Conference on Artificial Intelligence*, Austin, Texas, USA, 2015, pp. 1128-1135.

[26] A. Thomas, S. Barriere, L. Broseus, J. Brooke, C. Lorenzi, J.-P. Villemin, G. Beurier, R. Sabatier, C. Reynes, A. Mancheron, W. Ritchie, GECKO is a genetic algorithm to classify and explore high throughput sequencing data, *Communications biology*, Vol. 2, No.1, pp. 1-8, June, 2019.

[27] J. H. Yoo, H. I. Yoon, H. G. Kim, H. S. Yoon, S. S. Han, Optimization of Hyper-parameter for CNN Model using Genetic Algorithm, *1st International Conference on Electrical, Control and Instrumentation Engineering*, Kuala Lumpur, Malaysia, 2019, pp. 1-6.

[28] S. Mirjalili, Genetic algorithm, *Evolutionary algorithms and neural networks*, Springer, Cham, 2019, pp. 43-55.

[29] G. N. Xu, Z. Y. Xiao, J. S. Yu, Adaptive genetic algorithm for three-dimension container loading problem, *Control & Decision*, Vol. 22, No. 11, pp. 1280-1283+1288, November, 2007.

[30] R. G. Mantovani, A. L. Rossi, J. Vanschoren, B. Bischl, A. C. Carvalho, Effectiveness of random search in SVM hyper-parameter tuning, *International Joint Conference on Neural Networks*, Killarney, Ireland, 2015, pp.1-8.

[31] Q. Cao, S. S. Liao, X. Meng, H. Ye, Z. Yan, P. Wang, Identification of viable embryos using deep learning for medical image, *Proceedings of the 2018 5th International Conference on Bioinformatics Research and Applications*, Hong Kong, China. 2018, pp. 69-72.

[32] Z. Qu, J. Mei, L. Liu, D. Zhou, Crack Detection of Concrete Pavement With Cross-Entropy Loss Function and Improved VGG16 Network Model, *IEEE Access*, Vol. 8, pp. 54564-54573, March, 2020.

[33] I. Hammad, K. El-Sankary, Impact of Approximate Multipliers on VGG Deep Learning Network, *IEEE Access*, Vol. 6, pp. 60438-60444, October, 2018.

[34] H. Yuan, H. Zhou, Z. Cai, S. Zhang, R. Wu, Dynamic Pyramid Attention Networks for multi-orientation object detection, *Journal of Internet Technology*, Vol. 23, No. 1, pp. 79-90, January, 2022.

Biographies

Wenju Zhou received his B. Sc. degree and M. Sc. degree both from Shandong Normal University in 1990 and 2005 respectively, and obtained PhD degree from Shanghai University in 2014. He is now a professor in Shanghai University. His research interests include medical image processing, machine vision and the industrial applications of automation equipment.

Xiaofei Han received the B. Sc. degree in Electrical Engineering and Automation from the Ludong University, China in 2019. He is now pursuing the M. Sc. degree in control theory and control engineering at Shanghai University. His research interests include image process, machine vision and the industry applications of the automation equipment.

Yuan Xu, Doctor of Genetics, is now the technologist-in-charge of the department of Reproductive Medicine of the General Hospital affiliated to Shanghai Jiaotong University. His research interest was embryonic biology.

Rongfei Chen received the B.Sc. degree in Electrical Engineering and Automation from the Ludong University and the M.S. degree in electrical engineering from the YanShan University. He is now pursuing the Ph. D. degree in control theory and control engineering at Shanghai University. His research interests include image process, machine vision and multimodal sentiment analysis.

Zhenbo Zhang, doctor of Obstetrics and Gynecology of Fudan University, is now the doctoral supervisor of Shanghai Jiaotong University, the post-doctoral cooperation supervisor, the deputy director of the obstetrics and Gynecology clinical medicine center, the deputy director of the reproductive medicine center and the director of the embryo Laboratory of the first people's Hospital Affiliated to Shanghai Jiaotong University. His research interests include reproductive endocrinology, embryonic biology and reproductive toxicology.