Computer Science and Information Systems 22(4):1665–1685 https://doi.org/10.2298/CSIS241001066Y

Application of the Inception-ResNet-V2 algorithm to the analysis of embryo microscope images for the prediction model of assisted reproduction

Yu-Yu Yen[1], Weng Shao-Ping[2], Su Li-Jen[3], Kao Jui-Hung[4,⋆], and Chu Woei-Chyn[1,⋆]

[1] Department of Biomedical Engineering, National Yang Ming Chiao Tung University, Taipei, Taiwan
sheepkelly19.be11@nycu.edu.tw
wchu@nycu.edu.tw
[2] IHMED Reproductive Center, Taipei 106028, Taiwan
drweng@ihmed.com.tw
[3] Department of Biomedical Science and Engineering, National Central University, Taoyuan 320317, Taiwan
sulijen@gmail.com
[4] Department of Information Management, Shih Hsin University, Taipei, Taiwan
kjhtw@mail.shu.edu.tw

Abstract. The World Health Organization (WHO) estimates that approximately14 80 million men and women worldwide, with childbearing potential, need medical assistance due to fertility difficulties, which represents approximately 15% of the population. Similarly, about 15% of couples of maternal ages in Taiwan experience infertility problems. In clinical practice, in vitro fertilization (IVF) is the primary method of artificial reproduction. Using deep learning technology and an Inception-ResNetV2 model, we can create a reliable embryo classification and prediction system, which improves the selection of high-quality embryos and enhances pregnancy success rates. The classification and prediction model achieved 80% precision, AUC= 0.88, sensitivity 73% and 88% specificity. This exceeds the statistics of the Taiwanese National Health Service, where the average pregnancy rate for IVF in 2023 was 27.8 %. The results indicate that our model efficiently classifies embryos for successful implantation at a higher rate than the national statistics in Taiwan.

Keywords: Convolution neural networks, Deep learning; embryo, In vitro fertilization.

1. Introduction

The World Health Organization (WHO) estimates that approximately 80 million individuals worldwide experience infertility and require medical assistance. This figureaccounts for about 15% of the global reproductive population, more than three times the 1 population of Taiwan. Similarly, in Taiwan, around 15% of couples of childbearing age face fertility challenges. Based on this estimate, approximately 300,000 couples in Taiwan are experiencing varying degrees of fertility difficulties [1, 2]. Many factors can cause

⋆ Corresponding authors

infertility or low pregnancy ability. According to the Health Promotion Administration of the Ministry of Health and Welfare of Taiwan, the average pregnancy rate for IVF in Taiwan in 2023 was 27.8%, with a live birth rate of approximately 20.4%. Among these live births, 25.8% were twins, and 0.2% were triplets [3]. Clinical treatments for infertility often involve interventions at various stages, such as ovulation timing, intrauterine insemination (IUI), and more advanced techniques like in vitro fertilization (IVF).

Couples attempting to conceive without using contraception but who are unsuccessful in achieving pregnancy or carrying it to term are considered infertile. The likelihood of pregnancy generally declines with age, particularly after age 30. Conception and live birth rates decrease significantly as women age [4, 5]. The woman's age considerably influences the success of her conception and the quality of the oocyte. A woman's age significantly impacts her ability to conceive and the quality of her oocytes. As women age, the probability of conception decreases yearly while the risk of miscarriage increases. A woman produces approximately one million oocytes during her lifetime,although the number of viable oocytes decreases with age. By the late thirties, the natural conception rate drops below 10%. Among infertility cases, 40% are attributed to female factors, 40% to male factors, and the remaining 20% to combined factors or unexplained causes [6].

With recent advancements in reproductive technology, treatment methods have significantly progressed. In 1978, through the collaborative efforts of British physiologist Dr. Robert Edwards and obstetrician and gynecologist Patrick Steptoe, the first test tube baby was born [7]. This woman had a natural pregnancy and gave birth to a healthy baby. IVF is the breakthrough treatment for infertility. In recent years, reproductive technology has been changing with each passing day, such as the birth of new treatments in ovulation (Gonadotropin, recombinant follicular stimulating hormone (rFSH), etc.), improvement of ovulation stimulation (Gonadotropin-Releasing Hormone (GnRH) enhancer and antagonist treatment), the use of vaginal ultrasound to retrieve the oocyte, major advances in vitro fertilization (Intracytoplasmic sperm injection (ICSI), improvements of vitro culture technology(extended from three days of culture to five days), and changes in embryo implantation methods(replacement uterine implant from implantation of the fallopian tube), various technologies have become more mature, making test tube baby reproductive technology a very important auxiliary method for all infertile couples to seek a child.

IVF treatment is not a panacea; it is simply an artificial assisted reproductive technology that can provide infertile couples with the fastest goal. Many previously helpless problems, such as bilateral fallopian tubes, severe male infertility, etc., are currently only possible to achieve pregnancy through this artificial reproductive technology. According to statistics from the Taiwan Ministry of Health and Welfare, the average IVF pregnancy rate in Taiwan in 2017 was 27.1%, but the live birth rate was only approximately 20.4%, of which 25.8% were twins and 0.2% were triplets [3]. Similarly, data from the European Society of Human Reproduction and Embryology (ESHRE) in 2016 showed that the clinical pregnancy rate per embryo implantation in European countries was only 27.1% [8].

In the early stages of IVF treatment, the low efficiency required the transfer ofmultiple embryos to increase the probability of pregnancy. However, with the accumulation of knowledge and advancements in preimplantation embryonic development, the efficiency of IVF treatments has significantly improved. The transfer of multiple embryos, while increasing the likelihood of achieving pregnancy, also significantly raises the

risk of multiple pregnancies, which can exacerbate maternal and fetal complications during and after pregnancy. These complications include fetal death, developmental arrest, preeclampsia, eclampsia, placental abnormalities, and primary postpartum hemorrhage, all of which become more prevalent with multiple pregnancies [9-11]. Therefore, modern IVF treatment gradually emphasizes the safety of mother infant. The improvement in embryo selection technology is to reduce the number of embryos required for transplantation. Due to the improvement in embryo culture,technology, and culture medium, the efficiency of development from the fertilized egg to the blastocyst stage has increased. Advancements in embryo culture technology and media have significantly improved the development of embryos from fertilization to the blastocyst stage. By extending the in vitro culture period, the developmental potential of each embryo can be better assessed, and advances in cryopreservation techniques further enhance the ability to preserve embryos with high developmental capacity [12-14]. Related studies have compared embryo implantation and pregnancy potential on the basis of their morphology. The goal of selecting a single, optimal embryo for transfer is to achieve a singleton birth [15]. Although the positive impact of single embryo transfer (SET) on mother-infant safety is a well-known fact, it is still not possible to use single embryo transfer as the only method in most centers. The main reason is that current technology is not stable and efficient in selecting embryos with the best implantability, and to maintain an acceptable pregnancy rate, multiple embryo transfers are still necessary to have no choice [16]. Therefore, to increase the clinical results of single embryo transfer, how to further develop new technologies or markers to improve embryo screening remains one of the main topics today [17].

2. Materials and Methods

After ovulation stimulation and in vitro fertilization (IVF) during assisted reproductive technology cycles, embryos are cultured in vitro for two to five days before being implanted in the uterus. The fertilization process and subsequent development of each embryo vary greatly, with some embryos exhibiting robust growth while others may divide more slowly, stop dividing altogether, or develop cytoplasmic fragments. Given the limitations in pregnancy success rates associated with IVF treatment, many assisted reproduction facilities opt to implant a relatively large number of embryos to increase the likelihood of achieving pregnancy. This practice is often driven by the patient‘s strong desire for a successful pregnancy, which consequently increases their tolerance for the risks associated with multiple pregnancies. However, this approach significantly ncreases the risk of high-risk pregnancies and preterm births, placing a substantial burden on medical resources [18]. On this basis, it is considered necessary to establish a set of screening criteria for the treatment cycle of blastocyst-stage embryo implantation and to combine embryo culture values and embryo images using convolutional neural networks (CNN) algorithms. It is used to predict the potential of embryos for pregnancy and provide the clinician with appropriate counseling for the infertile couple after the embryos have been implanted following this standard of selection criteria.

2.1. Importance of embryo quality in IVF treatment

Many studies have pointed out that embryo types from the first day to the sixth day after fertilization can be used as a basis for embryo selection. Therefore, many scoring systems

have been proposed to improve the pregnancy success rate of IVF treatment [19-21]. The clinical guidelines provided by the American Society for Reproductive Medicine (ASRM) have evolved significantly over time, increasingly advocating for personalized approaches tailored to different age groups and varying embryo types, either on the third or fifth day of development. In other words, in clinical practice, the emphasis is on adapting to the patient's age to reduce the number of embryo implantations as much as possible, thus achieving the goal of reducing multiple pregnancies while maintaining a high success rate. At present, three days (division period) or five days (blastocyst period) represent the vast majority of in vitro culture worldwide, and most artificial reproduction centers still mainly have three days of culture. The selective single embryo transfer method during the cleavage stage on the third day is ineffective [22, 23]. This suggests that currently, it is not possible to directly select the embryo with the highest implantation potential for single embryo transfer to achieve the optimal pregnancy rate. Therefore, there are two follow-up solutions for artificial reproduction institutions. One is to continue cultivating embryos in the blastocyst stage before implantation, and the other is establishing screening criteria for embryos in the division stage.

The majority of assisted reproduction institutions focus on using embryos on the third day of the cleavage stage for implantation. Several factors drive this practice: First, extending the in vitro culture period to five days significantly increases the demand for manpower and incubator space by more than 40%. Additionally, the requirements for culture medium differ significantly. Pyruvate is the primary energy source during the cleavage stage, while glucose becomes essential during the morula and blastocyst stages. Consequently, the cost of maintaining cultures for five days increases substantially. Second, although culturing embryos for five days can help select more viable embryos, this process primarily enhances the implantation rate without necessarily improving pregnancy or live birth rates. Is it worth risking embryos due to prolonged culture time due to the risk of degradation? The insistence on extending the time of in vitro culture remains an issue that should be carefully considered by various artificial reproduction institutions. If it is relatively difficult to culture for five days to the blastocyst stage, a set of screening criteria is provided for embryos in the division stage. Especially in the United States and Taiwan, medical insurance does not cover IVF treatment [24, 25].

Patients are very passionate about pregnancy success, which relatively increases patients ardent expectation and degree of tolerance for a twin pregnancy, resulting in many artificial reproduction institutions with implants that include a larger number of embryos to increase their success rate of pregnancy. In view of this, the American Society of Reproductive Medicine provides guidelines for the number of embryo implants for its members, mainly based on age [26]. According to the clinical guidelines of the American Society of Reproductive Medicine on the number of embryos implanted in 2004, 2008, and 2009, patients continue to be divided into four age groups, namely under 35 years, 35 to 37 years, 38 to 40 years, and older than 40. The corresponding number of embryo implantations is suggested according to different age groups. The aim is to reduce the pregnancies of multiple births above triplets in the young population. Even in European countries where single embryo implantation is emphasized, single embryo implantation is restricted to young groups of people under the age of 38 years. The clinical use of the embryo scoring system on the third or fifth day of the type of embryo is relatively limited to young groups of people [27].

2.2. Selection and transfer in IVF treatment

During the fertilized period of fertilized oocytes, mitochondria are distributed mainly around the prokaryotic, especially where the two prokaryotics face each other. However, its distribution is not necessarily uniform. It is separated from the interface of the first split. Sometimes, there is more on one side and less on the other [28]. In prokaryotic stage embryos (fertilized oocytes), an uneven distribution of mitochondria will likely persist through subsequent developmental stages. During the two-cell stage,

mitochondria initially concentrate at the distal ends of the two nuclei before eventually surrounding them. Similarly, at the four-cell stage, mitochondria tend to concentrate at the distal ends of the nuclei in most cells. However, some cells may receive a greater number of mitochondria, while others may receive fewer, leading to variability in mitochondrial distribution across cells [29]. If the number of mitochondria in the cell is less, the ability to synthesize adenosine triphosphate is reduced. These embryonic cells will often stop dividing or even become embryonic fragments and die [28].

The oocyte must undergo several cell cycles (mitosis) to form a blastocyst and subsequent fetal tissue. In addition to precise gene regulation, this process also requires the supply of energy. In mammals, cell energy comes mainly from adenosinetriphosphate. The synthesis of adenosine triphosphate can be divided into two pathways: mitochondrial and non-mitochondrial. Before the stage of compaction or embryonic mulberry, oxidative phosphorylation is used as the pathway of the adenosine triphosphate production pathway, and the main energy source is pyruvate [30, 31]. After the embryonic stage of the mulberry, glucose is preferred as an energy source, but glycolysis is used as the production pathway of adenosine triphosphate [30]. Clinically, in vitro culture is typically divided into two stages. The first stage is enriched with pyruvate, which optimally supports embryo development during the first three days. The second stage is glucose-based, supporting development during the fourth and fifth days.

Embryos that advance to the eight-cell stage within the first three days will likely accumulate substantial amounts of free radicals and oxidative byproducts due to oxidative phosphorylation, making them particularly vulnerable to oxidative stress. Therefore, preserving the mitochondrial membrane potential is essential for supporting embryonic development during this critical early period. At this stage, pyruvate is preferred as the energy source over glucose. It is inferred that if the concentration of free radicals and oxidetives is too high in the first three days, the inner mitochondrial membrane potential will be affected and cannot be maintained, further affecting embryonic development. Suppose a reducing agent that removes the toxicity of free radicals can be provided. In that case, it may help embryo development even more, increase the rate of blastocyst development, and also help increase the pregnancy rate of IVF treatment [31]. The incidence of multiple pregnancies has been steadily increasing each year. In response, the American Society for Reproductive Medicine (ASRM) has developed comprehensive clinical guidelines for its members. These guidelines provide recommendations on the optimal number of embryos to implant, criteria for embryo selection based on maternal age, and assessments of embryo quality. By adhering to these protocols, clinicians aim to enhance reproductive outcomes while minimizing the risks associated with multiple gestations [26].

In recent years, due to the technological advancement of the embryo in vitro culture at National Taiwan University and even in artificial reproduction centers throughout Taiwan, in addition to culturing embryos until the third day for embryo implantation, some

institutions even cultivate blastocysts until the fifth day. Recent clinical guidelines of the American Society for Reproductive Medicine recommend the number of embryos implanted in the division stage on day 3 and the blastocyst stage on day 5, respectively [26]. All artificial reproduction agencies also expect to be able to establish their own guidelines for the pregnancy success rate of IVF treatments for embryo implantation on days 3 and 5. Finding suitable clinical guidelines based on age, embryo quality, number of implants, and other variables by improving embryo selection techniques to reduce the number of embryos required for transplantation. Due to the technological improvement of the embryo in vitro culture and culture medium, the efficient development of fertilized oocytes in the blastocyst stage has been improved. The developmental potential of individual embryos can be better recognized by prolonging in vitro culture time, and the advancement of vitrification and freezing technology can also maximize the preservation of fertile embryos. The embryos with the highest developmental capacity are chemically preserve [12-14].

Additionally, related research compares embryo implantation and pregnancy capacity with embryo morphology. Select the best single embryo transplant for single birth [15, 32]. Although the positive impact of single embryo transfer (SET) on maternal and infant safety is a well-known fact, it is still not possible to use single embryo transfer as the only method in most centers. The main reason is that current technology is not stable and does not efficiently screen embryos with optimal implantability. To maintain an acceptable pregnancy rate, multiple embryo transfer is sometimes not an option but is necessary. Therefore, to increase the clinical results of single embryo transfer (SET), how to further develop new technologies or markers to improve embryo screening remains one of the current important issues.

2.3. Related work

The reproductive technology for test tube babies basically consists of five procedures: ovulation stimulation, egg retrieval, in vitro fertilization, embryo culture, and embryo implantation into the mother's body. The capacity of each embryo to grow during fertilization varies, with some embryos growing well, others slowly dividing or even stopping, and others appearing as cytoplasmic fragments. The highest quality of the first three embryos can be applied to guide the physician and the spouse in coordinating the number of embryos to be implanted. Embryo quality has a significant impact on the success rate of an IVF pregnancy, and some embryos that develop in vitro stop dividing, which is considered to undergo cellular aging. The need to establish a set of selectioncriteria to divide embryos is essential when the 5-day growth period to the blastocyst stage is relatively challenging. The development of new techniques or markers to improve embryo screening is still one of the most important issues today [16], and the following is a description of this research.

Loewke et al. employed an AI model to predict pregnancy rates, achieving an area under the curve (AUC) between 0.6 and 0.7, outperforming traditional manual morphological classification at each stage. A bootstrap analysis predicted that implementing AI could enhance pregnancy rates by 5% to 12% per site, compared to manual classification using an inverted microscope. However, sites utilizing low magnification stereo zoom microscopes did not exhibit the anticipated improvements with AI implementation. Visualization techniques and attribution algorithms indicated significant overlap between the

features identified by the AI model and those used in the manual scoring system. Two sources of bias were identified-associated with the type of microscope and the embryo retention micropipette apparatus-and were subsequently mitigated. The analysis further revealed that a 0.1 (10%) increase in AI scores correlates with a corresponding increase in pregnancy rates [33].

Sujata et al. evaluated embryo quality by visual morphology during in vitro fertilization (IVF) to transfer potential embryos. However, the success rate of in vitro fertilization remains low due to differences in the selection process. The main objective is to improve the rate of implantation by predicting the quality of the embryos that are transferred from day 2 to day 3 [34].

Brás de Guimarães et al. developed an artificial neural network (ANN) supported by a decision tree to predict the probability of live birth after in vitro fertilization (IVF) and intracytoplasmic sperm injection (ICSI) treatments prior to the first embryo transfer. The study analyzed 26 demographic and clinical variables across 1,193 IVF/ICSI treatment cycles conducted at the Centro de Infertilidade e Reprodução Medicamente Assistida between 2012 and 2019. The ANN demonstrated an accuracy of 75.0%, with the area under the receiver operating characteristic (AUROC) curve measuring 75.2% (95% confidence interval: 72.5-77.5%) [35].

To enhance the precision of convolutional neural networks (CNNs) in image classification, C. Peng, Y. Liu, X. Yuan, and Q. Chen have undertaken a comparative analysis of distinct classification model structures. They have proposed an enhanced Inception-ResNet-v2 model, which is based on CNN. A multiscale depth-separable convolution has supplanted the original convolutional structure. This modification has reduced the number of parameters necessary to capture disparate sensory field features. The model has also been endowed with a channel filtering module to filter and merge channels, thereby enhancing the efficiency and accuracy of feature extraction. Data enhancement techniques and other methods optimize the model's performance. Experiments demonstrate that the proposed model surpasses most existing models in multiple datasets, achieving a maximum classification accuracy of 94.8% [36].

Barnett-Itzhaki et al. employed machine learning algorithms, specifically Support Vector Machines (SVM) and Neural Networks (NN), to predict outcomes such as the number of eggs retrieved, mature oocytes, fertilized oocytes, high-quality embryos, positive β-hCG results, clinical pregnancies, and live births. Using age, BMI, and clinical characteristics, these models outperformed traditional logistic regression models in predictive accuracy. The precision of the NN and SVM models ranged from 0.69 to 0.9 and 0.45 to 0.77, respectively, while the logistic regression model exhibited a lower precision range of 0.34 to 0.74 [37].

3. Results

3.1. IOTA Decentralized Ledger Technology

Inception-ResNet-v2 [38] is an enhanced variation of the earlier Inception V3 model, incorporating advancements inspired by Microsoft's ResNet architecture. This improved network architecture reduces computational complexity by fusing feature maps at different scales. Specifically, it replaces the 5x5 and 7x7 convolutions with multiple 3x3 convolutions, thus decreasing the computational effort required for processing.

3.2. Inception-ResNet-v1

The architecture consists of five Inception-ResNet-A modules, ten Inception-ResNet-B modules, five Inception-ResNet-C modules, and the Reduction-A and Reduction-B modules. These are sequentially processed after the stem, which connects the input image to the Inception-ResNet modules. Feature extraction is achieved by applying stride, padding, and max-aggregation in convolutional layers, enabling the capture of feature vectors from the facial images. Subsequently, these vectors are processed through average pooling, dropout, a fully connected layer, and normalization of L2 (as depicted in Figure 1).

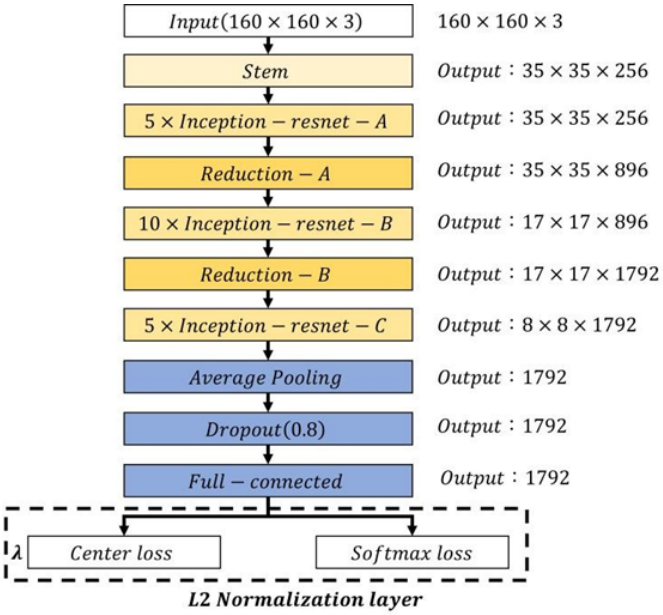


Fig. 1. Inception-ResNet-v1 architecture

3.3. Inception-ResNet-v2

The critical difference between Inception-ResNet-v1 and Inception-ResNet-v2 lies in their output dimensions, with Inception-ResNet-v2 having a more significant number of output dimensions, primarily attributed to variations in their stem structures. Additionally, Inception-ResNet-v2 contains a higher number of parameters within each module. The extraction of feature vectors from face images is achieved by applying stride, padding, and max pooling within the convolutional layers. This extraction process culminates in average pooling, dropout (to mitigate overfitting), a fully connected layer, and an L2 normalization layer. Inception-ResNet blocks, or residual inception blocks, are integral components of the Residual-Inception network. These blocks are computationally more efficient than the original Inception blocks. The Inception-ResNet architecture incorporates residual connections from ResNet into the Inception framework, allowing the output

of each Inception-ResNet layer to add its input value, thereby increasing the network's depth.

The Inception-ResNet module is a meticulously engineered convolutional block designed to produce distinct features while simultaneously reducing the number of parameters within the network. At the end of each Inception-ResNet layer, a 1x1 convolutional kernel is used for dimensionality enhancement, a feature absent in the Inception layer. This is significant since we used 1x1 convolutional cores for the purpose of reducing Inception's computation. We did not resize the image to 299×299. This does not have any change in the number of channels, but only in the size of the feature map generated during the process. After the convolution layer and the Inception module, the feature map size is 5×5 with 1792 dimensional vectors (number of channels). Kaiming He et al. [39], proposed Incep-tionResNet-v2. Architecturally, Inception-ResNet-v1 is quite similar to Incep-tion-ResNet-v2, but the difference lies in a deeper and more complex hierarchy, with more parameters corresponding to higher accuracy.

The main difference between the two is in the preprocessing part, the latter adopts a more complicated stem structure, and the 384-dimensional vector of the stem output dimension of Inception-ResNetv2 is larger than the 256-dimensional vector of Inception-ResNet-v1. The complicated stem structure caused a slightly slower training speed than Inception-ResNet-v1, but produced better performance. The following are the differences between Incep-tion-ResNet-v1 and Inception-ResNet-v2 in terms of stem, Inception-ResNet-A module, Inception-ResNet-B module, Inception-ResNet-C module, and differences between Reduction-A module and Reduction-B module. Although the structure of Reduction-A is the same, the discrepancies lie in the number of parameters. The number of parameters k, l, m, and n of Inception-ResNet-v1 is 192,192, 256, and 384, while the number of parameters k, l, m, and n of Inception-ResNet-v2 is 256, 256, 384, and 384 (as shown in Figure 2˜Figure 7).

Our findings indicate that scaling the residuals in a residual network can lead to instability during the early stages of training when the number of filters exceeds 1000. To address this issue, the learning rate is gradually adjusted to a stable level. At the same time, the residual scaling factor is maintained between 0.1 and 0.3 to ensure consistent and regular training (as illustrated in Figure 8).

The advancement of ResNet's residual learning propagation demonstrated that feed-forward and feedback signals could be transmitted directly through the network. As a result, the non-linear activation function (e.g., ReLU) in the shortcut connections was replaced by Identity Mappings. Furthermore, Inception-ResNet-V2 employs Batch Normalization in every layer, which, following normalization, simplifies the training process and improves the model's adaptability to uncertain data, surpassing the performance of earlier methods.

3.4. Research Methodology and Implementation Steps

In this research, Inception-ResNet-v2 pre-trained models were adopted to identify the success and failure of fertilization rate by employing the convolutional neural network in deep learning with migration learning. Each pre-trained model was verified by stratifying K-fold, and the assessment was performed by accuracy (ACC), area under the curve (AUC), sensitivity, and specificity. Consequently, the image modeling training and evaluation process are as follows (as shown in Figure 9).

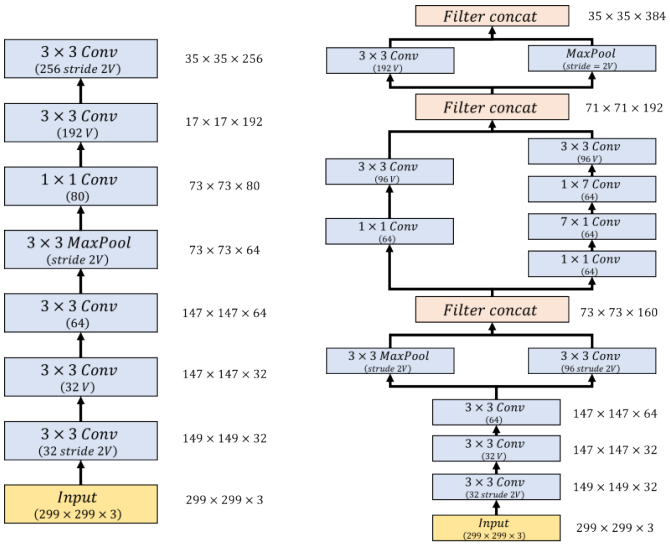


Fig. 2. Inception-ResNet stem differences

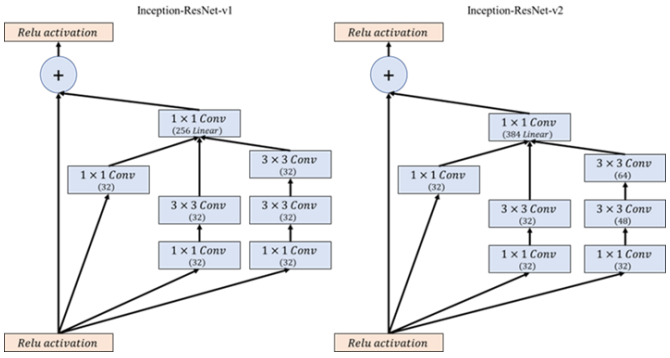


Fig. 3. Differences between the Inception-ResNet-A module

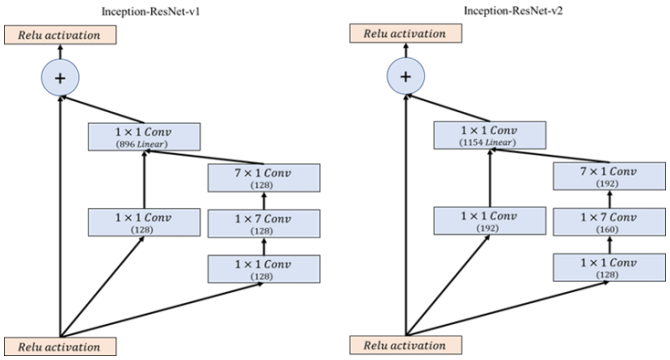


Fig. 4. Differences between the Inception-ResNet-B module

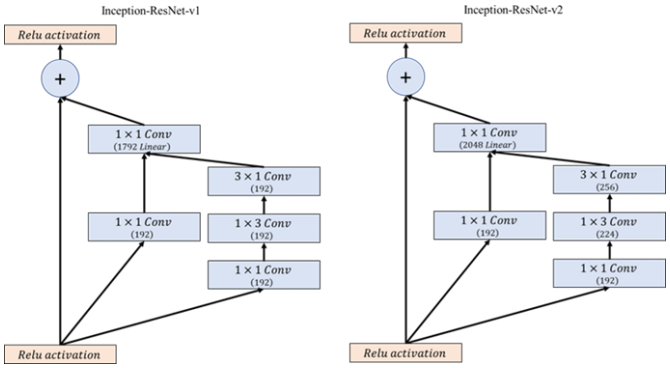


Fig. 5. Differences between the Inception-ResNet-B module

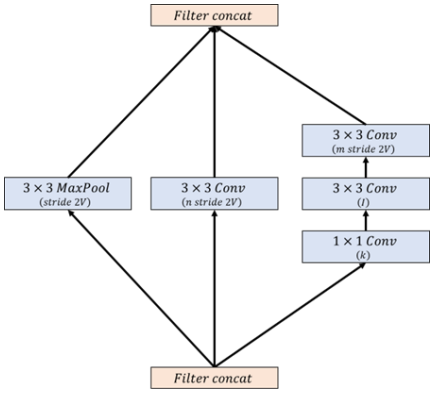


Fig. 6. Reduction-A Module Differences

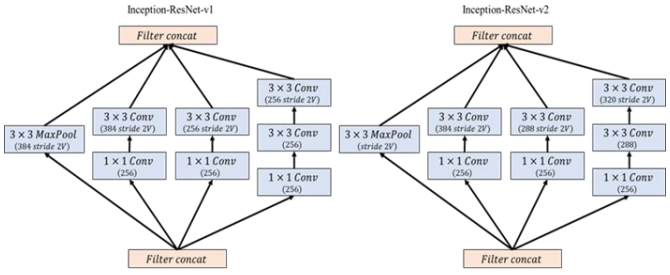


Fig. 7. Reduction-B Module Differences

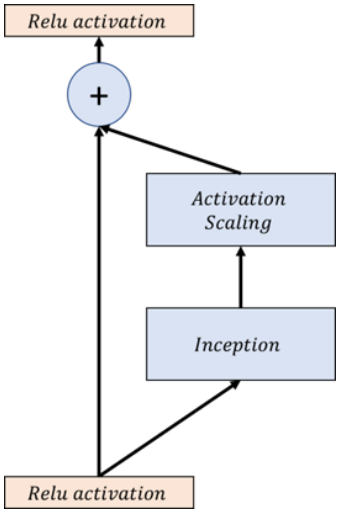


Fig. 8. Scaling of the Residuals

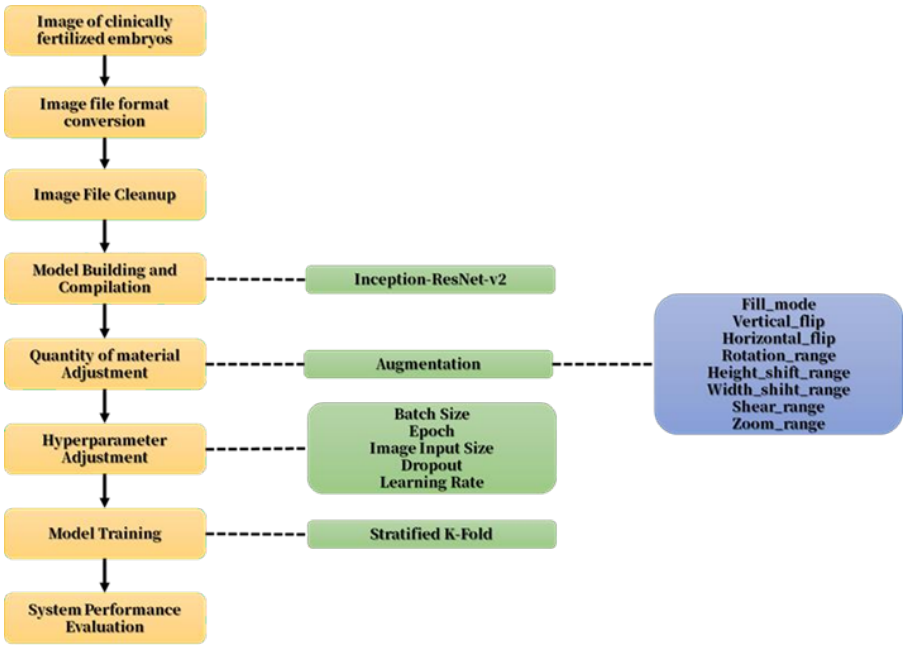


Fig. 9. Image-Modeling and Training Flywheel

3.5. Source and population of data

The data set for this research was acquired from Integrative Holistic Medicine, with a collection of 460 microscopic images of clinically conceived embryos. Of the 460 microscopic images, 150 were in PNG format (Portable Network Graphics), and 310 were in JPEG format (Joint Photographic Experts Group). Subsequently, the Python image library (PIL) was applied to convert the files to JPEG in a standardized manner. Among them, 230 embryos that did not conceive and 230 embryos that sustained a successful conception were also distinguished. The following are examples of embryo image collection (as shown in Figure 10).

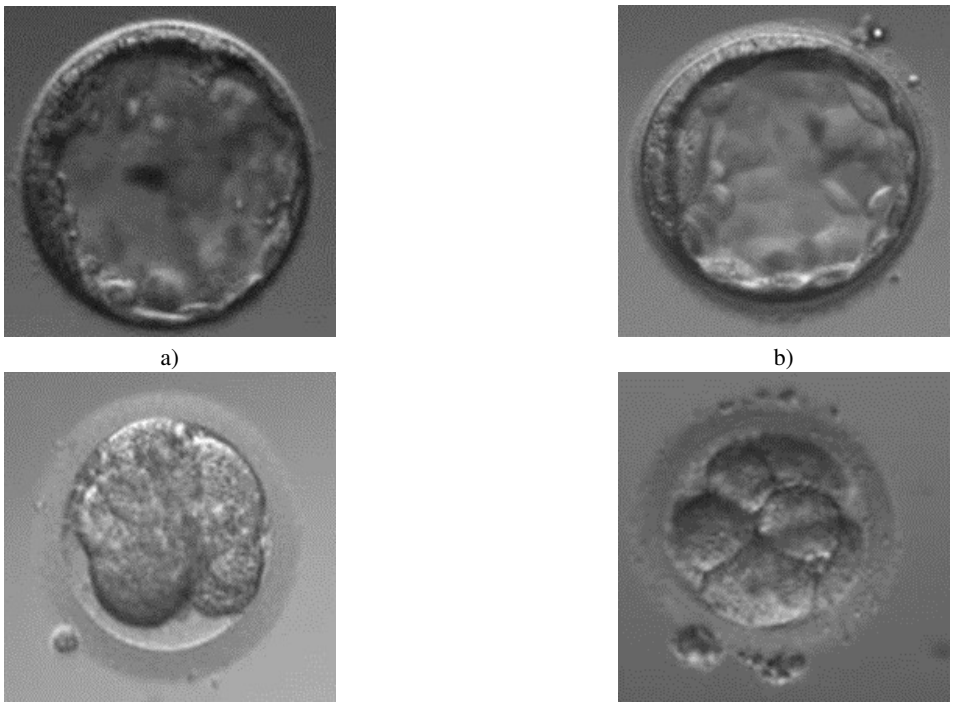


Fig. 10. Example of microscopic images of fertilized embryos selected from the database. (a) and (b) are fertilized embryos with failed implantation; (c) and (d) are fertilized embryos with successful implantation

4. Discussion

4.1. Experimental setup

This research adopts a transfer learning approach when constructing the model. At first, the convolutional substrate neural network weights are frozen, which prohibits updating

the weights while training is in progress. Meanwhile, the original classifier of the dense layer of the pre-trained model is abandoned, while the new classifier of the dense layer is established to compose a new model for retraining. The new model adopts relu and softmax for the activation function and adds batch normalization and dropout. When compiling the new model, the loss function of the new model adopted the categorical cross-entropy (CCE). The Optimizer is configured using Adam with a learning rate=0.001. Batch size of 16 randomly selected samples from the training set. To avoid overfitting, it is mandatory to stop training and keep the best model if there is no improvement after 10 epochs.

In total, 460 images were recorded, 230 without successful conception and 230 with successful conception, all in JPEG format (Joint Photographic Experts Group). The 230 data sets were randomly selected as 80% of the training set and 20% of the test set, while the original embryo image data sets had different sizes of image sources. Considering that the input image size is scaled, the calculation of feature point acquisition will change, which means that the model training results may be affected. Cut the image size to 512 pixels long * 512 pixels wide to facilitate subsequent data training. To deal with the problem of scarcity of data, we made use of the data enhancement method, which employed random horizontal_flip and vertical_flip, shear_range, zoom_range, rotation_range, width_shift_range, height_shift_range, and fill_mode. When we perform migration learning, we discard the original classifier of the dense layer and create a new classifier of the dense layer; therefore, we experiment with adjusting the input image size.

4.2. Accuracy analysis

In the field of machine learning, classification is a common task in supervised learning, for which binary classification is the most commonly applied approach. The actual output of the binary classification algorithm is a prediction score, which indicates the degree of certainty with which the system determines the class to which a given observation belongs to the positive category. For the person who uses this score, if the observation should belong to a positive or negative category, then he/she is required to select a classification threshold by comparing the scores with that value to interpret the scores. Any observation with a score above the threshold was predicted to be in a positive category. In contrast, those with a score below the threshold were predicted to be in a negative category.

As we know, accuracy is one of the most common indicators to evaluate classification models. In layman's terms, this refers to the proportion of results for which the model redicts the correct outcome, i.e., TP + TN divided by the number of all datasets. The alue is equal to the number of correctly predicted samples divided by the total number f the samples in the range [0,1].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

Nevertheless, when we encounter the problem of Imbalance Class Classification, if we only focus on accuracy as the main measurement of model performance, we have the so-called Accuracy Paradox, which makes the accuracy metric meaningless. Therefore, other statistics are available to help us more objectively evaluate whether the classification model is good or bad with unbalanced data sets.

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

To make the predicted results of the classifier more objective and accurate, precision and recall rates are employed (as one of the evaluation indicators, two metrics are extensively used in the field of information retrieval and statistical classification). The recall represents the percentage of data that are correctly predicted to be a positive category among all data that are in positive categories. It is described as follows:

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

Generally, we would like to get the precision and recall of a model that is not too poor, so we use the F1 indicator as a composite measure of the imbalance classification problem. The F1 indicator is the harmonic mean of precision and recall. It is described as follows:

$$F_\beta = (1+\beta^2)\frac{Precision \times Recall}{\beta^2 \times Precision + Recall} \tag{4}$$

Where β is the weight of the precision and recall control, and takes the range of values [0,2]. As we can see from the formula, the larger the value, the greater the emphasis on recall, while the smaller the value, the greater the emphasis on precision. The general classification task is usually both precision and recall, that is, the β-value is taken as 1, which means the F1-measure.

4.3. Confusion matrix

A confusion matrix is a visualization tool that is particularly suitable for supervised learning, which is the most widespread and fundamental way of evaluating classification models. Each column of the matrix represents a category prediction of the data, while each row indicates the actual category of the data. With this mechanism, it is easier to determine whether the model is confusing between the two distinct categories. Once the confusion matrix is obtained, it can be utilized to calculate the accuracy, precision, and recall of the corresponding categories in the model. In addition, we observe the performance of the model in each category. With the visualization tool, it is straightforward to observe which categories are less easy to distinguish, such as how many categories A have been assigned to category B. In the matrix, all correct predictions are on the diagonal, whereas incorrect predictions are presented outside the diagonal. In this regard, we can formulate target-oriented improvement strategies to make the model more distinguished for each category (as shown in Table 1).

Table 1. Confusion matrix

Confusion matrix	Positive	Negative
Positive	True Positive	False Positive
Negative	False Negative	True Negative

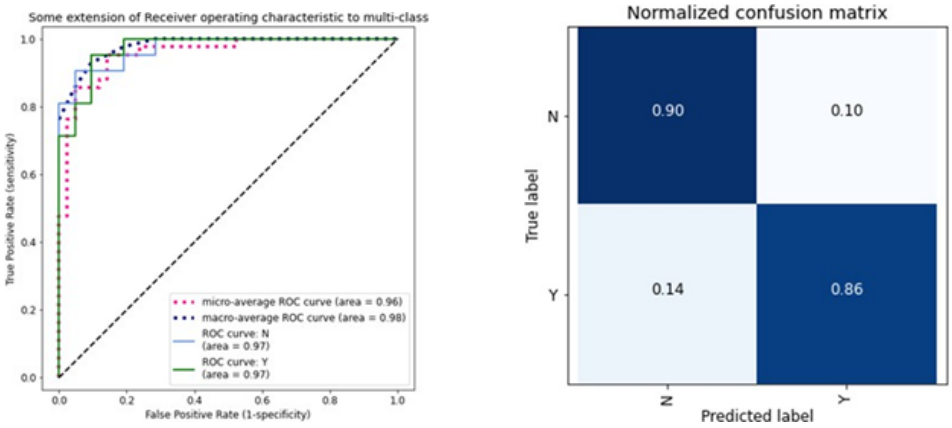
4.4. Analysis of results

Neural networks offer a different approach to pattern recognition and have been used in wide-ranging fields and have proven to be an effective diagnostic tool for many diseases or as a supplement for predicting treatment outcomes. In recent years, due to the popularity of deep learning research, various sophisticated neural network architectures have emerged; inception-resnet-v2 was proposed and applied in image recognition tasks, and the performance of these architectures that are highly task-dependent. In our research, we use InceptionResNetV2 to train a model based on embryo image datasets to evaluate the algorithm, classify, model, and train the embryos according to their morphological quality. For the dataset, it was classified into good and poor, using Group K-Fold to separate the training set and the test set. Furthermore, the test set section is split into validation and testing, with the files being recorded and then the values documented during backtesting. The following table shows the values of recall, precision, and F1-scores of the validation data, which indicate that precision N is 0.86, precision Y is 0.90, recall N is 0.90, recall N is 0.86, and both result in an F1-score of 0.88 (as shown in Table 2).

Table 2. Accuracy, sensitivity, and specificity of the validation set

Precision	recall	f1-score	support
N 0.90	0.90	0.88	21
Y 0.86	0.86	0.88	21
accuracy		0.88	42
macro avg 0.88	0.88	0.88	42
weighted avg 0.88	0.88	0.88	42

To present the values through visualizations, the following figure illustrates the chaotic matrix and the ROC curve of the validation data (as shown in Figure 11).



a) b)

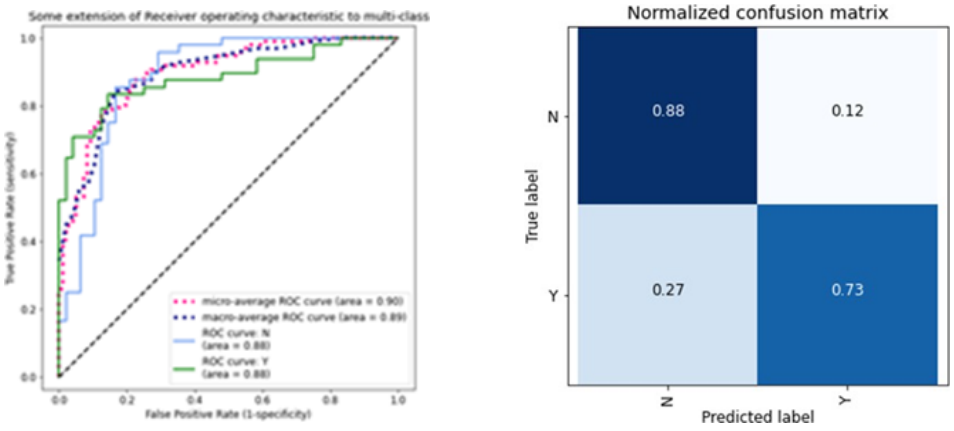
Fig. 11. a) Confusion matrix of the test set. b)Receiver operating characteristic curve plot

While the test data without the training can be used to realize the real situation of the trained model. The following table shows the recall, precision, and F1-score values of the test data. As can be seen, the precision N is 0.76, the precision Y is 0.85, the recall N is 0.88, the recall Y is 0.73 in the category F1-score of N is 0.82, and in the category, the F1-score of Y is 0.79 (as shown in Table 3).

Table 3. Accuracy, sensitivity, and specificity of the test set

Precision	recall	f1-score	support
N 0.76	0.88	0.82	48
Y 0.85	0.73	0.79	48
accuracy		0.80	96
macro avg 0.81	0.80	0.80	96
weighted avg 0.81	0.80	0.80	96

The following figure illustrates the chaotic matrix and the ROC curve for the verification test data (as shown in Figure 12).



a) b)

Fig. 12. a)Confusion matrix of the validation set. b) The curve of the receiver operating characteristic

With advances in machine learning, our future work will focus on many worthwhile applications and targets. For example, to predict the outcome of embryos at early time points of development, the utilization of research, neural networks help with daily clinical tasks and predict the outcome of networked embryos.

5. Conclusions

This study analyzed microscopic images of fertilized embryos using the Inception-ResNet-V2 algorithm. Inception-ResNet is a hybrid model that merges the Inception network with the Residual network. Currently, Inception-ResNet-V2 represents the most

advanced network architecture in the ImageNet dataset, incorporating transfer learning to build classification models. The model's accuracy was validated through K-Fold cross-validation and hierarchical analysis. Combining image enhancement techniques with hyperparameter adjustments, such as modifying image size, dropout layers, and learning rates, we developed a classification model that achieved a maximum accuracy of 80%, exceeding the typical success rate of manual selection.

Despite the promising results, few studies have applied deep learning techniques to the imaging of fertilized embryos. The classification model developed in this research provides a valuable reference for clinical applications, demonstrating improved accuracy in recognizing fertilized embryos. If further systematized, this classification model could serve as a highly effective tool to assist with the manual selection of embryos during the IVF process, thereby reducing workload and minimizing errors associated with manual selection. Ultimately, this would enhance the standardization and efficiency of the embryo selection process, benefiting infertility patients by increasing their chances of successful conception.

When deep learning is performed, the amount of data required is considerable because of the requirement for the machine to learn it within the model. Although the amount of data required for the adjustment of parameters is very large, it may not be possible to adjust the parameters effectively when the amount of data is too small. The number of datasets in this research is 460, which is still not enough for deep learning technology, yet the image quality is not consistent because the datasets are gathered in different years and the machine version has changed over time. Also, when capturing image data, some of them were manually cut, while others were programmed. While the time difference between the intercepted images may cause ambiguity or shift, the different sizes also affect the model training results. In the future, if we can acquire a higher quality and quantity of embryo images or obtain maternal biochemical information for supplemental analysis, it will help improve the precision of embryo selection.

References

1. Zegers-Hochschild, F., et al., *The International Committee for Monitoring Assisted Reproductive Technology.* Human reproduction, **24**(11): p. 2683–2687.(2009)
2. Kissin, D.M., et al., *Assisted reproductive technology surveillance* : Cambridge University Press.(2019)
3. Health Promotion Administration, M.o.H. and Welfare, *The Assisted Reproductive Technology Summary 2021 National Report of Taiwan*, Taiwanese Government Taipei.(2023)
4. Gnoth, C., et al., *Definition and prevalence of subfertility and infertility.* Human reproduction, **20**(5): p. 1144–1147.(2005)
5. Gurunath, S., et al., *Defining infertility|a systematic review of prevalence studies.* Human reproduction update, **17**(5): p. 575–588.(2011)
6. Yu Ng, E.H., et al., *High serum oestradiol concentrations in fresh IVF cycles do not impair implantation and pregnancy rates in subsequent frozen–thawed embryo transfer cycles.* Human Reproduction, **15**(2): p. 250–255.(2000)

7. Warnock, M., *Report of the committee of inquiry into human fertilisation and embryology*. Vol. 9314: HM Stationery Office.(1984)
8. Geyter, C.D. *European pregnancy rates from IVF and ICSI 'appear to have reached a peak'*.[cited 2025 9/4]; Available from: https://www.eshre.eu/Annual-Meeting/Vienna-2019/Media/2019-Press-releases/EIM. (2009)
9. Wennerholm, U.-B., et al., *Incidence of congenital malformations in children born after ICSI.* Human reproduction, **15**(4): p. 944–948.(2000)
10. Sullivan, E.A., et al., *Single embryo transfer reduces the risk of perinatal mortality, a population study.* Human Reproduction. **27**(12): p. 3609–3615.(2012)
11. Lieberman, B., *An embryo too many?* Human reproduction (Oxford, England), **13**(10): p. 2664–2666.(1998)
12. Gardner, D.K., et al., *Culture and transfer of human blastocysts increases implantation rates and reduces the need for multiple embryo transfers.* Fertility and sterility, **69**(1): p. 84–88.(1998)
13. Rienzi, L., et al., *embryo and blastocyst cryopreservation in ART: Systematic review and meta-analysis comparing slow-freezing versus vitrification to produce evidence for the development of global guidance.,23.* DOI: https://doi. org/10.1093/humupd/dmw038: p. 139–155.(2017)
14. Gardner, D.K., *Blastocyst culture: toward single embryo transfers.* Human Fertility, **3**(4): p. 229–237.(2000)
15. Ahlström, A., et al., *Trophectoderm morphology: an important parameter for predicting live birth after single blastocyst transfer.* Human reproduction, **26**(12): p. 3289–3296.(2011)
16. Gardner, D.K., et al., *Diagnosis of human preimplantation embryo viability.* Human reproduction update, **21**(6): p. 727–747.(2015)
17. Gardner, D.K. and B. Balaban, *Assessment of human embryo development using morphological criteria in an era of time-lapse, algorithms and 'OMICS': is looking good still important?* MHR: Basic science of reproductive medicine, **22**(10): p. 704–718.(2016)
18. Lee, T.-H., *The effect of microenvironmental free radicals on embryo development for assisted reproduction technology cycles*. (2012)
19. Henderson, I., et al., *Predicting the outcomes of assisted reproductive technology treatments: a systematic review and quality assessment of prediction models.* F&S Reviews, **2**(1): p. 1–10.(2021)
20. Vestergaard, E.T., et al., *The follicle-stimulating hormone (FSH) and luteinizing hormone (LH) response to a gonadotropin-releasing hormone analogue test in healthy prepubertal girls aged 10 months to 6 years.* European Journal of Endocrinology, **176**(6): p. 747–753.(2017)
21. Visser, D. and F.l.R. Fourie, *The applicability of the cumulative embryo score system for embryo selection and quality control in an in-vitro fertilization/embryo transfer programme.* Human Reproduction (Oxford, England), **8**(10): p. 1719–1722.(1993)
22. Papanikolaou, E.G., et al., *In vitro fertilization with single blastocyst-stage versus single cleavage-stage embryos.* New England Journal of Medicine, **354**(11): p. 1139–1146.(2006)
23. Majumdar, G., et al., *Relationship between morphology, euploidy and implantation potential of cleavage and blastocyst stage embryos.* Journal of human reproductive sciences, **10**(1): p. 49–57.(2017)
24. Jain, T., B.L. Harlow, and M.D. Hornstein, *Insurance coverage and outcomes of in vitro fertilization.* New England Journal of Medicine, **347**(9): p. 661–666.(2002)
25. Howell, E.P., et al., *Preconception evaluation before in vitro fertilization.* Obstetrical & gynecological survey, **75**(6): p. 359–368.(2020)
26. Medicine, P.C.o.t.A.S.f.R. and P.C.o.t.S.f.A.R. Technology, *Guidelines on number of embryos transferred.* Fertility and Sterility, **92**(5): p. 1518–1519.(2009)
27. Penzias, A., et al., *Guidance on the limits to the number of embryos to transfer: a committee opinion.* Fertility and sterility, **107**(4): p. 901–903.(2017)

28. Van Blerkom, J., P. Davis, and S. Alexander, *Differential mitochondrial distribution in human pronuclear embryos leads to disproportionate inheritance between blastomeres: relationship to microtubular organization, ATP content and competence.* Human reproduction, **15**(12): p. 2621–2633.(2000)
29. Cozzolino, M., D. Marin, and G. Sisti, *New Frontiers in IVF: mtDNA and autologous germline mitochondrial energy transfer.* Reproductive Biology and Endocrinology, **17**(1): p. 55.(2019)
30. Colavita, M. and G. Tanzer, *A cryptanalysis of IOTA′s curl hash function.* White paper, p. 1–13.(2018)
31. Wilding, M., et al., *Energy substrates, mitochondrial membrane potential and human preimplantation embryo division.* Reproductive Biomedicine Online, **5**(1): p. 39–42.(2002)
32. Lou, H., et al., *Does the sex ratio of singleton births after frozen single blastocyst transfer differ in relation to blastocyst development?* Reproductive Biology and Endocrinology, **18**(1): p. 72.(2020)
33. Loewke, K., et al., *Characterization of an artificial intelligence model for ranking static images of blastocyst stage embryos.* Fertility and sterility, **117**(3): p. 528–535.(2022)
34. Sujata, P.N., S. Madiwalar, and V. Aparanji. *Machine learning techniques to improve the success rate in in-vitro fertilization (IVF) procedure*. in *IOP Conference Series: Materials Science and Engineering*. IOP Publishing.(2020)
35. Brás de Guimarães, B., et al., *Application of artificial intelligence algorithms to estimate the success rate in medically assisted procreation.* Reproductive Medicine, **1**(3): p. 181–194.(2020)
36. Peng, C., et al., *Research of image recognition method based on enhanced inception-ResNet-V2.* Multimedia Tools and Applications, **81**(24): p. 34345–34365.(2022)
37. Barnett-Itzhaki, Z., et al., *Machine learning vs. classic statistics for the prediction of IVF outcomes.* Journal of assisted reproduction and genetics, **37**(10): p. 2405–2412.(2020)
38. Szegedy, C., et al. *Rethinking the inception architecture for computer vision*. in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016)
39. He, K., et al. *Identity mappings in deep residual networks*. in *European conference on computer vision*. Springer.(2016)

Yu-Yu Yen, she has been working as a lecturer in the Center for General Education at Shih Hsin University since 2022. She is also currently enrolled in a PhD program in the Department of Bio- medical Engineering, National Yang Ming Chiao Tung University.

Shao-Ping Weng, is a specialist in reproductive medicine at IHMED International Healthcare. His clinical and research expertise includes recurrent in vitro fertilization (IVF) failure, advanced maternal age infertility, immunological infertility, and premature ovarian insufficiency.

Li-Jen Su, is the Director of the Service Learning Development Center at National Central University (NCU). He also serves as the Director of the Education and Research Center for Technology-Assisted Substance Abuse Prevention and Management at NCU, as well as the Core Facilities for High Throughput Experimental Analysis in the Department of Biomedical Sciences and Engineering at NCU.

Jui-Hung Kao, is an associate professor at Shih Hsin University since 2023. The empirical research topics focus on spatial data analysis, medical management research, and long-term medical policy.

Woei-Chyn Chu, earned his Ph.D. in electrical and computer engineering from the University of California at Irvine in 1991, he holds the position of distinguished professor in the Department of Biomedical Engineering at National Yang Ming Chiao Tung University in Taipei, Taiwan, ROC.Dr. Chu has made significant contributions through his patented work, particularly in integrating RFID technologies for clinical applications and real-time tracking.

Received: October 01, 2024; Accepted: May 18, 2025.