International Journal of
Applied Data Science in Engineering and Health
https://ijadseh.com

Enhancing IVF Success: Deep Learning for Accurate Day 3 and Day 5 Embryo Detection from Microscopic Images

Shaghayegh Mahmoudiandehkordi [1,*], Maryam Yeganegi [2], Mostafa Shomalzadeh [3], Younes Ghasemi [4], Maryam Kalatehjari [5]

[1] Department of Obstetrics and Gynecology, Isfahan University of Medical Sciences, Isfahan, Iran

shaghayeghmahmoudian@gmail.com

[2] Department of Obstetrics and Gynecology, Iranshahr University of Medical Sciences, Iran

[3] Shahid Beheshti University of Medical Sciences, Tehran, Iran

[4] Islamic Azad University, Tehran Medical Branch, Tehran, Iran

[5] Reproductive Sciences and Sexual Health Center, Isfahan University of Medical Sciences, Iran

Received date: July 3, 2024; Accepted date: August 10, 2024

Abstract

Embryo selection is a critical factor in the success of in vitro fertilization procedures, directly influencing the likelihood of a successful pregnancy. Accurate classification and detection of embryos at key developmental stages, particularly on Day 3 (cleavage stage) and Day 5 (blastocyst stage), are essential for making informed decisions regarding embryo transfer. Traditional methods of embryo assessment rely on manual morphological evaluation, which can be subjective and prone to inter-observer variability. To address these limitations, this study investigates the application of advanced deep learning models for the automated detection of embryo state using microscopic images captured on Day 3 and Day 5. We evaluated five state-of-the-art convolutional neural network (CNN) architectures—VGG19, DenseNet, ResNet50, InceptionV3, and EfficientNetV2—based on their performance in distinguishing between 'Day 3' and 'Day 5' embryos. The results indicate that EfficientNetV2 outperformed the other models, achieving the highest accuracy (94.34%), precision (93.29%), and recall (94.31%). This superior performance suggests that EfficientNetV2 is the most reliable model for embryo state classification, offering the potential to significantly enhance the accuracy and consistency of embryo selection in IVF clinics.

Keywords*:* In Vitro Fertilization (IVF); Day 3 Cleavage Stage; Day 5 Blastocyst Stage; Deep Learning.

Introduction

In vitro fertilization (IVF) is one of the most commonly used assisted reproductive technologies (ART) and has revolutionized the treatment of infertility. By enabling the fertilization of eggs outside the body, IVF provides a pathway to pregnancy for couples facing various reproductive challenges. However, the success of IVF is contingent upon multiple factors, with the selection of high-quality embryos for transfer being among the most critical. The process of selecting embryos involves evaluating their morphological characteristics at different developmental stages, notably on Day 3 (cleavage stage) and Day 5 (blastocyst stage). On Day 3, embryos typically consist of 6-8 cells, and embryologists assess criteria such as the number of cells, the uniformity of cell size, the degree of fragmentation, and the overall symmetry. By Day 5, embryos reach the blastocyst stage, characterized by a fluid-filled cavity, and the differentiation of cells into the trophectoderm and inner cell mass (ICM). The quality of the blastocyst, including the appearance and structure of these components, plays a crucial role in determining the embryo's viability. [1-3]

Despite the importance of these assessments, traditional embryo grading is inherently subjective. Embryologists use their expertise to visually inspect embryos under a microscope, but this process can be influenced by individual experience and interpretation, leading to variability between observers. This inter-observer variability can result in inconsistent embryo selection, affecting the chances of a successful pregnancy. Additionally, manual grading systems often fail to capture subtle, yet critical, features that may be predictive of an embryo's potential for implantation and development. The classification of embryos on Day 3 and Day 5 is particularly important as these stages represent key milestones in embryo development, providing crucial information about the embryo's viability and potential for successful implantation. On Day 3, embryos are typically at the cleavage stage, consisting of 6-8 cells. The number of cells, the rate of cell division, and the uniformity of cell size are key indicators of the embryo's health. Embryologists assess these features, along with the degree of fragmentation and overall symmetry, to identify embryos with the highest potential for further development. Based on this assessment, decisions can be made about whether to transfer the embryo on Day 3 or continue culturing it to Day 5 for further evaluation. By Day 5, embryos that continue to develop reach the blastocyst stage, characterized by the differentiation of cells into the trophectoderm and inner cell mass (ICM). The quality of these structures is a critical indicator of the embryo's ability to implant and result in a successful pregnancy. Blastocysts generally have a higher implantation potential compared to earlier-stage embryos, making Day 5 classification an important step in optimizing success rates. Additionally, the blastocyst stage is often preferred for preimplantation genetic testing (PGT) to screen for chromosomal abnormalities before transfer, allowing for the selection of the healthiest embryos. Given the subjective nature of manual grading, there is a growing interest in developing more objective and standardized methods for embryo assessment. [3-8]

Analyzing imaging data in medicine is of paramount importance as it enables the extraction of critical diagnostic information, facilitates the early detection and monitoring of diseases, and supports the development of personalized treatment plans [9,10]. Automated systems based on advanced imaging techniques and artificial intelligence (AI), particularly deep learning, offer promising solutions to these challenges. These systems can analyze microscopic images of embryos in a consistent and reproducible manner, potentially identifying features that are not easily discernible to the human eye. By reducing human error and standardizing the evaluation process, AI-driven approaches have the potential to improve embryo selection, thereby increasing the success rates of IVF procedures. The advent of deep learning and its applications in medical imaging has opened new avenues for enhancing the accuracy and consistency of embryo classification. Deep learning models, particularly convolutional neural networks (CNNs), have demonstrated remarkable success in various image classification tasks, including medical diagnostics, by learning intricate patterns and features from large datasets. In the context of IVF, the application of deep learning to embryo classification offers the potential to standardize assessments, reduce human error, and ultimately improve the selection of embryos most likely to result in successful pregnancies. [10-12]

Deep convolutional neural networks (DCNNs) are powerful tools for extracting meaningful features from images, which are essential for tasks like object recognition, identification, and image similarity searches [13-15]. These networks represent a deep learning architecture that leverages multiple layers of artificial neurons to process data in a non-linear manner, allowing the system to learn and recognize complex patterns from the input data. In recent years, the application of machine learning and deep learning algorithms in detecting and classifying embryo analysis tasks has seen a significant rise [16-18]. Various computer-assisted techniques, such as pattern recognition and time-lapse image analysis, have been explored for embryo grading and selection, underscoring ongoing efforts to improve the efficiency and accuracy of assisted reproductive technology (ART). These techniques typically utilize morphological features as key discriminators in assessing embryo analysis, to predict the likelihood of successful implantation and pregnancy.

This study aims to develop and evaluate a deep learning-based model for the automated classification of embryos on Day 3 and Day 5 based on microscopic images. By focusing on these critical stages of development, the model seeks to assist embryologists in making more informed decisions during the embryo selection process. The proposed method leverages state-of-the-art CNN architectures to extract meaningful features from the images, enabling the accurate differentiation of embryos with varying developmental potential.

Methods and Materials

In this study, we employed several convolutional neural network (CNN) architectures to classify Day 3 and Day 5 embryos based on microscopic images. CNNs have become the cornerstone of modern image classification tasks due to their ability to automatically learn hierarchical representations of visual data. Among the models chosen for this work, VGG16 and VGG19 stand out for their simplicity and depth [19]. These networks, consisting of 16 and 19 layers respectively, utilize small 3x3 convolutional filters throughout the architecture, enabling the extraction of fine-grained image features. VGG networks are particularly known for their consistent performance across various image classification tasks, including medical imaging applications such as tumor detection and organ segmentation, where capturing subtle details is crucial. Another model we explored is the Residual Network, or ResNet [20], which revolutionized deep learning by introducing residual learning. This concept allows for the training of very deep networks, such as ResNet50 and ResNet101, without the issue of vanishing gradients that often plague deep networks. ResNet’s architecture consists of residual blocks that effectively learn identity mappings, thus enabling the network to converge faster and achieve higher accuracy. ResNet has been widely adopted in medical imaging tasks, proving its robustness in complex pattern recognition tasks such as disease classification and anomaly detection. We also integrated DenseNet [21], or Densely Connected Convolutional Networks, which further pushes the boundaries of CNN architectures by connecting each layer to every other layer in a feed-forward manner. This dense connectivity encourages feature reuse across the network, reducing the number of parameters and improving computational efficiency.

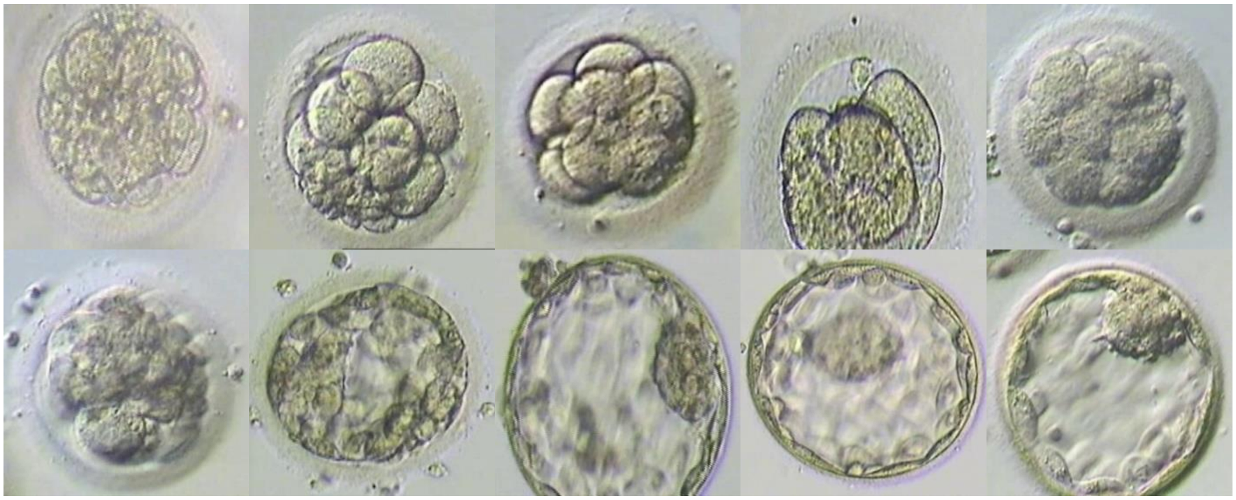

Figure 1. First row: 'Day 3' Embryos. Second Row: 'Day 5' Embryos

Models like DenseNet121 and DenseNet169 are particularly advantageous in medical imaging contexts where the relationships between features are intricate and require detailed analysis, such as in radiological image classification and anomaly detection. The Inception network [22], also known as GoogLeNet, was another architecture used in this study. Inception networks are distinctive for their inception modules, which apply convolutions of varying sizes in parallel within the same layer. This approach allows the network to capture features at multiple scales simultaneously, making it highly effective in tasks where multi-scale feature extraction is essential. In medical imaging, Inception networks have been successfully applied to tasks such as tumor detection and the classification of lesions, where the ability to process features at different scales is critical. Lastly, we employed EfficientNet [23], a family of CNNs designed with a novel compound scaling method that optimizes network depth, width, and resolution simultaneously. This model series, ranging from EfficientNet-B0 to B7, offers a scalable solution that balances accuracy and computational efficiency. EfficientNet has achieved state-of-the-art performance in various image classification benchmarks and is particularly suited for medical applications where computational resources may be constrained, such as in mobile health diagnostics.

Experiments

Dataset

The embryo classification dataset comprises 1,020 images in JPG format, each labeled with a prefix that indicates its corresponding developmental stage. Images of Day 3 embryos are labeled with the prefix "D3," while those of Day 5 embryos are labeled with "D5," allowing for straightforward identification of the embryo's stage. In this study, we introduce deep learning models designed to classify these embryo images accurately, assigning a value of 1 for 'Day 5' and 0 for 'Day 3' embryos. The models will be trained using the provided training dataset and then applied to classify embryo state in the test dataset. Each image is identified by a unique ID in the ID column, and participants will generate the Class column based on the model's classification outcomes. [24]

Training

Each of these models was trained using a comprehensive dataset of labeled embryo images. The training process involved standard techniques such as backpropagation with cross-entropy loss and optimization via stochastic gradient descent (SGD) with momentum.

Metrics

In evaluating the performance of the embryo classification models, three key metrics are used: accuracy, precision, and recall. Accuracy measures the overall effectiveness of the model by calculating the ratio of correctly classified images (both 'Day 3' and 'Day 5' embryos) to the total number of images. It provides a general sense of how often the model makes correct predictions. However, accuracy alone may not fully capture the model's performance, especially in cases where the data is imbalanced. Precision is a more specific metric that assesses the model's ability to correctly identify true-state embryos. It is defined as the ratio of true positive predictions (correctly identified true-state embryos) to the sum of true positives and false positives (images incorrectly classified as true-state embryos). High precision indicates that the model is reliable in identifying true-state embryos, with fewer misclassifications. Recall, also known as sensitivity, measures the model's ability to identify all relevant instances of true embryos.

Table 1. Classification performance of the different models

Model	Accuracy	Precision	Recall
VGG19	81.87±0.0062	81.42±0.045	80.28±0.0024
DenseNet	85.17±0.0034	86.80±0.0027	85.6±0.0032
ResNet50	89.11±0.0084	90.05±0.0025	90.34±0.095
InceptionV3	92.22±0.0069	92.27±0.0057	92.45±0.0066
EfficientNetV2	94.34±0.0038	93.29±0.0048	94.31±0.0029

It is calculated as the ratio of true positive predictions to the sum of true positives and false negatives (actual true state embryos that the model failed to identify). High recall indicates that the model is effective in capturing most of the true state embryos, even if it means including more false positives.

Results

The results presented in Table 1 offer a comprehensive comparison of the performance of several CNN-based models—VGG19, DenseNet, ResNet50, InceptionV3, and EfficientNetV2—in the task of embryo classification. Each model was evaluated using key metrics: accuracy, precision, and recall, providing insights into their ability to distinguish between ‘Day 3’ and 'Day 5' embryos based on microscopic images. EfficientNetV2 emerges as the most effective model, achieving the highest scores across all metrics. With an accuracy of 94.34%, it surpasses the other models by a notable margin. The precision (93.29%) and recall (94.31%) of EfficientNetV2 are also impressive, indicating that the model is not only accurate in its predictions but also reliable in consistently identifying true positives and minimizing false negatives. The superior performance of EfficientNetV2 can be attributed to its optimized architecture, which balances depth, width, and resolution effectively, allowing it to learn and generalize complex patterns in the embryo images more efficiently than the other models. On the other end of the spectrum, VGG19 delivered the lowest performance, with an accuracy of 81.87%. Although VGG19 has been a popular architecture in the past due to its simplicity and effectiveness in various image classification tasks, it falls short in this specific application. The precision (81.42%) and recall (80.28%) are relatively lower compared to the other models, reflecting its limited capacity to accurately and consistently identify true-state embryos. While VGG19 can perform reasonably well, it struggles with the complex, nuanced patterns found in embryo images, which might require more advanced architectures to capture effectively. DenseNet and ResNet50 show a marked improvement over VGG19, with DenseNet achieving an accuracy of 85.17% and ResNet50 reaching 89.11%. Both models demonstrate better precision and recall than VGG19, with ResNet50 performing particularly well in balancing these metrics. DenseNet's improvement can be linked to its dense connectivity, which promotes feature reuse and reduces the number of parameters, thus enhancing its ability to capture the necessary details in the images. ResNet50's use of residual learning helps it maintain high accuracy and precision, making it a strong contender in this comparison. InceptionV3 also shows strong performance, with an accuracy of 92.22%, and precision and recall closely aligned at around 92%. The model's ability to handle multi-scale feature extraction effectively allows it to excel in this task. However, it still falls slightly short of EfficientNetV2, which suggests that while InceptionV3 is highly capable, the optimized scaling in EfficientNetV2 gives it an edge in performance.

Briefly, the analysis indicates that EfficientNetV2 outperforms the other models, making it the most suitable choice for accurate and consistent embryo classification. On the other hand, VGG19, despite its historical significance, exhibits the weakest performance in this context, likely due to its simpler architecture, which may not be sufficient to capture the complex features necessary for high-quality embryo state detection. The results highlight the importance of using advanced, optimized models like EfficientNetV2 for tasks that require a high degree of accuracy and reliability.

Conclusion

This study explored the application of several state-of-the-art CNN-based deep learning models—VGG19, DenseNet, ResNet50, InceptionV3, and EfficientNetV2—for the classification of embryo state based on Day 3 and Day 5 microscopic images. The performance of these models was evaluated using key metrics such as accuracy, precision, and recall, to determine their effectiveness in distinguishing between 'Day 3' and 'Day 5' embryos. The results demonstrate that EfficientNetV2 is the most effective model, achieving the highest accuracy, precision, and recall among the tested architectures. Its superior performance can be attributed to its balanced and optimized architecture, which effectively captures and generalizes complex patterns in the embryo images. This makes EfficientNetV2 the most suitable choice for clinical application in the classification of embryo state assessment, offering the potential to significantly improve the consistency and accuracy of embryo selection in IVF procedures. In contrast, VGG19, while historically significant in the field of image classification, showed the weakest performance, indicating that more advanced architectures are necessary to effectively handle the nuanced features present in embryo images. DenseNet, ResNet50, and InceptionV3 performed better, with InceptionV3 approaching the high performance of EfficientNetV2, but still falling short in comparison.

References

[1] Bódis J, Gödöny K, Várnagy Á, Kovács K, Koppán M, Nagy B, Erostyák J, Herczeg R, Szekeres-Barthó J, Gyenesei A, Kovács GL. How to reduce the potential harmful effects of light on blastocyst development during IVF. Medical Principles and Practice. 2020 May 29;29(6):558-64.

[2] Mastenbroek S, Van Der Veen F, Aflatoonian A, Shapiro B, Bossuyt P, Repping S. Embryo selection in IVF. Human Reproduction. 2011 May 1;26(5):964-6.

[3] Ziaei R, Ghasemi-Tehrani H, Movahedi M, Kalatehjari M, Vajdi M, Mokari-Yamchi A, Elyasi M, Ghavami A. The association between Diet Quality Index–International score and risk of diminished ovarian reserve: a case–control study. Frontiers in Nutrition. 2023 Nov 29;10:1277311.

[4] Allameh Z, Hajiahmadi S, Adibi A, Ebrahimi Oloun Abadi Z, Mahmoodian Dehkordi S. Diagnostic Value of Ultrasonography and MR in Antenatal Diagnosis of Placenta Accreta Spectrum. Journal of Fetal Medicine. 2020 Dec;7:275-81.

[5] Bori L, Meseguer F, Valera MA, Galan A, Remohi J, Meseguer M. The higher the score, the better the clinical outcome: retrospective evaluation of automatic embryo grading as a support tool for embryo selection in IVF laboratories. Human reproduction. 2022 Jun 1;37(6):1148-60.

[6] Sherkat R, Shahshahan Z, Kalatehjari M, Yaran M, Nasirian M, Najafi S, Zangeneh NP, Montazerin SM. Cytomegalovirus specific cell-mediated immunity status in women with preeclampsia: A case-control study. Advanced Biomedical Research. 2023 Jan 1;12(1):10.

[7] Wu C, Yan W, Li H, Li J, Wang H, Chang S, Yu T, Jin Y, Ma C, Luo Y, Yi D. A classification system of day 3 human embryos using deep learning. Biomedical Signal Processing and Control. 2021 Sep 1;70:102943.

[8] Milewski R, Milewska AJ, Czerniecki J, Leśniewska M, Wołczyński S. Analysis of the demographic profile of patients treated for infertility using assisted reproductive techniques in 2005-2010. Ginekologia Polska. 2013;84(7).

[9] Afrazeh F, Ghasemi Y, Shomalzadeh M, Rostamian S. The Role of Imaging Data from Different Radiologic Modalities During the Previous Global Pandemic. International Journal of Applied Data Science in Engineering and Health. 2024 Jul 27;1(1):9-17.

[10] Zaninovic N, Rosenwaks Z. Artificial intelligence in human in vitro fertilization and embryology. Fertility and Sterility. 2020 Nov 1;114(5):914-20.

[11] Litjens G, Kooi T, Bejnordi BE, Setio AA, Ciompi F, Ghafoorian M, Van Der Laak JA, Van Ginneken B, Sánchez CI. A survey on deep learning in medical image analysis. Medical image analysis. 2017 Dec 1;42:60-88.

[12] Shao D, Dai Y, Li N, Cao X, Zhao W, Cheng L, Rong Z, Huang L, Wang Y, Zhao J. Artificial intelligence in clinical research of cancers. Briefings in Bioinformatics. 2022 Jan;23(1):bbab523.

[13] Zadeh FS, Molani S, Orouskhani M, Rezaei M, Shafiei M, Abbasi H. Generative Adversarial Networks for Brain Images Synthesis: A Review. arXiv preprint arXiv:2305.15421. 2023 May 16.

[14] Kumar R, Kumbharkar P, Vanam S, Sharma S. Medical images classification using deep learning: a survey. Multimedia Tools and Applications. 2024 Feb;83(7):19683-728.

[15] Gong Y, Qiu H, Liu X, Yang Y, Zhu M. Research and Application of Deep Learning in Medical Image Reconstruction and Enhancement. Frontiers in Computing and Intelligent Systems. 2024 Apr 10;7(3):72-6.

[16] Yang HY, Leahy BD, Jang WD, Wei D, Kalma Y, Rahav R, Carmon A, Kopel R, Azem F, Venturas M, Kelleher CP. BlastAssist: a deep learning pipeline to measure interpretable features of human embryos. Human Reproduction. 2024 Apr 1;39(4):698-708.

[17] Sharma A, Dorobantiu A, Ali S, Iliceto M, Stensen MH, Delbarre E, Riegler MA, Hammer HL. Deep learning methods to forecasting human embryo development in time-lapse videos. bioRxiv. 2024:2024-03.

[18] Garg K, Dev A, Bansal P, Mittal H. An Efficient Deep Learning Model for Embryo Classification. In2024 14th International Conference on Cloud Computing, Data Science & Engineering (Confluence) 2024 Jan 18 (pp. 358-363). IEEE.

[19] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[20] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[21] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. InProceedings of the IEEE conference on computer vision and pattern recognition 2017 (pp. 4700-4708).

[22] Xia, X., Xu, C., & Nan, B. (2017, June). Inception-v3 for flower classification. In 2017 2nd international conference on image, vision and computing (ICIVC) (pp. 783-787). IEEE.

[23] Mingxing. Tan, Quoc V. Le. Efficientnetv2: Smaller models and faster training. In International conference on machine learning, pp. 10096-10106. PMLR. 2021, July.

[24] https://www.kaggle.com/datasets/gauravduttakiit/embryo-classification-based-on-microscopic-images