

3d Vision in Deep Learning Approaches On Assisted Reproductive Technologies of IVF

J.Deepa^{1*}, A.Akila²

^{1*,2} Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai, Tamil Nadu, India.

Mail Id: ^{1*} pushdeepa3@gmail.com, ² akila.scs@vistas.ac.in

Abstract

In vitro fertilization (IVF) is an important assisted reproductive technology (ART) that relies on accurate embryo assessment and follicle tracking to improve success rates. Recent advances in deep learning and 3D visualization have provided promising solutions to automate and improve both embryo and follicle assessment. This study proposes a deep learning framework for IVF that includes preprocessing, segmentation, and classification techniques. Preprocessing includes non-local means (NLM) filtering and normalization to reduce noise while preserving important morphological details in 3D embryo and follicle imaging. This step ensures improved contrast and clarity, which enables better downstream processing. For segmentation, a U-Net-based framework is used to precisely define reproductive structures such as oocytes, embryos, and follicles, which facilitates accurate localization and feature extraction. Segmentation plays a key role in identifying regions of interest and aiding subsequent classification. By focusing on the segmented regions, the R-CNN model distinguishes between viable and non-viable embryos, as well as follicle maturity stages, and automates the grading process with high accuracy. In this approach used as two datasets as 3D ultrasound images and 3D OCT images. The proposed 3D deep learning approach provides an automated, objective, and efficient method for embryo and follicle assessment, which reduces the subjectivity of manual assessment. This study highlights the importance of deep learning-based 3D vision techniques in revolutionizing IVF procedures and advancing reproductive medicine.

Keywords: OCT, ultrasound images, NLM, Normalization, Follicle monitoring, Faster RCNN

1) Introduction

A widespread trend toward activity and career-centric lifestyles is seen in the modern period of fast socioeconomic advancement. It seems that the conventional focus on family life particularly about having and raising children has diminished. Among contemporary women, this phenomenon is especially noticeable. Infertility is a significant worry in our age because of the apparent worldwide trend of delaying motherhood. According to real-world statistics, one in six couples globally is thought to struggle with sterility¹.

¹ Fu K, Li Y, Lv H, Wu W, Song J, Xu J. Development of a model predicting the outcome of in vitro fertilization cycles by a robust decision tree method. *Front Endocrinol* 2022; 13:877518.

According to the National Center for Health Statistics, infertility affects around 8.8% of people in the United States ². In terms of non-US locations, infertility affects 8–12% of couples, with certain regions seeing a significant increase in occurrence, up to 30% ³. Infertility affects 10% to 15% of couples in Taiwan, which has some of the lowest birth rates on the earth. The standard time to diagnose sterility is 2.9 years, which is much longer than the WHO's one-year criterion, and treatment frequently begins 1.5 years later, according to a 2021 study ⁴. Of the 49,652 treatment cycles in 2021, 8.3% were due to male factors, 55.2% to female factors, 32.2% to combination factors, and 4.3% to unexplained reasons, according to data from the Ministry of Health and Welfare.

Assisted reproductive technology, or ART, is the term used to describe fertility treatments that use eggs or embryos. To do this, a woman's ovaries must have their eggs surgically removed, fertilized in a laboratory, and then given back to the original or a different woman. The Fertility Clinic Success Rate and Certification Act of 1992 states that techniques like timed intercourse and intrauterine insemination, in which only sperm are handled or no egg retrieval is expected, are exempt from these standards. As such, this evaluation does not address IUI or timed intercourse. IVF success rates have skyrocketed in spite of significant technological breakthroughs. In cycles without preimplantation intrinsic testing for aneuploidy, the implantation rate is just about 50%. At its peak, this probability rises to over 60% even with PGT-A ⁵. The live birth rate, which typically stays at about 30% each IVF cycle, emphasizes the challenges that persist even with advancements in IVF technology ⁶.

² Datta J, Palmer MJ, Tanton C, Gibson LJ, Jones KG, Macdowall W, et al. Prevalence of infertility and help-seeking among 15,000 women and men. *Hum Reprod* 2016;31(9):2108e18.

³ Yigit P, Bener A, Karabulut S. Comparison of machine learning classification techniques to predict implantation success in an IVF treatment cycle. *Reprod Biomed Online* 2022;45(5):923e34.

⁴ Chen W-A, Wu C-L, Ho H-Y, Chang F, Yang J-H, Kung F-T, et al. Social determinants of health that impact the time to diagnosis and treatment of infertility in Taiwan. *J Formos Med Assoc* 2024. S0929-6646(24)00238-9.

⁵ Pirtea P, Cedars MI, Devine K, Ata B, Franasiak J, Racowsky C, et al. Recurrent implantation failure: reality or a statistical mirage: a consensus statement from the July 1, 2022 Lugano Workshop on recurrent implantation failure. *Fertil Steril* 2023;120(1):45e59.

⁶ McLernon DJ, Steyerberg EW, Te Velde ER, Lee AJ, Bhattacharya S. Predicting the chances of live birth after one or more complete cycles of in vitro fertilization: population-based study of linked cycle data from 113 873 women. *BMJ* 2016;355: i5735.

To improve the success of ART, a detailed examination of each step of the IVF process is necessary. IVF is a complicated, multi-phase procedure that requires a range of resources. It has challenges since it is time-consuming, labour-intensive, and subject to significant variance across observers. The effectiveness and reproducibility of ART are impacted by these challenges.

The medical sector is increasingly utilizing artificial intelligence for several reasons. It facilitates accurate clinical decision-making, helps with genetic analysis for individualized therapies, and enhances the interpretation of medical imaging. AI also facilitates patient monitoring, expedites medication research, and improves surgical techniques with robotic support. By leveraging the benefits of automation, AI has the potential to increase ART's productivity, reproducibility, and constancy ⁷. By automating repetitive and time-consuming procedures in IVF, including ovarian stimulation or workflows in the embryology lab, artificial intelligence holds promise for reducing the strain on medical professionals and embryologists. AI also continuously improves accuracy through machine learning, which lowers the possibility of errors in these procedures while ensuring best practices and outcomes ⁸.

Contribution of the paper

- Implementing non-local means filtering to reduce noise while preserving important morphological features in embryo images.
- Using normalization techniques to improve contrast and consistency, ensuring better segmentation and classification accuracy.
- Development of an optimized U-Net-based segmentation model for identifying and defining key reproductive structures such as embryos and eggs in 3D imaging.
- Improved segmentation performance, enabling accurate regional localization for downstream classification.
- Integration of a region-based convolution neural network for automatic embryo grading and reliability assessment.
- Efficient classification of embryos based on morphological and structural features extracted from segmented regions
- A complete deep-learning pipeline integrating preprocessing, segmentation, and classification to improve embryo selection in IVF.
- Experimental validation demonstrates improved accuracy and reliability compared to traditional manual assessment.
- Reducing subjectivity and variability in manual embryo assessment.

⁷ Abdullah KAL, Atazhanova T, Chavez-Badiola A, Shivhare SB. Automation in ART: paving the way for the future of infertility treatment. *Reprod Sci* 2023;30(4):1006e16.

⁸ Letterie G, Mac Donald A. Artificial intelligence in vitro fertilization: a computer decision support system for day-to-day management of ovarian stimulation during in vitro fertilization. *Fertil Steril* 2020;114(5):1026e31.

Organization of the paper

This study explores the application of deep learning and 3D vision techniques in embryo and follicle assessment for in vitro fertilization (IVF). Section 2 describes various current literatures on embryo and follicle assessment, highlighting advances in machine learning, deep learning, and 3D imaging techniques in reproductive medicine. Section 3 discusses the proposed methodology, describing the preprocessing (NLM filtering and normalization), segmentation (U-Net), and classification (R-CNN) techniques used to automate and improve embryo and follicle assessment. Section 4 focuses on the performance analysis and experimental results, evaluating the model using precision, recall, accuracy, Dice coefficient, MIoU, and ROC-ACC score. A comparative analysis with random forest and linear regression models is also presented, and the precision and ROC-ACC comparisons are illustrated in Figure 2. Finally, Section 5 concludes the study, summarizing the main findings, emphasizing the impact of deep learning-based 3D vision on IVF.

2) Related work

By automating procedures, improving clinical results, and lowering human error, artificial intelligence (AI) is transforming in vitro fertilization (IVF) labs. It provides individualized treatment programs and prognostic insights. But it also brings up social, legal, and ethical issues like algorithmic prejudice and data security. To guarantee the appropriate application of AI in reproductive medicine, the study highlights the need for further investigation and ethical supervision ⁹. To create a deep learning model for determining ploidy position in time-lapse recordings, retrospective research was conducted. Eighty percent of the time-lapse movies and twenty percent of the known preimplantation genetic testing findings for aneuploidy were old to train the model. With an AUC of 0.74, the model showed promise in identifying embryo ploidy. Exclusion criteria should be optimized and conducted on a greater scale in future studies ¹⁰. With a focus on sperm separation and analysis. We provide a range of microfluidic sperm separation procedures, both natural and non-natural. Additionally discussed is some recent advancement in on-chip fertilization ¹¹.

⁹ Yaling Hew, Duygu Kutuk, Tuba Duzcu, Yagmur Ergun and Murat Basar, “Artificial Intelligence in IVF Laboratories: Elevating Outcomes Through Precision and Efficiency”, *Biology*, Vol. 13, Issues. 12, 2024.

¹⁰ Chun-I Lee, Yan-Ru Su, Chien-Hong Chen, T Arthur Chang, Esther En-Shu Kuo, Wei-Lin Zheng, Wen-Ting Hsieh, Chun-Chia Huang, Maw-Sheng Lee, Mark Liu, “End-to-end deep learning for recognition of ploidy status using time-lapse videos”, *Journal of Assisted Reproduction and Genetics*, Vol. 38, Issues. 7, pp.1655–1663, 2021.

¹¹ Bouloorchi Tabalvandani, M., Saeidpour, Z., Habibi, Z. et al. “Microfluidics as an emerging paradigm for assisted reproductive technology”, *A sperm separation perspective. Biomed Microdevices* 26, Vol 23, (2024).

In direct to bridge the gap between the technological and medical elements of microfluidic sperm collection, the author proposed a solution. Here, we offer a current list of microfluidic sperm selection techniques and how they are used in labs that use assisted reproductive technologies. Nevertheless, additional characteristics like chemotaxis and particularly rheotaxis may be exploited on microfluidic devices ¹². The KIDScore and iDAScore systems predict blastocyst embryo live birth occurrences. Higher KID5 values are more predictive of successful outcomes than higher iDA5 levels, according to the results, which indicate a positive association ¹³.

The difficulty of tracking cell locations within a micro fluidic chip because of the microscope's field of view (FOV) limits was proposed. The suggested technique makes use of an area control strategy with visual feedback to guarantee that the cell stays in the center of the picture. The experimental findings demonstrate the robustness of the loss of eyesight feature, the success rate, and the consistent placement of the tiny objects/cells at each step. With its embryologist-centered design and standardized robotic manipulation, our intelligent manipulation system, when combined with innovative manipulation techniques, presents a potential option for in vitro fertilization ¹⁴.

Building federated learning ecosystems and cognition-inspired learning pipelines are two examples of authentic distributed computing and state-of-the-art machine learning techniques that we identify and evaluate for their usefulness. Lastly, talk about current initiatives and potential research avenues to address current issues and raise the efficacy of AI/ML applications in the healthcare industry ¹⁵. Algorithmic bias, Data security, and human-machine interaction in healthcare decision-making are just a few of the ethical, legal, and social issues that arise when AI is integrated. Enhancing patient results and increasing reproductive medicine's accuracy. It emphasizes the need for further investigation and moral supervision to provide equitable and open applications in this delicate area, guaranteeing the proper application of AI in reproductive medicine ¹⁶.

¹² Ali Reza Jahangiri, Niloofar Ziarati, et al., "Microfluidics: The future of sperm selection in assisted reproduction", WILEY, Vol. 12, Issue. 6,2024.

¹³ Papamentzelopoulou, MS., Prifti, IN., Mavrogianni, D. et al., "Assessment of artificial intelligence model and manual morphogenetic annotation system as embryo grading methods for successful live birth prediction", a retrospective monocentric study. *Reprod Biol Endocrinol* 22, Vol 27, (2024).

¹⁴ S. Miao, Y. Jia, Z. Jiang, J. Xu and X. Li et al., "Cell Cryopreservation in a Microfluidic Chip With Vision-Based Fluid Control and Region Reaching," in *IEEE Transactions on Automation Science and Engineering*, doi: 10.1109/TASE.2024.

¹⁵ E. Zeydan, S. S. Arslan and M. Liyanage et al., "Managing Distributed Machine Learning Lifecycle for Healthcare Data in the Cloud," in *IEEE Access*, vol. 12, pp. 115750-115774, 2024, doi: 10.1109/ACCESS.2024.

¹⁶ Yaling Hew, Duygu Kutuk, et al., "Artificial Intelligence in IVF Laboratories: Elevating Outcomes Through Precision and Efficiency", *Biology*, Vol 13(12), 2024.

Optimizing IVF procedures including drug dosage, timing, and embryological evaluations is being greatly aided by recent developments in AI, such as ML and predictive modeling. Artificial intelligence-powered diagnostic testing, treatment result prediction modeling, schedule optimization, dose and protocol collection, hormone and follicular monitoring, and enhanced embryo selection are some of the innovations ¹⁷. It suggests a similar stream fusion network that divides embryo microscopic images semantically using a shallow, low-cost architecture. With less than 0.7 million trainable parameters, the PSF-Net's shallow design combines the advantages of depth-wise concatenation for feature aggregation with element-wise summing to provide reliable identification. With a mean Jaccard index of 87.69%, the suggested approach also shows encouraging segmentation capabilities for all blastocyst compartments when compared to the most sophisticated techniques. The PSF-Net design's effectiveness is further confirmed by the ablation studies ¹⁸.

BELA predicts quality ratings via multitask learning, which is subsequently employed to forecast ploidy status. Using the Weill Cornell dataset, euploidy and aneuploidy embryo discrimination achieve an area under the receiver in commission feature curve of 0.76, BELA performs comparably to models trained on the manual scores of embryologists. BELA serves as an illustration of how these models can speed up the evaluation of embryos, but it cannot replace preimplantation genetic testing for aneuploidy ¹⁹. To identify the fundamental visual characteristics that underlie image-based classification algorithms, the authors introduce DISCOVER, a generative model. Disentangled latent representations are learned by DISCOVER, where every latent feature encodes a distinct visual attribute that drives categorization. By producing jumbled, exaggerated counterfactual explanations, its design permits "human-in-the-loop" interpretation ²⁰.

- ¹⁷ Pavlovic, Zoran J, Jiang, Victoria S, Hariton, Eduardo, et al., "Current applications of artificial intelligence in assisted reproductive technologies through the perspective of a patient's journey", Vol 36, Issue 4,2024.
- ¹⁸ Muhammad Arsalan, Adnan Haider, et al., "Deep Learning-Based Detection of Human Blastocyst Compartments with Fractal Dimension Estimation", MDPI, Vol 8, Issue 5,2024.
- ¹⁹ Suraj Rajendran, Matthew Brendel, et al., "Automatic ploidy prediction and quality assessment of human blastocysts using time-lapse imaging", Nature Communications volume 15, Article number: 7756 (2024).
- ²⁰ Oded Rotem, Tamar Schwartz, Ron Maor, et al., "Visual interpretability of image-based classification models by generative latent space disentanglement applied to in vitro fertilization", Nature Communications volume 15, Article number: 7390 (2024).

Convolution neural networks are used to pinpoint critical pre-implantation human development windows that are associated with embryo viability and, as a result, appropriate for the early grading of IVF embryos. The findings on how to improve the overall evaluation of embryo viability using machine learning models developed at different developmental stages. By utilizing the well-known properties of transfer learning, we demonstrate CNN models' effectiveness on a small number of datasets, opening the door for clinic-by-clinic implementation that takes local data heterogeneity into account ²¹. compared to algorithms that just looked at time-lapse system (TLS) videos, were much higher for algorithms trained on multi-centric clinical data. The algorithms were trained and validated using a dataset of 9986 embryos from 5226 patients recorded using three distinct TLSs from 14 European fertility clinics (two countries) (95.60% known clinical pregnancy result, 32.47% frozen transfers). The hybrid model's average AUC was considerably higher than the video model's in all seven-folds (0.727 vs. 0.684, respectively; P = 0.015; Wilcoxon test) ²². Blockchain technology safeguards patient data, while 5G integration ensures reliable and fast connection for real-time data transfer. However, developments in networking and cyber security are necessary to handle problems like network problems and security threats. The IoT in the IVF lab should not replace the experience and knowledge of clinical embryologists for safety and supervision ²³. In order to improve healthcare, this chapter critiques the AI systems already in use for GYN/OB diagnostics and shows how these systems were developed and how they may be employed in clinical settings. It also sheds light on the several ways AI is being used in GYN/OB. Many machine learning methods have been successfully used to collect unique data related to fetal cardiocography, previous research on maternal-fetal disorders, and other topics ²⁴.

Two-stage deep supervised 3D convolutional neural networks for the automated detection of ovarian follicles in ultrasound images are presented in this article. The CNNs are trained using transfer learning and complete follicle detection. The results show that the suggested methods estimate follicles up to 2.9% more correctly than the comparative methods and perform around 7.6% better than automatic detection methods. The training is dependable and sufficiently flexible to be applied to a variety of issue areas ²⁵.

²¹ Camilla Mapstone, Helen Hunter, Daniel Brison, et al., "Deep learning pipeline reveals key moments in human embryonic development predictive of live birth after in vitro fertilization", *Biology Methods and Protocols*, Volume 9, Issue 1, 2024.

²² A Duval, D Nogueira, N Dissler, et at., "A hybrid artificial intelligence model leverages multi-centric clinical data to improve fetal heart rate pregnancy prediction across time-lapse systems", *Human Reproduction*, Volume 38, Issue 4, April 2023.

²³ Giles A. Palmer, Giles Tomkin, et al., "The Internet of Things in assisted reproduction" *Reproductive Biomedicine Online*, Volume 47, Issue 5, 2023.

²⁴ Chander Prabha et al., "Role of artificial intelligence in gynecology and obstetrics", *Artificial Intelligence and Machine Learning for Women's Health Issues*, 2024.

²⁵ Bozidar Potocnik and Martin Savc, "Deeply-Supervised 3D Convolutional Neural Networks for Automated Ovary and Follicle Detection from Ultrasound Volumes", *Applied science*, Vol. 12, No. 3, 2022.

One technique for treating infertility in human embryos is in vitro fertilization, or IVF. Deep learning-based methods, however, are costly and imprecise. Semantic segmentation is used to autonomously identify human blastocyst compartments using a parallel stream fusion network. The network's shallow design, which combines element-wise summation and feature aggregation, enables precise detection with 0.7 million trainable parameters. With a mean Jacobian index of 87.69%, the approach exhibits good segmentation performance, and ablation experiments validate its efficacy ²⁶. Introduced EVATOM, an artificial confocal microscopy (ACM)-based machine-learning-assisted embryo health monitoring tool. The technology grades embryos into healthy/intermediate or sick groups using new quantitative embryo health indicators and a label-free nucleus identification approach. The models' weighted F1 scores on in-distribution tests are 1.0 and 0.99, while on out-of-distribution tests, they are 0.9 and 0.95 ²⁷. Medical picture segmentation has greatly improved because to deep learning technologies; the most often used architecture is the U-shaped Network (UNet). On the other hand, U-Net's basic skip connection may result in feature fusion instability and semantic gaps ²⁸

3) Proposed methodology

Follicle tracking in IVF based on assisted reproductive technology (ART) and machine learning approach involves a comprehensive deep learning pipeline. Pre-processing techniques such as non-local means (NLM) filtering and normalization are used to improve image quality by reducing noise and enhancing contrast. For segmentation, a U-Net framework is used to precisely define follicle structures, which ensures accurate localization. After segmentation, a faster region-based convolutional neural network (R-CNN) is used for classification that enables automatic follicle estimation. The performance of the proposed method is evaluated using precision, recall, accuracy, Dice coefficient, mean intersection over union (MIoU) and ROC-ACC score, which ensures robust and reliable results. The overall workflow of a follicle tracking system that integrates ART technology with machine learning is illustrated in Figure 1, which provides a clear representation of the sequential steps in pre-processing, segmentation and classification.

²⁶ Muhammad Arsalan, Adnan Haider, Jiho Choi, and Kang Ryoung Park, “Detecting Blastocyst Components by Artificial Intelligence for Human Embryological Analysis to Improve Success Rate of In Vitro Fertilization”, *Journal of personalized medicine*, pp. 1-16, 2024.

²⁷ Neha Goswami, Nicola Winston, Wonho Choi, Nastasia Z. E. Lai, Rachel B. Arcanjo, Xi Chen, Nahil Sobh, Romana A. Nowak, Mark A. Anastasio & Gabriel Popescu, “EVATOM: an optical, label-free, machine learning assisted embryo health assessment tool”, *communications biology*, Vol. 7, pp. 1-15, 2024.

²⁸ Xia Zhao, Jiahui Wang, Jing Wang, Jing Wang, Renyun Hong, Tao Shen, Yi Liu, and Yuanjiao Liang, “DTLR-CS: Deep tensor low-rank channel cross fusion neural network for reproductive cell segmentation”, *PLOS ONE*, pp. 1-17, 2023.

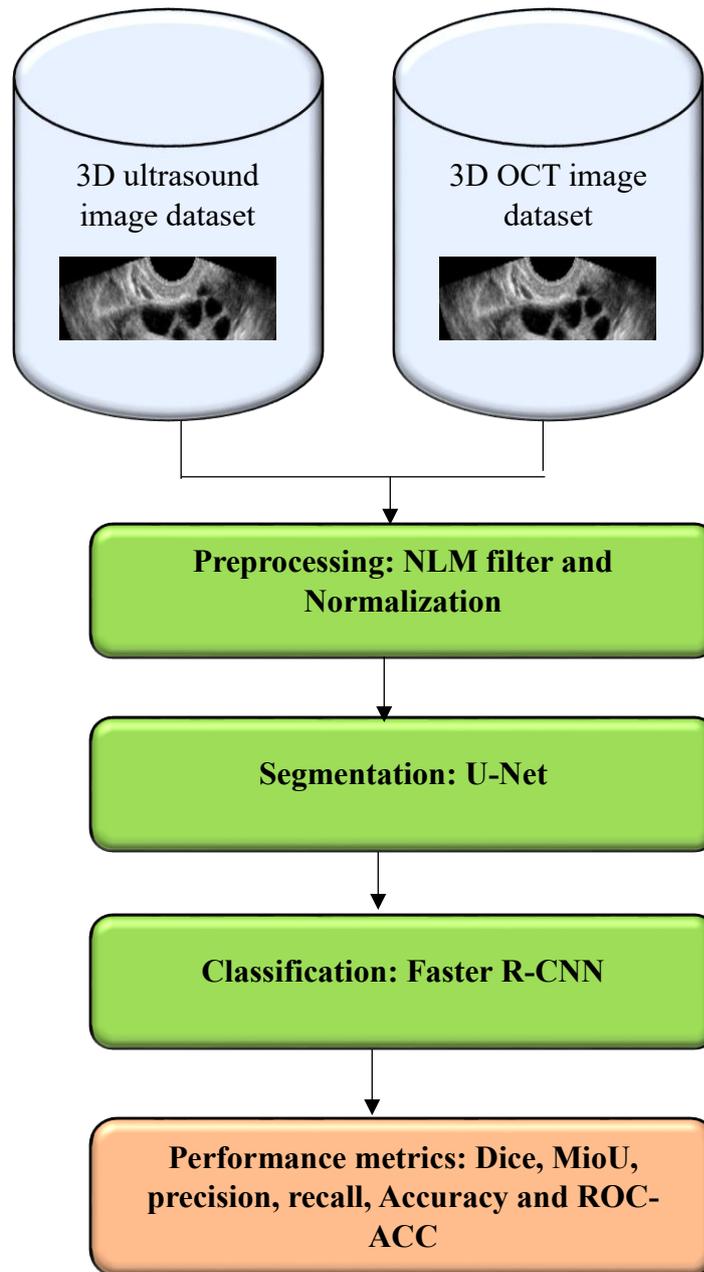


Figure 1. overall flow diagram of follicle monitoring

Preprocessing

The gathered images contain both ovaries in a single file and are saved in several file formats. To include just one ovary in a single file, image cropping is used. Every picture is an RGB image. Normalization is finished once the texture of the picture is restored and speckle noise is indifferent using the non-local mean filter.

NLM filter

The NLM filter typically replaces the worth of the pixel of attention with the standard heaviness of the equally chosen picture areas to produce the de-noised image. Eq. (1) is used to get the pixel weight for each pixel pi of the noisy image I.

$$w(pi, pj) = \frac{1}{W(pi)} e^{-\frac{|I(N_{pi}) - (N_{pj})|_2^2, \sigma}{h^2}} \tag{1}$$

w(pi, pj) – weights of pixels pi and pj

|I(N_{pi}) – (N_{pj})|₂², σ – vector distance and the range parameter

The normalized feature W (pi) is considered in below the equation,

$$W(pi) = \sum e^{-\frac{|I(N_{pi}) - (N_{pj})|_2^2, \sigma}{h^2}} \tag{2}$$

The denoised importance meant for the pixel pi is known by,

$$NLMF(I(pi)) = \sum_{pj \in I} \left(\frac{1}{W(pi)} e^{-\frac{|I(N_{pi}) - (N_{pj})|_2^2, \sigma}{h^2}} \right) (I(pj)) \tag{3}$$

Normalization is completed when the noise has been removed from the picture. The photos are reshaped into 256 × 256 since they are in various forms.

Algorithm

Algorithm steps:

1. Enter the picture.
2. Calculate the value of sigma.
3. Set the patch size to 5 × 5, the patch space to 3 × 3, and the h constant to 1.15*estimated_sigma.
4. With the aid of Equations (1), (2), and (3), calculate the de-noised significance NLMF (I(pi)).
5. Use the image's de-noised pixel significance in place of the original pixel value.
6. Continue steps 4 and 5 until every pixel is achieved.
7. accumulate the de-noised photo in a wallet

Normalization

It is the pre-processing method most frequently employed, transforming the image data into a variety of 0 to 1. The normalize methods utilized here are z-score normalization and zero means, and it is defined as follows in equation (4):

$$X'_i = \frac{X_i - mean(X)}{SD(x)} \tag{4}$$

X'_i – normalization information Where SD (X) denotes the input X's standard deviation and mean denotes the mean value. The mathematical expression for standard deviation is Eq. (5):

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (X_i - \text{mean}(X))^2} \quad (5)$$

The model deviation for the given contribution parameters is given by σ in this case. Furthermore, image scaling is carried out.

Segmentation: U-Net

Individual of the first and best-known methods for segmenting medical pictures is U-Net, which was first put out to solve the cell wall segmentation problem. U-Net is a completely symmetrical U-shaped construction that is divisible in half. In the first section, a systolic route is represented using a standard CNN architecture. Each block of the systolic route consists of two successive 3×3 convolutions, a maximum pooling layer, and a ReLU activation unit. U-Net's extended route is unique in that it up-samples feature maps on each level using 2×2 convolution, after which they are cropped and stitched to the feature maps that were up-sampled from the relevant layers in the shrinkage route. Following two successive 3×3 convolutions and ReLU activations, a further 1×1 convolution is utilized to reduce the feature map to the necessary numeral of channel segmented images. Network trimming of the feature map is also required to exclude pixel features that have minimal contextual information at their borders. More importantly, it allows the network to employ context to distinguish objects from larger overlapping regions by spreading contextual information across the network. The following formula provides U-Net's energy function:

$$E = \sum w(x) \log(P_{k(x)}(x)) \quad (6)$$

$$p_k(x) = \frac{e^{a_k(x)}}{\sum_{k=1}^k e^{a_k(x)}} \quad (7)$$

Here, $a_k(x)$ symbolizes the kth channel's activation function, and pk stands for the softmax purpose that was functional to the network's production feature map.

U-Net uses the skip relation structure to send the low-level feature map to the advanced feature map component and ingeniously provides a one-to-one correlation between the encoder and decoder modules. For processing, it combines the low-level and high-level feature maps, and these actions enable U-Net to perform exceptionally well in the segmentation of medical images. U-Net does, however, have several obvious shortcomings. The relationship between the two forms of information cannot be adequately explained by a straightforward fusion approach that splices the lower-level in order with the higher-level information.

Classification: Faster RCNN (region-based convolution neural network)

The Fast R-CNN network employs SVD to deconstruct its final parallel, fully connected layers, which may concurrently produce box regression and classification results. This eliminates the

need for extra feature storage space, speeds up detection, lowers computational complexity, and enables end-to-end multitasking training. By first substituting Softmax for the SVM classifier and then the ROI pooling layer for the final max pooling layer, Fast R-CNN has enhanced R-CNN by mapping region suggestions of varying sizes to the same dimension. This removes R-CNN's needless feature extraction processes.

The primary cause of the Fast R-CNN's discernible speed advantage over the R-CNN is the ROI pooling layer. In SPP nets, the spatial pyramid pooling layer is only a specific case of a one-level pyramid. Each ROI's height (h , w) and upper-left corner (r , c) are specified by the four-tuple (r , c , h , w). Finding the ROI in the picture and comparing it to the feature map patch is one of its primary objectives. The alternative is to use a single-layer SPP layer to downsample (max pool) the feature map patch into a fixed-size feature before sending it to the fully-connected layer FC.

Following Fast R-CNN, the CPU implementation of the region proposal methods Edge Boxes (0.2s/image) and Selective Search (2s/image) presents the biggest obstacle to enhancing object detection performance. To create a full end-to-end network, the Faster R-CNN substitutes the RPN layer for the Selective Search layer. Building on the concepts of SPPnet and ROI, Region Proposal Networks (RPN) project the CNN feature map onto the original ultrasound and OCT images, then apply boxes of varying sizes to the image. We refer to the box as an anchor. Based on the results of determining the IOU between these anchors and the ground truth, classify the anchor into positive and negative categories. The IOU value is categorized as negative if it is less than 0.3 and as positive if it is greater than 0.7. Based on the training numerical value, a predetermined number of positive and negative samples are selected in order to regress and fix the box borders.

At first, each anchor will provide two scores, uniformly represented by p , one of which will indicate the item's likelihood and the other its chance of not being the object. Four coordinate values are then generated to reflect the positive and negative samples that the RPN collected, together with the anchor's position coordinates in a predefined number of anchors. Consequently, the loss function of the RPN layer is a multi-task loss function consisting of two parts: the Softmax loss of the object probability and the smooth L1 Loss of the coordinate position between the anchor and the ground truth. The following is a precise definition of the loss function:

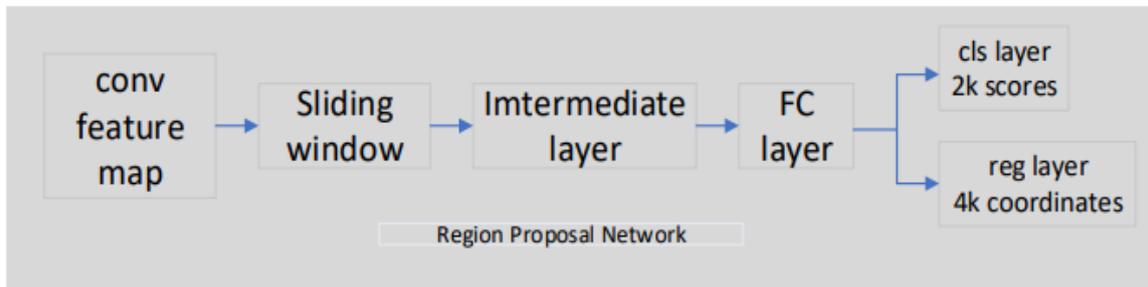


Figure 2. RPN structure

The Loss function is specifically defined as follows:

$$L(\{p_i\}\{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) \quad (8)$$

i is the i -th fasten in a mini-batch in this instance, and p_i is the predicted probability that anchor i will be an article. The ground-truth label p_i^* is 1 when the anchor is positive and 0 when it is negative. It is positive when the anchor's label is positive, and negative when it is negative. The vector t_i^* represents the four parameterized coordinates of the predicted bounding box, and t_i^* represents the coordinates of a ground-truth box that is connected to a positive anchor. Given that (x, y, w, h) represent the width, height, and center coordinates of the box (the same is true for $y, w,$ and h), the variables (x, x_a, x^*) are for the prediction box, anchor box, and ground-truth box, respectively. Making the two values equal is the learning goal, as seen in Figure 3:

$$t_x = (x - x_a)/w_a, t_y = (y - y_a)/h_a \quad (9)$$

$$t_w = \log(w/w_a), t_h = \log(h/h_a) \quad (10)$$

$$t_x^* = (x^* - x_a)/w_a, t_y^* = (y^* - y_a)/h_a \quad (11)$$

$$t_w^* = \log(w^*/w_a), t_h^* = \log(h^*/h_a) \quad (12)$$

In method (2), softmax Loss contains two categories and is given below,

$$L_{cls}(p_i, p_i^*) = -\log(p_i^* p_i + (1 - p_i^*)(1 - p_i)) \quad (13)$$

L_{reg} is the smooth loss for two offsets and the given is below,

$$L_{reg}(t_i, t_i^*) = R(t_i - t_i^*) \quad (14)$$

R is the function which is given below,

$$R(x) = smooth_{L1}(x) = \begin{cases} 0.5x^2 & |x| \leq 1 \\ |x| - 0.5 & otherwise \end{cases} \quad (15)$$

λ Is the hyperparameter that controls the equilibrium connecting the categorization loss and degeneration loss. N_{cls} And N_{reg} are used to regularize the categorization loss L_{cls} and the degeneration loss L_{reg} correspondingly.

Almost half of the network's time is spent operating at the complete connection layer. Consequently, the FC computation is accelerated by using the reduced SVD.

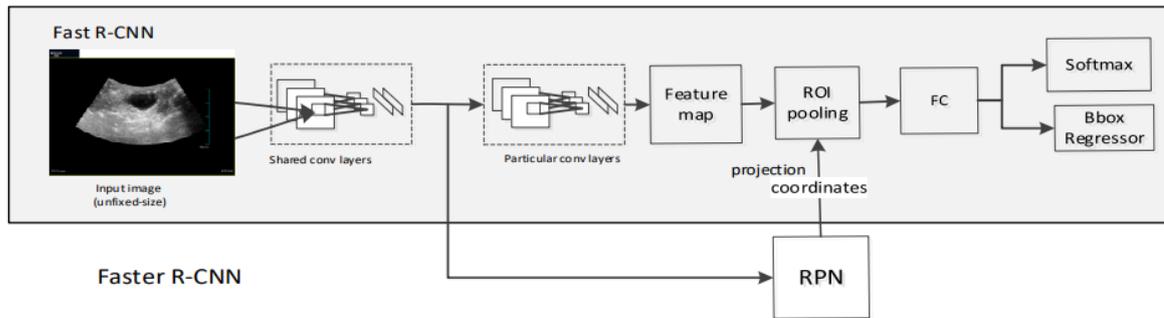


Figure 3. Faster RCNN

4) Result and discussion

Follicle tracking in IVF, based on assisted reproductive technology (ART) and machine learning approach, involves a comprehensive deep learning pipeline. Pre-processing techniques such as non-local means (NLM) filtering and normalization are used to improve image quality by reducing noise and enhancing contrast. For segmentation, a U-Net framework is used to precisely define follicle structures and ensure accurate localization. After segmentation, a Faster region-based convolutional neural network (R-CNN) is used for classification that enables automatic follicle estimation. Performance analysis includes pre-processing, segmentation, and classification, with comparative analysis with existing follicle tracking techniques to highlight improvements. The proposed model is evaluated using precision, recall, accuracy, Dice coefficient, Mean Intersection over Union (MIoU), and ROC-ACC score. To ensure robust and reliable performance, the entire algorithm is implemented using the Python programming language, version 3.11, utilizing advanced deep learning libraries for optimal performance.

Performance metrics

Dice

The assessment criteria for the segmentation procedure are the dice similarity coefficient (DSC). Dice is a measure of aggregated similarity. With values ranging from 0 to 1, the measure is typically used to determine how similar two samples are to one another. The top segmentation effect is 1, and the most unpleasant is 0. The following is the method for manipulative dice:

$$dice = \frac{2TP}{2TP+FP+FN} \quad (16)$$

MioU

One common measure for semantic segmentation is Mean Intersection over Union (MIoU). To get the average value, MIoU is utilized to compute the ratio of the connection and concatenation of the network's real and forecasted outcomes. The following is the calculating formula:

$$MIoU = \frac{TP}{TP+FP+FN} \quad (17)$$

Precision

In the fields of statistical classification and information retrieval, precision is a commonly used measure. The measure, which is used to evaluate the quality of the findings, shows the ratio of samples that are anticipated to be accurate to folks that are projected to be accurate:

$$precision = \frac{TP}{TP+FP} \quad (18)$$

Recall

The ratio of accurately predicted samples to the number of activist samples in the sample is known as recall. The quality of the outcomes is also assessed using the measure. The following is the calculation.

$$recall = \frac{TP}{TP+FN} \quad (19)$$

Where FP is the numeral of predicted pixel sites that receive an inaccurate target category classification, FN is the numeral of predicted pixel points that receive an incorrect non-target category classification, and TP is the number of pixels that are expected to receive a correct target class classification.

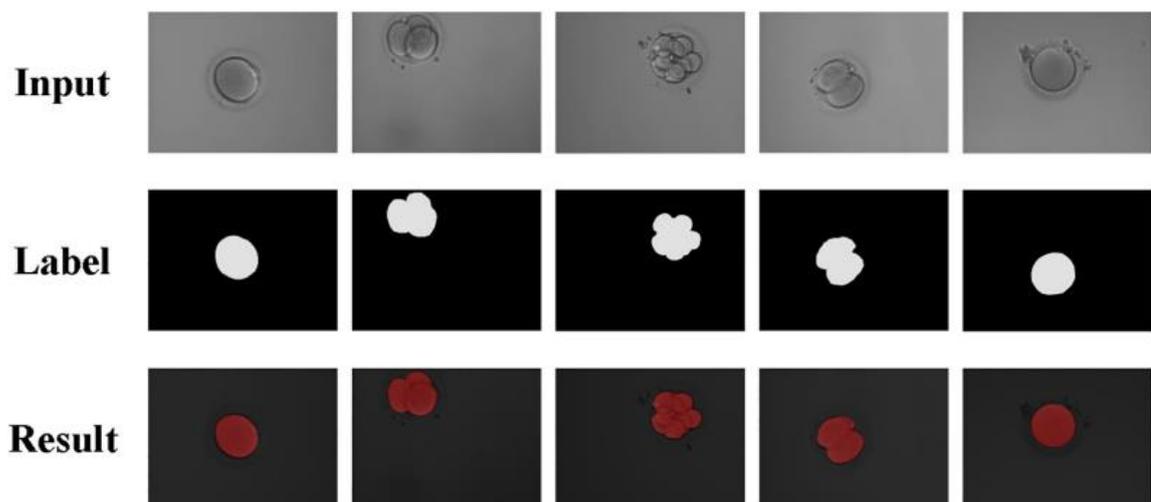


Figure 4. Test segmentation results

The results of cell segmentation utilizing the approach suggested in this study are shown in Figure 4, which shows that the method can precisely distinguish the cells from the background, leading to improved cell segmentation test results.

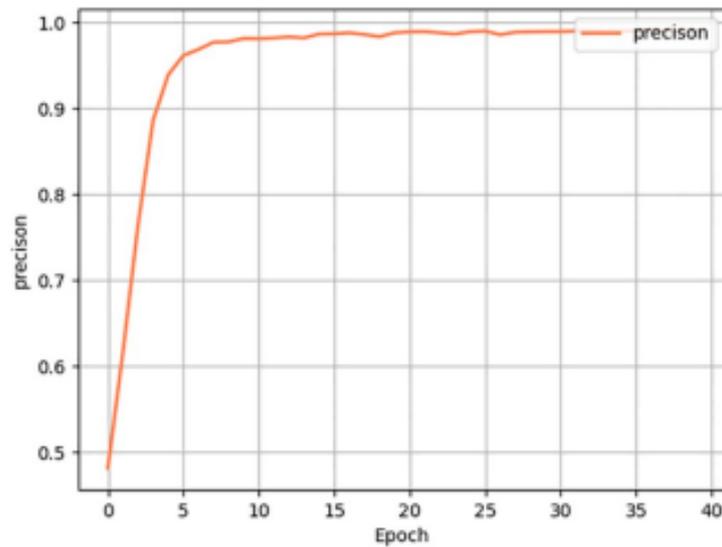


Figure 5. Precision metric results

As shown in Figure 5, as the number of training rounds increased, the accuracy continued to increase and then flattened out, indicating that the model had reached convergence and further training did not yield significant improvements.

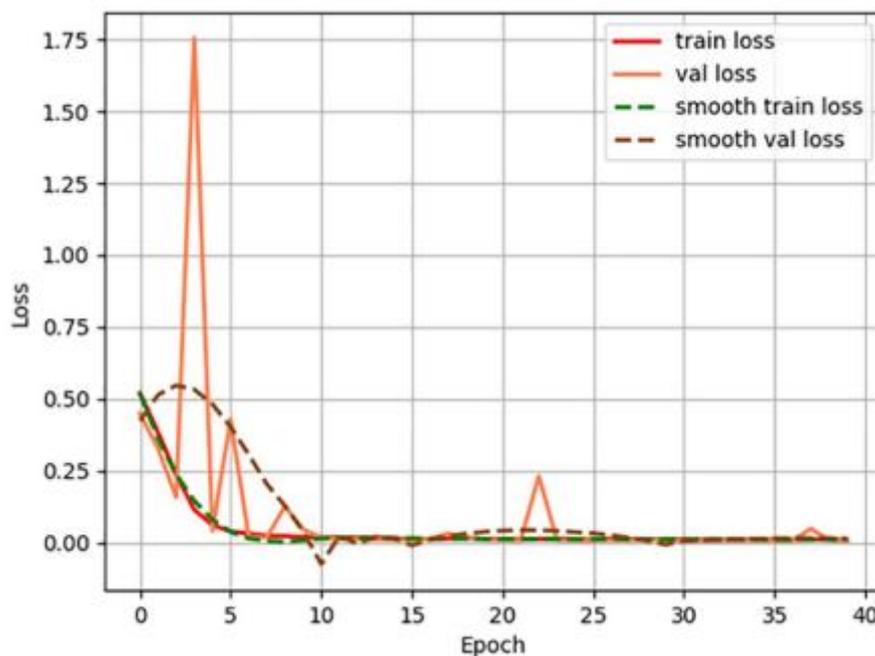


Figure 6. Losses values for 50 epochs

The loss rate also kept down at the same period, and the curve briefly fluctuated before leveling out. Because of this, our approach has performed well in picture classification utilizing faster R-CNN, as seen in figure 6.

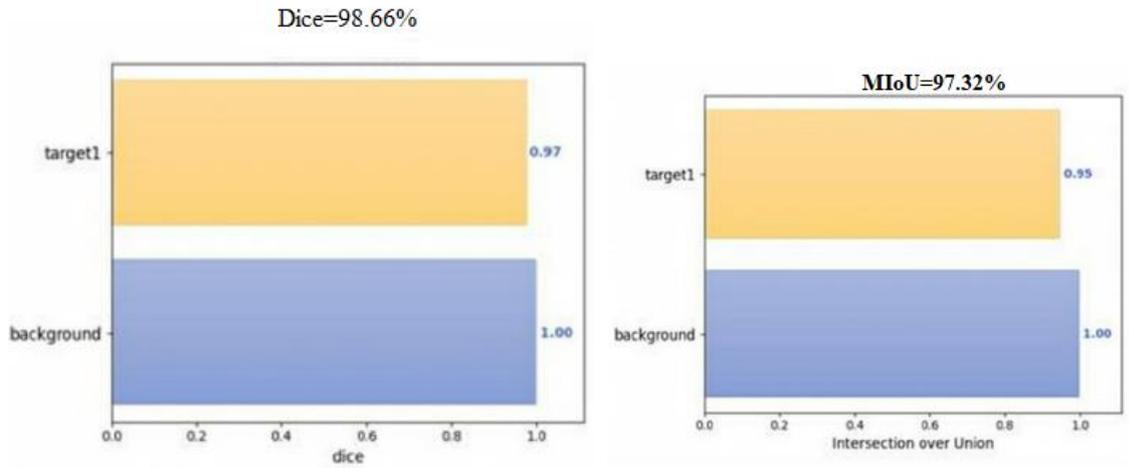


Figure 7. Dice and MIoU metrics of RCNN based on ART

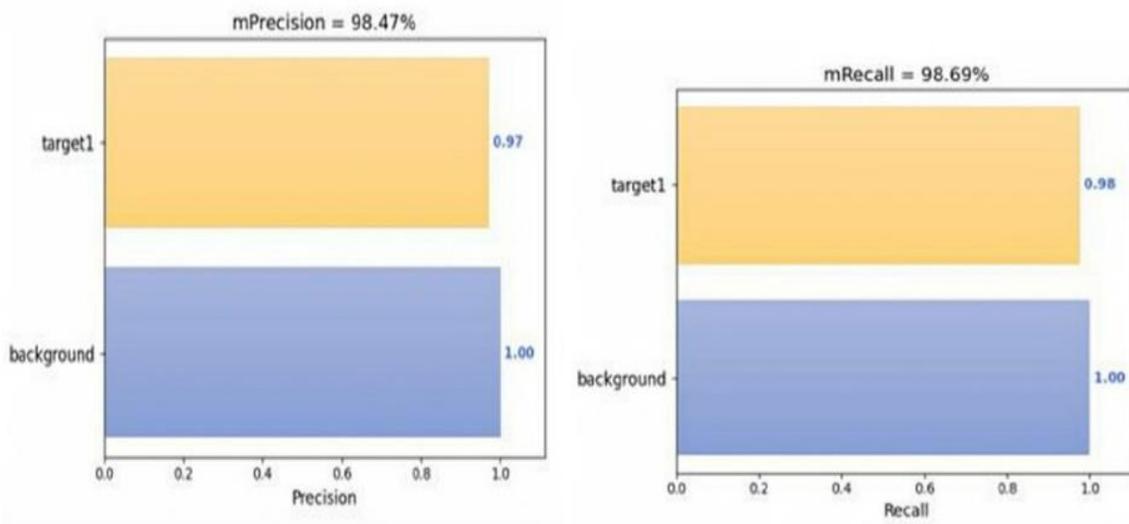


Figure 8. Precision and recall metrics of RCNN based on ART

Figures 7 and 8 present the Dice, MIoU, Precision, and Recall parameters of our method. After 50 rounds of the epoch, the Dice metric is 98.66%, the MIoU metric is 97.32%, the Precision metric is 98.47%, and the Recall metric is 98.69%. This demonstrates that the cell picture segmentation results produced by our model are rather accurate. We evaluated our method's overall performance by contrasting it with alternative methods.

Table 1. Comparison of performance with other models as CNN, LR

Method	Dice (%)	MioU (%)
CNN	91.35	91.18
LR	93.05	90.31
R-CNN	98.66	97.32

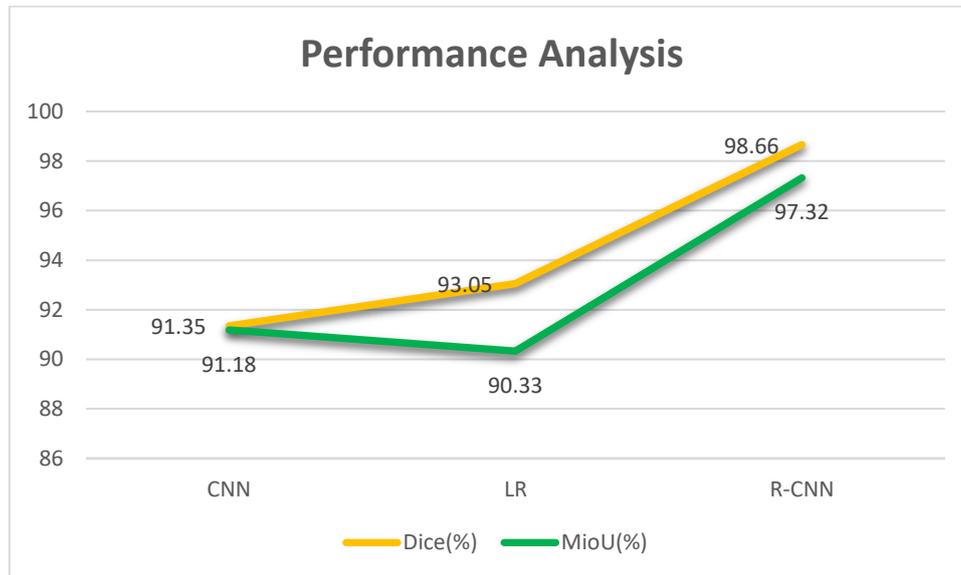


Figure 9. Dice and MIoU metrics analysis

We contrasted our approach with other algorithms to examine its overall performance. We contrasted our approach with LR and CNN. The findings of comparing the Dice and MIoU metrics across the various methods were displayed in Table 1. Our approach outperformed the others in terms of Dice and MIoU assessment measures, as shown in Table 1 and figure 9. The approach suggested in this research has a Dice of 98.66% and a MIoU of 97.32%. The performance increase of our suggested strategy is 6.14% in MIoU and 7.31% in Dice when compared to the CNN method. Our technique's Dice and MIoU both improved by 5.61% and 7.01%, respectively, in comparison to the LR method. Consequently, the fasterRCNN approach suggested in this research performs better in terms of segmentation and classification

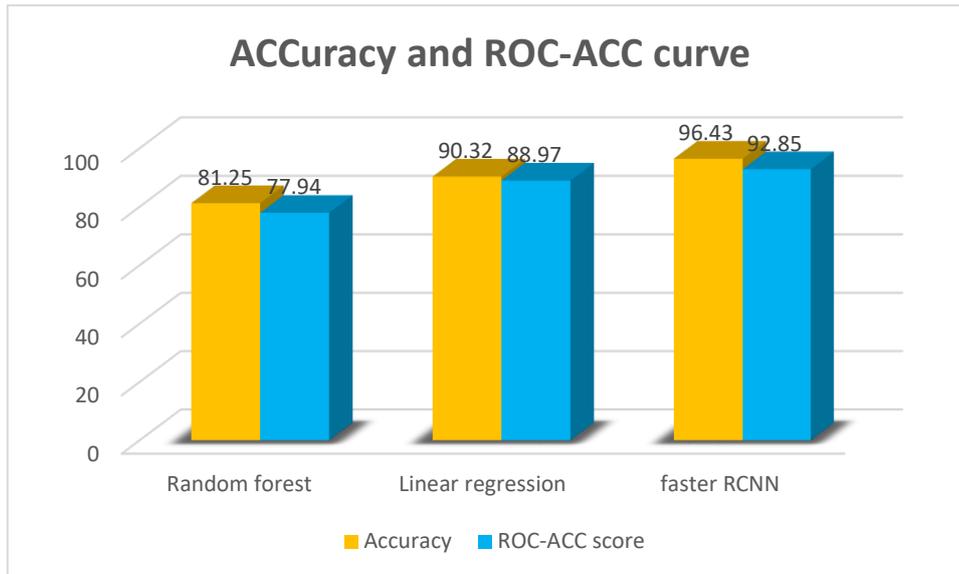


Figure 10. Accuracy and ROC-ACC curve

The proposed model is evaluated using, accuracy, and ROC-ACC score to ensure robust and reliable performance. Furthermore, the accuracy and ROC-ACC scores of follicle monitoring in IVF using R-CNN are compared with random forest and linear regression models, demonstrating the improved performance of the deep learning approach. This comparative analysis is presented in Figure 10, which illustrates the performance differences between the models.

5) Conclusion

The proposed a deep learning-based 3D vision framework for embryo assessment and follicle tracking, which integrates preprocessing, segmentation, and classification techniques in in vitro fertilization (IVF). Non-local means (NLM) filtering and normalization were used to improve image quality, while U-Net was used for accurate segmentation of reproductive structures. A region-based convolutional neural network (R-CNN) effectively classified embryos and follicles, ensuring automatic, objective, and accurate assessment. Performance analysis demonstrated that the proposed method outperformed traditional machine learning models, such as random forest and linear regression, with high accuracy and ROC-ACC scores, as shown in Figure 2. Evaluation metrics, including precision, recall, Dice coefficient, and MIoU, confirmed the robustness of this approach. By automating the grading process, this method reduces subjectivity, improves decision-making in assisted reproductive technology (ART), and improves IVF success rates.