

Visualized hysteroscopic artificial intelligence fertility assessment system for endometrial injury: an image-deep-learning study

Bohan Li^a, Hui Chen^{b,c} and Hua Duan^a

^aDepartment of Minimally Invasive Gynecologic Center, Beijing Obstetrics and Gynecology Hospital, Capital Medical University, Beijing Maternal and Child Health Care Hospital, Beijing, China; ^bSchool of Biomedical Engineering, Capital Medical University, Beijing, China; ^cBeijing Advanced Innovation Center for Big Data-based Precision Medicine, Capital Medical University, Beijing, China

ABSTRACT

Objective: Asherman's syndrome (AS) is a significant cause of subfertility in women from developing countries. Over 80% of AS cases in these regions are linked to dilation and curettage (D&C) procedures following pregnancy. The incidence of AS in patients with infertility and recurrent miscarriage can be as high as 10%, while the pregnancy rate in cases of moderate to severe adhesions can be as low as 34%. We aimed to establish a hysteroscopic artificial intelligence system using image-deep-learning algorithms for fertility assessment.

Methods: This diagnostic study included 555 cases with 4922 hysteroscopic images from a Chinese intrauterine adhesions cohort clinical database (NCT05381376). The study evaluated two image-deep-learning algorithms' effectiveness in predicting pregnancy within one year, using AUCs and decision curve analysis. The models' performance was evaluated for two-year prediction via concordance index and cumulative time-dependent ROC. A quantifiable visualization panel of the system was established.

Results: The proportional hazard CNN system accurately predicted conception, with AUCs of 0.982, 0.992, and 0.990 in three randomly assigned datasets, superior to the InceptionV3 framework, and achieved a net benefit of 69.4% for subfertility assessment. The system fitted well with c-indexes of 0.920–0.940 and was time-stable. The quantifiable visualization panel displayed four intrauterine pathologies intuitively. The performance was comparable to senior hysteroscopists, with a kappa value of 0.84–0.89.

Conclusions: The CNN based on the proportional hazard approach accurately assesses fertility postoperatively. The quantifiable visualization panel could assist in intrauterine pathologies assessment, optimize treatment strategies, and achieve individualized and cost-efficient practices.

ARTICLE HISTORY

Received 23 August 2023

Revised 17 May 2024

Accepted 4 January 2025



KEYWORDS


Image deep learning; hysteroscopy; endometrial injury; Asherman's syndrome; subfertility; ART triage

Introduction

Artificial intelligence (AI) can improve health conditions and enhance mechanistic discoveries [1]. Advances in technologies and algorithms have allowed quick diagnosis through deep learning techniques, thus expanding the application of artificial intelligence in gynecology [2]. Hysteroscopy is usually used to examine intrauterine anatomy. Nonetheless, exploration of AI is relatively scarce in hysteroscopy. Asherman's syndrome (AS) is a syndrome characterized by the presence of hysteroscopically confirmed intrauterine adhesions (IUA), which causes morphological anomalies and subfertility, resulting in infertility and recurrent pregnancy loss [3–6]. AS

can promote subfertility through injury/destruction of the endometrium, preventing blastocyst implantation [7]. D&C following pregnancy is reported to be the leading cause of AS, accounting for over 80% of cases. However, AS can also occur in women with infertility or recurrent miscarriage, with a prevalence as high as 10% [7]. This is accompanied by a high postoperative recurrence rate (33%–66%) and a low conception rate (34%–54%) [8,9]. Assisted reproduction techniques (ART) can prompt conception for patients before adhesions recur [10], however, leading to increased treatment costs [11,12]. Thus, it is essential to assess the necessity of ART interventions [13].

CONTACT Hua Duan  duanhua@ccmu.edu.cn  Department of Minimally Invasive Gynecologic Center, Beijing Obstetrics and Gynecology Hospital, Capital Medical University, Beijing Maternal and Child Health Care Hospital, No. Qihelou Street, Dongcheng District, Beijing, 100006, China

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/07853890.2025.2478473>.

© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

Machine learning has emerged as a promising tool in the field of AS diagnosis and treatment. Building on prior research, our study utilizes decision trees and XGBoost, powerful machine learning algorithms, to predict pregnancy outcomes for AS patients based on clinical data [14,15]. 3D sonohysterography is a non-invasive imaging technique that provides a three-dimensional view of the uterine cavity, which aids in diagnosing AS [16]. In addition to clinical data, recent studies have explored the use of 3D ultrasound parameters for AI-based prediction of AS outcomes. This multi-modal approach offers a valuable new perspective for assessing prognosis [17]. Although these studies have demonstrated good predictive results, there are some limitations. Previous AI approaches in AS prognosis relied on physician ratings and manual measurements, which can introduce bias and inconsistency due to subjectivity. Furthermore, the lack of validation on external datasets limits their real-world applicability in clinical settings [15,17].

Hysteroscopy is recognized as the gold standard for diagnosing AS. It is also a method used to evaluate morphological anomalies and endometrial receptivity [13,18]. Nevertheless, clinicians require a long learning cycle to evaluate fertility, and empirical observations need better consistency. Deep-learning approaches can ameliorate such issues. Despite the emergence of several studies on the application of deep learning algorithms, a noticeable research gap persists in the investigation of these algorithms in the analysis of hysteroscopic images (2). Furthermore, there is a paucity of research exploring the potential of deep learning approaches in the subfertility field, underscoring the need for further investigation to evaluate the feasibility and effectiveness of these techniques in this clinical context.

Nevertheless, the limitation of deep-learning models is that most studies have focused on the accuracy of the models ignoring their interpretability, leading to the lack of suitable scenarios in the clinical setting [1,19–21]. The intuitive visualisation of results reduces the cost of doctor-patient communication and facilitates individualised treatment [22]. For this, the study established two types of deep-learning frameworks for systematic subfertility risk management. Deep learning networks based on time as a dependent variable can learn the highly intricate and linear/nonlinear associations between prognostic clinical characteristics and an individual's risk of subfertility [23]. Therefore, DeepSurv, a deep network proportional hazard model, was tried out with CNN to establish an image deep learning framework as the proportional hazard CNN.

This study hypothesizes that our hysteroscope-specific CNN integrated with PH modeling can provide superior predictions of ongoing pregnancy outcomes

within one year. Specifically, we postulate that this approach can stably predict the probability of pregnancy at each time point compared to traditional transfer learning models for subfertility risk assessment. The proposed method could be a reliable tool for clinical decision-making in infertility treatment if validated. Besides, the system was designed to encompass a quantifiable visualization panel for intrauterine pathologies, which allows clinicians to develop appropriate clinical strategies depending on the AI results.

Materials and methods

Database establishment and data collection

Data were obtained from the IUADB (ClinicalTrials.gov; NCT05381376). The study was approved and supervised by the Research Ethics Committee of the Beijing Obstetrics and Gynecology Hospital. The Chinese Ethics Committee of Registering Clinical Trials approved the revised version (protocol number; IEC-B-03-v01-FJ1 and ChiECRCT20220180). Each participant had provided written informed assent for the Collection and Application of Clinical Sample and Medical Data certified. All patients (732) were recruited from the establishment of the database (December 2018) to January 2020. The clinical information was collected from all patients (Supplementary Table 1).

Study design and participants

Inclusion criteria mainly followed the principles of database case (Table 1).

We followed the patients for at least two years after their second-look hysteroscopy procedure. During this follow-up period, we considered a successful ongoing pregnancy, confirmed by ultrasound at or beyond 12 weeks, as a positive outcome [18]. The sample size was analyzed by the power of test efficacy (Supplementary Table 2). The design schematic of the study is shown in Figure 1.

Table 1. Inclusion and exclusion criteria.

Inclusion criteria
1. Patients with previously diagnostic hysteroscopy confirmed adhesions combined with infertility or recurrent pregnancy loss (early pregnancy loss >2);
Patients desiring to conceive (20–45 years);
2. Patients without gynecological endocrine disorders (e.g. PCOS) and AMH > 1ng/ml;
3. No sperm abnormalities in the male partner;
4. Patients who signed informed consent form;
Exclusion criteria
1. Patients without second-look hysteroscopic images
2. Patients with substandard hysteroscopic image angles
3. Patients with low-resolution hysteroscopic images
4. Patients with black and white images
5. Patients with distorted scanned images of paper reports.

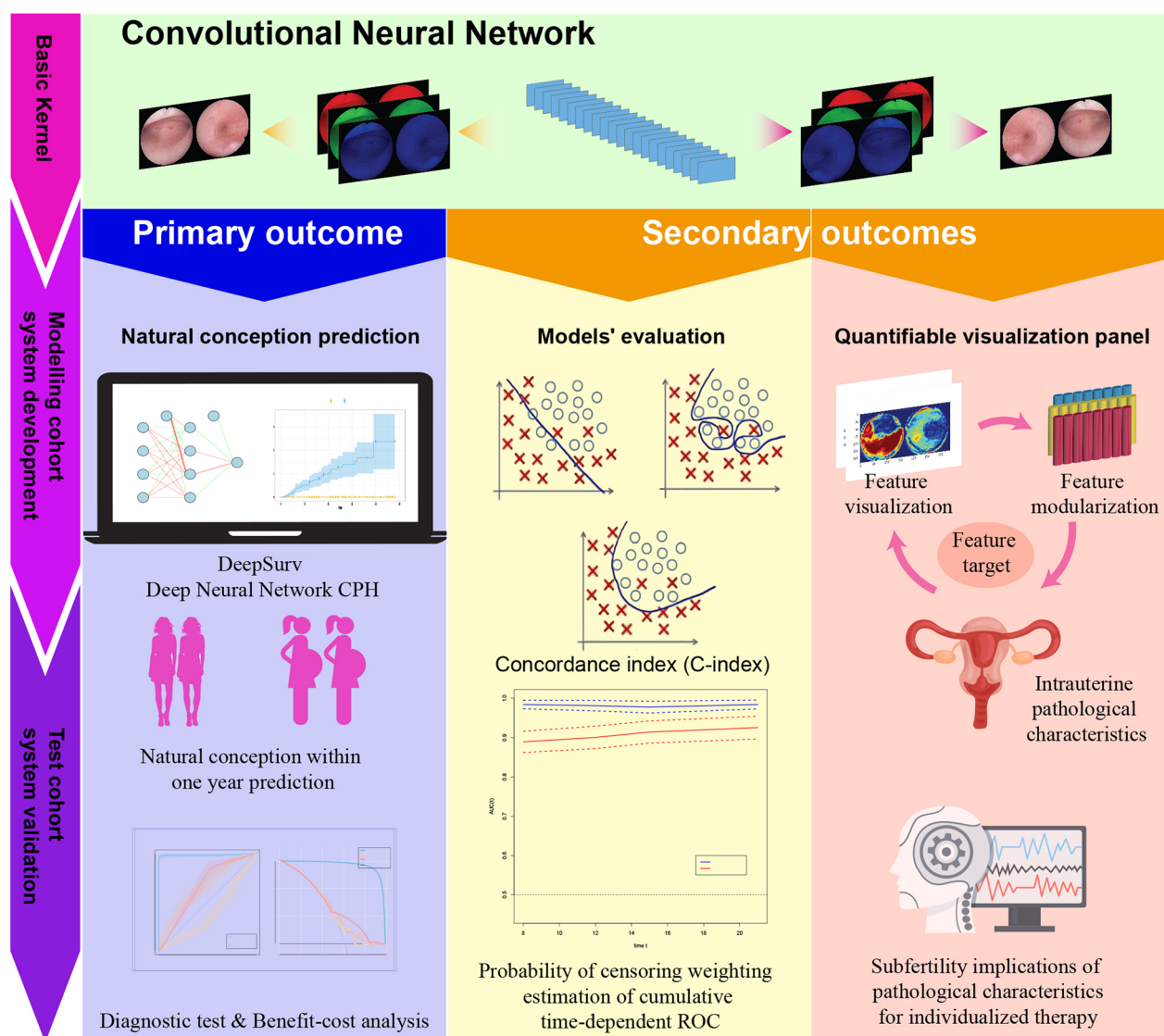


Figure 1. The comprehensive study can be divided into two phases: modelling and testing the system. The selected hysteroscopic images are processed by CNN. Differences in processing algorithms ultimately achieve three outcomes (one primary outcome and two secondary outcomes). Proportional hazard CNN system comprised the convolutional neural network combined with deep neural network CPH (Deepsurv). The system was trained (training set) and tuned (validation set) using the images from the modelling cohort. To assess the system's performance and advantages, we conducted diagnostic tests and cost-benefit analyses in a separate test cohort. This evaluation compared the system's accuracy to established methods like the InceptionV3 classifier and other relevant indicators. The secondary outcomes were the models' evaluation and a quantifiable visualization panel. Concordance index (c-index) and cumulative time-dependent ROC curve were performed to evaluate the two models for their fitting and time-stable. To achieve the system's quantifiable visualization panel, unsupervised learning, and Grad-CAM algorithms were employed to visualize and modularize extracted features clustered to corresponding intrauterine pathological characteristics. The clustering analysis revealed distinct patient subgroups based on their features. This suggests that the identified clusters could be valuable for guiding personalized treatment approaches.

The study was reported according to the STARD [24] (Supplementary Table 3).

Diagnosis and treatment of Asherman's syndrome

The diagnosis of AS was conducted following established guidelines, and then confirmed *via* hysteroscopy. Experienced endoscopic surgeons performed

hysteroscopic adhesiolysis in all patients under general anesthesia to remove intrauterine adhesions visualized during the procedure. Olympus S70 series equipment and physiological saline perfusion were used. General anesthesia was administered *via* endotracheal intubation combined with intravenous anesthesia. Routine cervical preparation was performed in the later stages before surgery [25,26]. The surgery was guided by

transabdominal ultrasound. During hysteroscopy, the morphology of uterine cavity and extent of adhesions were fully observed, tissue adhesion was examined, and scar tissue was excised to ensure that the residual endometrium allowed the uterine cavity to return to a normal shape without adhesions. After adhesiolysis, a suitable balloon was placed to create an intrauterine physical barrier, which was removed after 5–7 days [27]. After 3 months of postoperative estrogen and progesterone therapy, hysteroscopic postoperative exploration was conducted, and patients were encouraged to attempt conception. Patients who failed to conceive after 1 year and had no evident adhesion recurrence upon re-examination were encouraged to undergo assisted reproductive technology (ART) treatment.

Convolutional neural network model development

A total of 4922 hysteroscopic images were screened and processed for model training and validation. The detailed process of image collection is shown in [Supplementary Table 4](#). This study compared two types of CNN models: the proportional hazard and the InceptionV3 framework CNN classifier. Models were constructed in TensorFlow 2's Keras-based environment.

The proportional hazard CNN combines convolutional neural networks and proportional hazard models, allowing it to learn and predict how fertility outcomes change over time. This makes it particularly suitable for fertility assessments. InceptionV3 is a pre-trained deep learning model that has shown outstanding performance in image classification tasks. Therefore, its performance in predicting fertility outcomes was compared with that of the designed proportional hazard CNN.

For the proportional hazard CNN: Appropriate methods for proportional hazard CNN were explored by designing and training an autoencoder for flatbed images (14112 abstract features). To train our proportional hazards deep learning model, we employed the DeepSurv algorithm. This algorithm adjusted the model's internal parameters to optimize its ability to predict the likelihood of events (like pregnancy) over time. The loss function is the average negative log partial likelihood with regularization:

$$l_{\theta} = -\frac{1}{N_{E=1}} \sum_{i \in E_1=1} \left(\hat{h}_{\theta}(x_i) - \log \sum_{j \in \mathcal{N}(T_i)} e^{\hat{h}_{\theta}(x_j)} \right) + \lambda \cdot \|\theta\|_2^2 \quad [26,28].$$

This establishes a time-related deep learning model.

For the InceptionV3: We directly applied the pre-trained InceptionV3 model to predict fertility outcomes, while fine-tuning its last layer to adapt to our task.

Patients in the modeling cohort were randomly divided into training and validation sets (8:2) for parameter adjusting ([Supplementary Table 5](#)). To

optimize the performance of our model, we performed hyperparameter tuning by adjusting model parameters based on performance on the validation set to find the optimal model configuration. We employed grid search method to explore different combinations of hyperparameters to find the best hyperparameter settings. Moreover, we utilized cross-validation to assess the stability of the model across different data intervals within the training set ([Supplementary Table 6](#)).

Statistical analysis

Z tests were used to determine the significance of the AUC of the receiver operating characteristic (ROC) curve via the pROC package in R. Decision curve analysis (DCA) and clinical impact curve (CIC) were performed using the DecisionCurve package in R [29]. Missing values (less than 10%) were imputed using multiple imputations (mice package; in R, version 3.6.2).

Quantifiable visualization panel

Gradient-weighted Class Activation Mapping (Grad-CAM) is a technique for creating visual explanations that use the gradients of an arbitrary target concept that feed into the final convolutional layer to create a coarse localization map that highlights essential regions in the image for concept prediction. We leveraged this capability to develop visualization panels for proportional hazard CNN that can highlight intrauterine pathologies.

We performed a quantitative analysis of intrauterine pathologies using an unsupervised learning algorithm. Specifically, we clustered the codes with high correlation in the encoding matrix using the weighted correlation (Pearson approach, $a_{ij} = |cor(x_i, y_j)|^{\beta}$). We then connected the clustered network with clinical traits that correlated highly with the codes. This process enabled us to produce quantifiable image features reflecting intrauterine conditions [30,31]. Feature correlation modules corresponded to branches of the resulting hierarchical clustering tree (dendrogram) [32].

Results

Patients

The Five hundred fifty-five patients were randomly divided into the modeling and test cohorts (3:1) for at least two years of follow-up. The baseline clinical characteristics are shown in [Supplementary Table 7](#). Nevertheless, patients in whom severe adhesions recurred during follow-up, requiring a second

procedure, were terminated from the cohort and classified as treatment failures. Finally, 379 and 102 cases in the modeling (303 for parameter training and 76 for hyper-parameter validating) and test cohorts, respectively, were included for analysis (Supplementary Figure 1). The modeling cohort was divided into 303 cases for parameter training (training set) and 76 cases for validating hyperparameter tuning (validation set).

Of the 303 women assigned to the training set, 112 (36.96%) achieved ongoing pregnancy during the follow-up period, including 71 within one year.

Of the 76 women assigned to the validation set, 36 (47.37%) achieved ongoing pregnancy during the follow-up period, including 18 within one year.

Of the 102 women assigned to the test cohort, 39 (38.24%) achieved ongoing pregnancy during the follow-up period, including 24 within one year.

Primary outcome

The primary outcome of this study was predicting the natural conception within one year, which can effectively screen out subfertility patients. The study tested two types of CNN networks: proportional hazard and InceptionV3 framework network. AUCs of the proportional hazard approach achieved 0.982(95%CI; 0.967–0.994) in the training set, 0.992(95%CI; 0.979–1) in the validation set, and 0.990(95%CI; 0.974–1) in the test cohort. Those were significantly superior to the InceptionV3 framework classifier [AUCs of 0.917 (95%CI; 0.893–0.941), 0.779 (95%CI; 0.668–0.891), 0.885 (95%CI; 0.828–0.943)] and the traditional predictors ($P_{\text{all}} < 0.05$) (Figure 2A–C). After excluding subjects who had undergone ART treatment in the training set ($n=4$), validation set ($n=2$), and test cohort ($n=1$) within one year, the proportional hazard approach achieved AUCs of 0.982 (95% CI; 0.971–0.995), 0.992 (95% CI; 0.978–1), and 0.989 (95% CI; 0.973–1), respectively, which was also significantly better than others (Supplementary Figure 2). Similarly, the proportional hazard approach performed remarkably in other diagnostic tests, including sensitivity, specificity, Youden index (YI), positive predictive value (PPV), and negative predictive value (NPV), as detailed in Table 2. These results suggest that the proportional hazard approach can effectively predict natural conceptions within one year compared with the transfer learning approach and traditional predictors.

Accurate indicators represent a favorable clinical benefit rate reflected by clinical decision and impact curves. The conception failure rate of 76.47% within one year was used as the risk (intervention) threshold for net benefit prediction. The proportional hazard CNN approach showed that patients in the test cohort

had a net benefit of 66.8 (95% CI; 59.3, 73.5)%, 68.2 (95% CI; 56.4, 82.9)%, and 69.4(95% CI; 54.6–82.4) % at this threshold, outperforming others (Figure 2D–F).

Stratified analyses of the test cohort were conducted based on clinical traits that may affect outcomes to validate the stability of the model. Cox univariate analysis identified 14 indicators that may affect outcomes (Supplementary Figure 3). No clinical traits could significantly affect the system accuracy. A sensitivity analysis of the predicted results was also conducted by sequentially dropping each operator, furtherly demonstrating the model's robustness (Supplementary Figure 4).

Secondary outcomes

Secondary outcomes of this study include models' evaluations and establishing a quantifiable visualization panel for the image-deep-learning system. When time were added for dependent variable, concordance index (c-index) and inverse probability of censoring weighting estimation of cumulative time-dependent ROC curve should be performed to evaluate the two models. C-indexes of the proportional hazard vs. InceptionV3 model were 0.940 (95% CI; 0.924–0.953) vs. 0.991 (95% CI; 0.980–1), 0.920 (95% CI; 0.889–0.950) vs. 0.829 (95% CI; 0.728–0.930), 0.925 (95% CI; 0.901–0.949) vs. 0.932 (95% CI; 0.889–0.974) in the training set, validation set, and test cohort, respectively, all with high performance. AUCs exhibited that, with timepoint variations, the prediction results of proportional hazard CNN were more stable than those of the InceptionV3 classifier during the follow-up period (2years) (Figure 3). This indicates that the proportional hazard CNN is capable of accurately predicting the probability of pregnancy for patients at each time point. This approach avoids the limitations of the InceptionV3 classifier, which can only provide predictions for a single point in time. Our proportional hazard CNN, in contrast, offers a significant advantage by predicting fertility outcomes across time. This allows clinicians to gain a more comprehensive understanding of a patient's fertility potential and make more informed treatment decisions in managing subfertility caused by AS.

The quantifiable visualization panel offered the results of imaging and quantitative evaluation of hysteroscopic features in patients with or without successful conception, respectively. Figure 4A shows the imaging results of the pathologies performed by Grad-CAM. Patients with failed conception are highlighted red for unfavorable factors, indicating morphological anomalies and occluded fallopian tube ostia by adhesion bands. In contrast, patients with successful

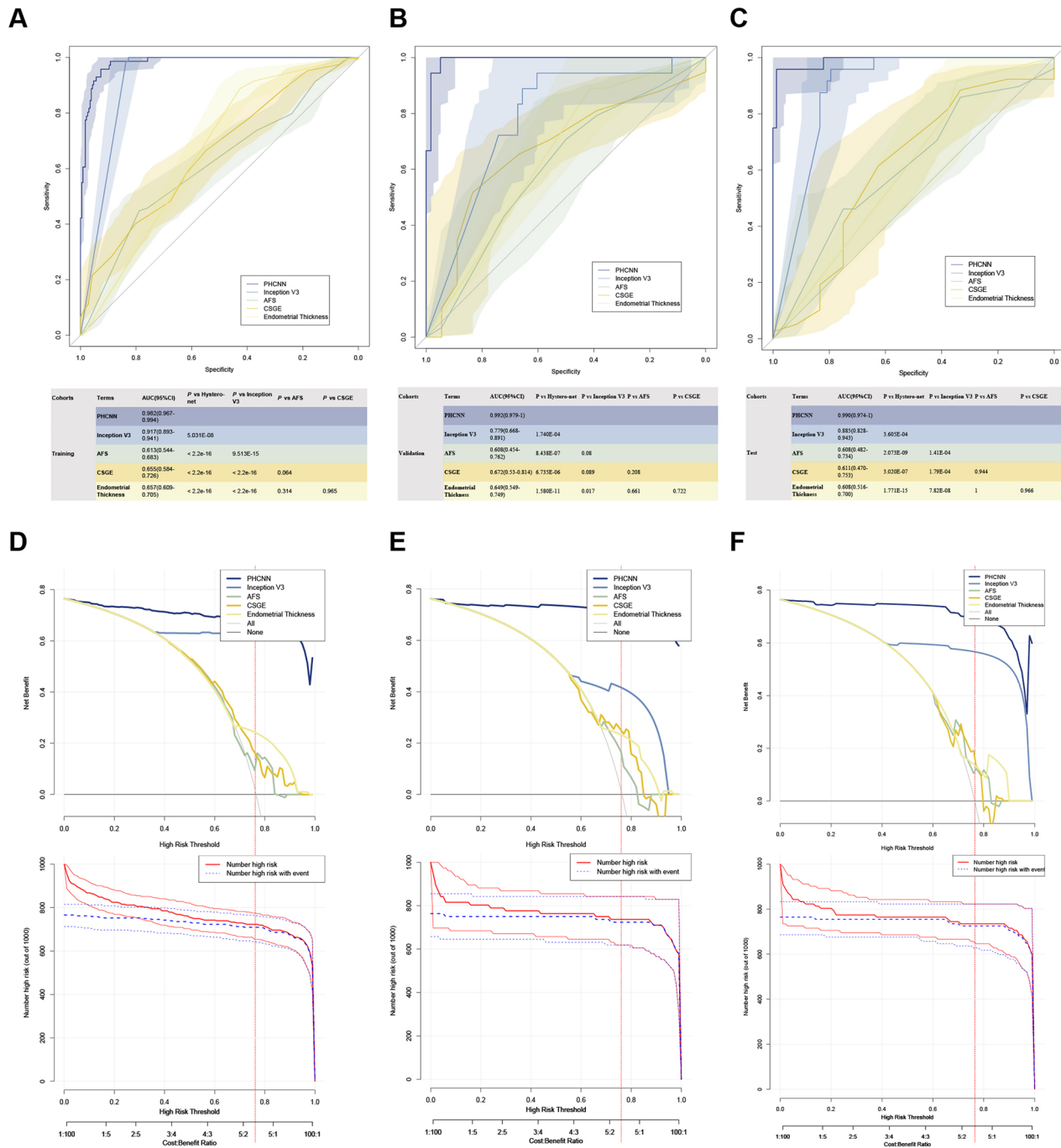


Figure 2. One-year conceiving prediction evaluations. ROC of proportional hazard CNN, InceptionV3, AFS, CSGE, and endometrial thickness for predicting of successful conception within one year in the training set (A), validation set (B), and test cohort (C). The corresponding AUCs for each method are presented in the table below. Decision curve analysis (DCA) and clinical impact curve (CIC) of the proportional hazard CNN system for the training set (D), validation set (E), and test cohort (F). The X-axis represents the threshold probability for infertility outcome; the Y-axis represents the net benefit. The net benefit of all machine learning models was larger over the range of clinical threshold (subfertility rate) compared to the reference model and other indicators. AFS=American Fertility Society score, CSGE=Chinese Society of Gynecological Endoscopy score.

conception are highlighted blue for favorable aspects, representing healthy endometrium and adequate blood supply. The subfertility weights were highly clustered as Uterine cavity morphology features (UCMFs, $p=4\times 10^{-12}$), Fallopian tube ostia features (FTOFs, $p=10^{-11}$), Endometrial thickness features (ETFs, $p=10^{-11}$), and

Blood supply features (BSFs, $p=2\times 10^{-44}$). The coefficient of determination (R^2) of these characteristics for subfertility was 15.21% ($p=6\times 10^{-15}$), 36% ($p=3\times 10^{-38}$), 11.56% ($p=2\times 10^{-11}$), and 16.81% ($p=4\times 10^{-17}$), in order.

Figure 4B presents the model results alongside two examples of its quantifiable visualization panels.

Table 2. Diagnostic test of each evaluation metrics in the three datasets.

Cohorts	Term	Sensitivity (%)	Specificity (%)	YI	PPV (%)	NPV (%)
Training	PHCNN	95.77 (88.1 – 99.1)	92.67 (88.5 – 95.7)	0.8845	80 (71.6 – 86.4)	98.6 (95.9 – 99.5)
	InceptionV3	100 (94.9 – 100.0)	82.76 (77.3 – 87.4)	0.8276	64 (57.2 – 70.2)	100
	AFS	44.83 (38.3 – 51.5)	78.87 (67.6 – 87.7)	0.237	87.4 (81.2 – 91.7)	30.4 (27.0 – 34.1)
	CSGE	54.93 (42.7 – 66.8)	66.38 (59.9 – 72.4)	0.2131	33.3 (27.5 – 39.8)	82.8 (78.6 – 86.3)
	Endometrial thickness	88.73 (79.0 – 95.0)	42.67 (36.2 – 49.3)	0.314	32.1 (29.2 – 35.2)	92.5 (86.4 – 96.0)
Validation	PHCNN	100 (81.5 – 100.0)	94.83 (85.6 – 98.9)	0.9483	85.7 (66.6 – 94.8)	100
	Inception V3	94.44 (72.7 – 99.9)	60.34 (46.6 – 73.0)	0.5479	42.5 (34.5 – 50.9)	97.2 (83.7 – 99.6)
	AFS	70.69 (57.3 – 81.9)	50 (26.0 – 74.0)	0.2069	82 (73.6 – 88.2)	34.6 (22.3 – 49.4)
	CSGE	51.72 (38.2 – 65.0)	83.33 (58.6 – 96.4)	0.3506	90.9 (77.6 – 96.7)	34.9 (27.7 – 42.9)
	Endometrial thickness	88.89 (65.3 – 98.6)	41.38 (28.6 – 55.1)	0.3027	32 (26.4 – 38.2)	92.3 (75.8 – 97.9)
Test	PHCNN	95.83 (78.9 – 99.9)	98.72 (93.1 – 100.0)	0.9455	95.8 (76.6 – 99.4)	98.7 (91.9 – 99.8)
	Inception V3	97.44 (86.5 – 99.9)	96.83 (89.0 – 99.6)	0.9426	95 (82.9 – 98.7)	98.4 (89.8 – 99.8)
	AFS	71.79 (55.1 – 85.0)	49.21 (36.4 – 62.1)	0.21	46.7 (39.0 – 54.5)	73.8 (61.7 – 83.1)
	CSGE	69.23 (52.4 – 83.0)	50.79 (37.9 – 63.6)	0.2002	46.6 (38.6 – 54.7)	72.7 (61.1 – 81.9)
	Endometrial thickness	87.18 (72.6 – 95.7)	47.62 (34.9 – 60.6)	0.348	50.7 (44.2 – 57.3)	85.7 (71.8 – 93.4)

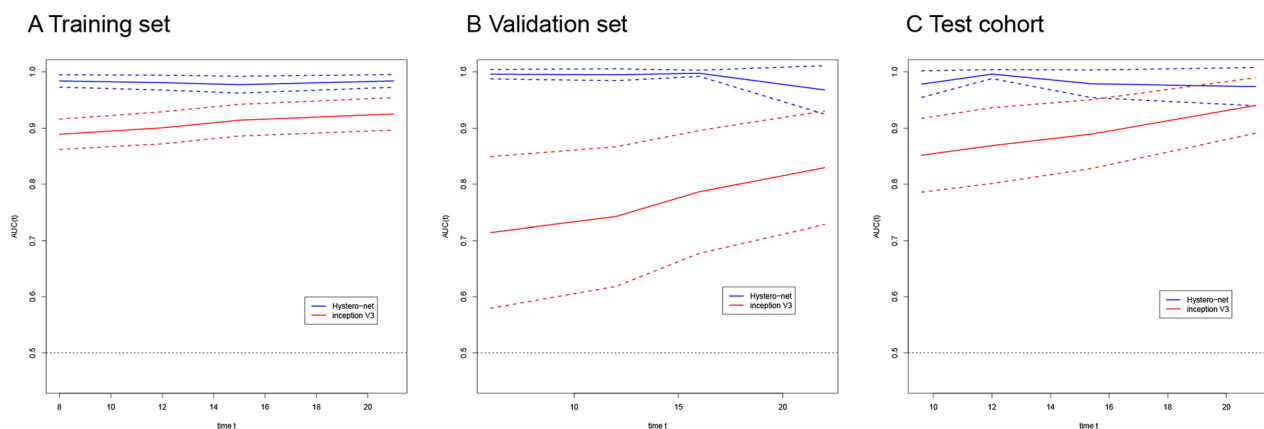


Figure 3. Inverse probability of censoring weighting estimation of cumulative time-dependent ROC curve of (A) training set, (B) validation set, (C) test cohort. The horizontal axis represents the duration of follow-up time, while the vertical axis represents the area under the curve (AUC) used for predicting ongoing pregnancy at each time point. This graph illustrates the relationship between the AUC and the duration of follow-up time, providing insights into the predictive ability of ongoing pregnancy.

The process of training the panel and region of interest (ROI) within the convolutional layers were described in detail in [Supplementary Figure 5](#). The model analyzed intrauterine pathologies and calculated assessment scores or probabilities based on cluster characteristics. These scores were then compared to the assessment made by two hysteroscopsists. To facilitate analysis, each score was classified as either low or high-risk, according to the cut-off value. With respect to risk factors indicating the presence of conception difficulties, there was high agreement between the weights quantification scheme of the image-deep-learning system and the assessment of the senior hysteroscopist. The kappa values were 0.89 (95%CI; 0.842–0.937) [Inadequate blood supply], 0.86 (95%CI; 0.81–0.91) [Insufficient endometrial thickness], 0.84 (95%CI; 0.79–0.891) [occluded fallopian tube ostia], and 0.837 (95%CI; 0.783–0.890) [morphological anomalies], respectively.

Discussion

For Asherman's syndrome, the lack of effective methods for assessing fertility after adhesiolysis often contribute to over-reliance on ART, resulting in potential medical resources waste and healthcare costs increase. This study is a novel attempt to develop an image deep-learning system, based on hysteroscopy to predict postoperative patients' fertility. Additionally, a panel was designed to quantitatively visualize the intrauterine pathologies of at-risk populations to support individualized therapy.

This study selected hysteroscopic images for deep learning to predict reproductive outcomes for AS. It is generally caused by damage to the endometrium's basal layer. Potential risk factors include D&C, other intrauterine procedures, and infections. Studies have demonstrated a relatively high prevalence of AS (19.4%) in women who have undergone uterine fibroid

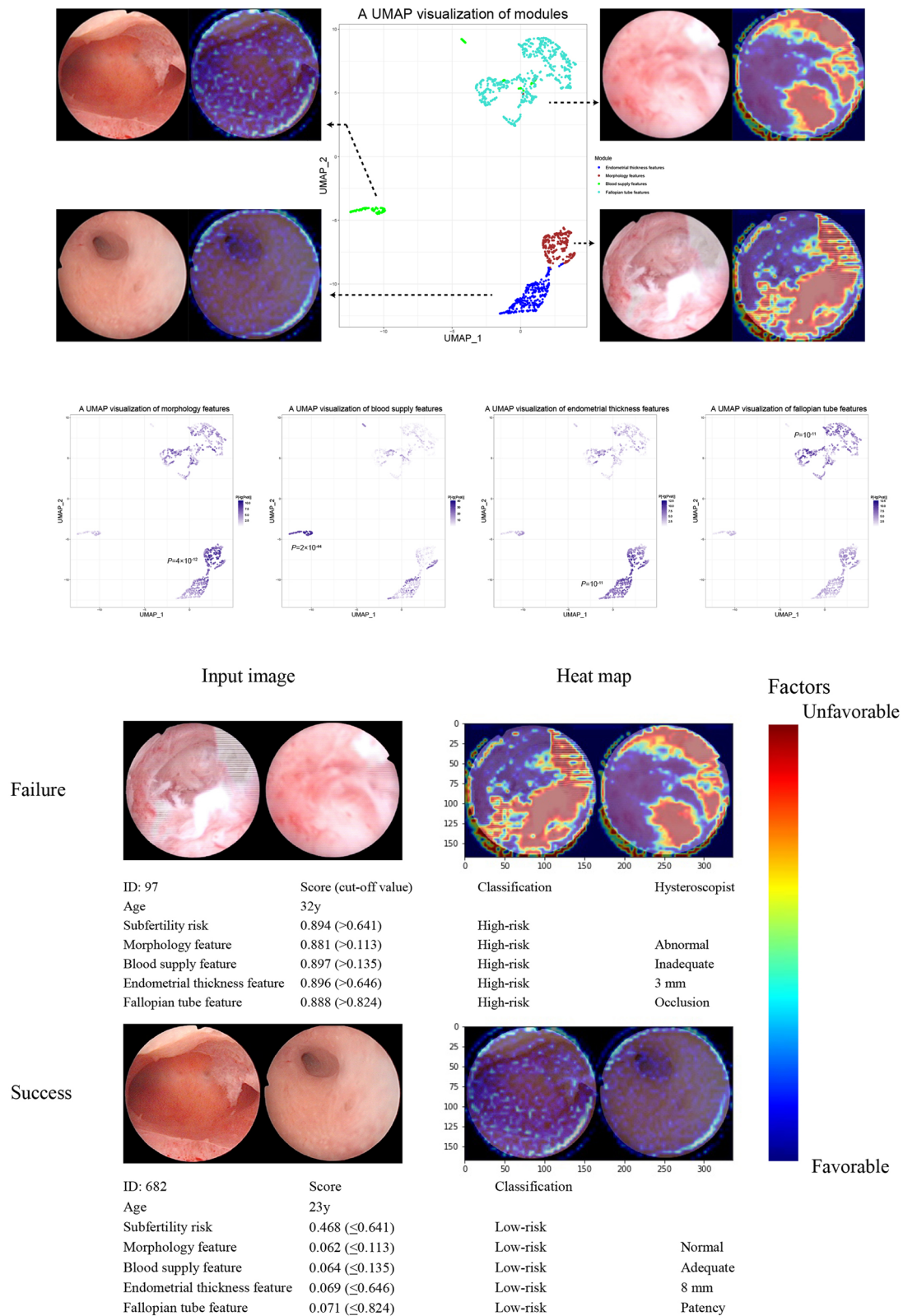


Figure 4. Examples of quantifiable visualization panel presentations (two examples). Both cases were from the test cohort. The case with successfully conceiving was 35-yo, while the failed one was 24-yo—both with an AFS score of 10 preoperatively. (A) UMAPs of each intrauterine pathological feature. Darkens of blue represent a significant relationship between the module and the corresponding pathology. (B) Visualization examples. The two examples are the input pictures (left) and the weighted heat map (right) for conception failure (top) and success (bottom) subjects, respectively. Patients with failed conception are highlighted red for unfavorable factors, while patients with successful conception are highlighted blue for favorable aspects. UCMF=uterine cavity morphology feature, FTOF=fallopian tube ostia feature, ETF=endometrial thickness feature, BSF=blood supply feature.

removal [32]. This high prevalence increases the risk of infertility in women [7]. Hysteroscopy is an essential part of fertility evaluation. A limited number of studies have reported on the ability of endometrial status and endometrial vascular distribution by direct intrauterine visualization to assess endometrial receptivity, all with favorable results. For the two methods, sensitivities were 75% and 71.43%, while specificities were 71.43% and 61.11%, respectively [13]. Computer-aided diagnostic techniques can effectively shorten the learning curve. The accuracy of hysteroscopic assessment is further advanced.

In the field of reproductive health, previous studies have demonstrated the potential value of AI models. GhoshRoy et al. [33,34] proposed a male fertility prediction model based on support vector machine, random forest, extreme gradient boosting (XGB), and SMOTE techniques, visualising the model's interpretation using SHAP. Similarly, in our previous research focusing on AS, we also developed predictive models based on clinical data [14,15,35], utilising algorithms such as decision trees and XGBoost, which demonstrated high levels of predictive efficiency and model interpretability. With regard to data analysis, AI models have the potential to enhance data augmentation, feature extraction, and interpretability [36], particularly in the context of computer vision in this field, which still requires further exploration. This study presents a preliminary exploration of the prediction of fertility using AS hysteroscopic images and visualises the risk of intrauterine areas affecting fertility based on model interpretation. These explorations contribute to a better understanding of the role of artificial intelligence in fertility analysis and the potential to improve reproductive health outcomes.

The advantage of deep learning for images (CNN approach) over traditional machine learning models is that tens of thousands of parameters are generated as the number of convolutional layers increases, which allows CNN to fit high-performance models with small sample sizes. As reported, an InceptionV3 framework model was developed using 123 ureteroscopy images to discriminate the types of urinary calculi, which achieved an AUC of 99% [37]. Besides, A CNN to predict chemotherapy response rates in rectal cancer used 303 cases as modeling data getting a high performance as well [19]. The issue for CNN models is frequently their low accuracy in other datasets due to overfitting instead of underfitting. To this end, this study took advantage of prospective data, which can standardize the image acquisition process. This reduced the heterogeneity issue caused by the different acquisition angles and locations, resulting in a promising

performance in all three data tests. Furthermore, we designed virtual image frameworks to standardize the position of graphic acquisitions [25,26,38], which could further reduce training costs and increase clinical applicability in subsequent studies.

We investigated two models for predicting hysteroscopic reproductive outcomes: a proportional hazard-based CNN and a transfer learning classifier. Transfer learning, where a pre-trained model is adapted to a new task, is a common approach in computer vision. InceptionV3, a popular pre-trained model, excels at image classification due to its extensive training on large datasets [26,39]. However, these classification models may not effectively capture the temporal aspects crucial for predicting fertility outcomes. In this study, we compared two deep learning approaches for predicting hysteroscopic reproductive outcomes: a proportional hazard model and a transfer learning classifier based on InceptionV3. Notably, the proportional hazard model achieved superior performance, particularly when considering time-dependent factors. This advantage was due to the model's inherent ability to incorporate time as a dependent variable. This allows for more stable and accurate predictions across the entire follow-up period, making it more suitable for risk stratification compared to the InceptionV3-based transfer learning model.

Traditional AS evaluation systems (AFS, ESGE, or CSGE) can stratify and predict reproductive outcomes for intrauterine adhesions *via* a risk-stratification system [40,41]. A retrospective study by Cao et al. [41] analyzed the effectiveness of different AS evaluation systems in assessing live birth rates, with AUCs for AFS and CSGE being 0.663 and 0.684, respectively, which were consistent with our results. Endometrial thickness is also an essential indicator with high sensitivity and low specificity for predicting fertility [13]. This was proven by our results, as well. In this study, the proportional hazard CNN system was superior in diagnosis, compared to the aforementioned methods. Additionally, the model's net benefit of prejudgment was 3–5 times higher than other indicators. All of these results demonstrate their potential utility for clinical applications.

Application of deep learning in healthcare face two major hurdles: limited generalizability and lack of interpretability [1]. Clinicians require clinically interpretable data to develop effective treatment plans [19]. This aligns with medical ethics principles, where doctors often explain treatment options and patient conditions in terms of probabilities rather than absolute certainties. Previous deep-learning classifiers do not adequately overcome these challenges [42]. To address these issues, we developed a quantifiable visualization panel to

visually and quantitatively characterize the four pathological characteristics: Uterine cavity morphology, fallopian tube ostia, endometrial thickness, and blood supply. It has been suggested that all of these pathological characteristics may play some role (>15%) in embryo implantation by Zilberberg et al. [43]. In this study, these pathological factors were also found to contribute to an increased risk of subfertility. The quantifiable visualization panel allows subfertility classification based on hysteroscopic pathological characteristics, which can assist in individualized therapy postoperatively.

Because there is no practical method to prevent the recurrence of IUA, timely ART treatment may achieve a successful pregnancy before re-adhesions. However, the criteria for ART triage and the timing of intervention remain controversial. The study identified 90–180 days after the procedure as the optimal window for IVF success, but this timeframe coincides with a significant increase in treatment costs. To select high-risk patients for ART triage, the proportional hazard CNN system, which is flexible and accurate, would be a reasonable choice. Additionally, appropriate intervention options are of equal clinical concern. Previous studies have shown that retreatment of severe uterine cavity abnormalities that persist after surgery may be

one of the options to consider for improving the success rate of ART [10]. Regarding postoperative occluded fallopian tube ostia, IVF is undoubtedly the preferred solution [44]. Insufficient endometrial thickness is, obviously, also an essential factor contributing to subfertility. The use of the EMT as a tool for deciding whether to terminate the cycle, freeze all embryos, or forego further IVF treatment remains controversial. Current data suggest that the EMT has a limited ability to identify women with a low chance of becoming pregnant after IVF. Regimens in ongoing clinical trials, such as amnion grafts, stem cell therapy, and platelet-rich plasma, appear to be alternative options for endometrial repair [45–47]. Blood supply is also critical to successful ongoing pregnancies, and the success rate of IVF can be more than twice as high with an adequately perfused uterus as with a poorly perfused one. Most current treatment modalities are synergistic with endometrial repairs [48]. Identifying these pathological characteristics allows subgrouping and individualizing treatment for the subfertility population. We designed the corresponding clinical application flow according to the model, as shown in Figure 5.

AI-based visualization interfaces can also enhance traditional hysteroscopy training. While surgeons require substantial hands-on experience to develop

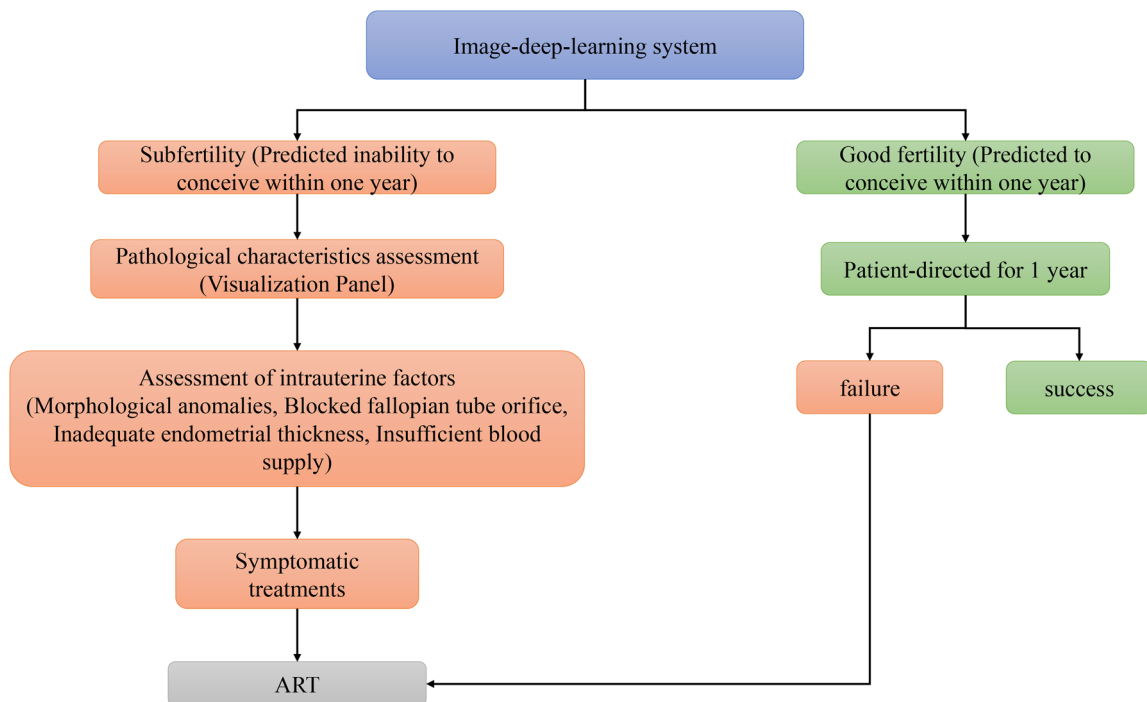


Figure 5. Image-deep-learning system clinical application flow. Postoperative patients with better fertility predicted by proportional hazard CNN can attempt to conceive spontaneously for one year, while patients with predicted subfertility may recommend ART triage. Based on available evidence, individualized treatment may be helpful if intrauterine pathologies (morphological abnormalities, occluded fallopian tube ostia, insufficient endometrial thickness, and inadequate blood supply) are detected negatively to affect successful conception. ART=assisted reproductive technology.

proficiency in diagnostic techniques [43], AI can provide real-time feedback during the learning process. These intuitive interfaces can highlight potential indicators of infertility and intrauterine abnormalities, expediting skill development.

Despite the promising findings, this study has some limitations. While this pilot study highlights the potential of image-deep learning model, it was based on a small sample size. Larger datasets typically are needed to improve the generalizability of the model. Therefore, we plan to collect additional data for further validation. Moreover, the study is currently at the stage of partial analysis, used for model exploration. However, sample size calculations indicated sufficient power (over 90%) to draw reliable conclusions from this initial phase. In future, we shall recruit more data to conduct a comprehensive validation to assess the model's generalizability in real-world clinical use.

Conclusion

This study presents a novel deep learning model utilizing proportional hazard CNN. The model demonstrates promising accuracy in predicting fertility outcomes for patients with AS who have undergone hysteroscopic adhesiolysis. Our study achieved two key results. First, the model significantly outperformed existing methods in predicting pregnancy outcomes for patients. Second, the accompanying quantitative visualization panel offers clinicians an intuitive tool. This tool allows them to not only understand the AI model's predictions but also leverage them to develop personalized treatment plans for each patient. This study offers a promising initial demonstration of how AI technology can be applied in fertility assessment. The model and visualization panel provide a novel perspective and valuable tools for precision medicine in the clinical management of fertility.

Acknowledgements

We are equally grateful to the database entry staff and hysteroscopists of Beijing Obstetrics and Gynecology Hospital and Sir Run Run Shaw Hospital, especially Prof. Yinshu Guo, Yanzhen Peng, Ying Zhang, Xin Wang, Jinjuan Wang, and Fengxian Fu for their professional reports and photographic acquisition. In addition, we appreciate Prof. Hui Chen from the School of Biomedical Engineering, Capital Medical University for her professional guidance on database statistics.

Authors contributions

BL, HC and HD designed the study and developed the conceptual ideas. BL collected all the input sources and additional data, annotated the images, BL and HC implemented

the main algorithms and other computational analysis, analyzed the results. BL and HD have accessed and verified all the data in the study. HD had final responsibility for the decision to submit for publication. All authors read and approved the final version of the manuscript.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Ethical approval and consent to participate

The study was approved and supervised by the Research Ethics Committee of the Beijing Obstetrics and Gynecology Hospital. The Chinese Ethics Committee of Registering Clinical Trials approved the revised version (protocol number; IEC-B-03-v01-FJ1 and ChiECRCT20220180). Each participant had provided written informed assent for the Collection and Application of Clinical Sample and Medical Data certified.

Funding

This study was financially supported by the National Key Research and Development Program of China (2018YFC1004803).

Data availability statement

The raw data supporting the conclusions of this article will be made available on request from the corresponding author, without undue reservation.

References

- [1] Rajpurkar P, Chen E, Banerjee O, et al. AI in health and medicine. *Nat Med*. 2022;28(1):31–38. doi:10.1038/s41591-021-01614-0.
- [2] Akazawa M, Hashimoto K. Artificial intelligence in gynecologic cancers: current status and future challenges – a systematic review. *Artif Intell Med*. 2021;120:102164. doi:10.1016/j.artmed.2021.102164.
- [3] Vitale SG, Riemma G, Carugno J, et al. Postsurgical barrier strategies to avoid the recurrence of intrauterine adhesion formation after hysteroscopic adhesiolysis: a network meta-analysis of randomized controlled trials. *Am J Obstet Gynecol*. 2022;226(4):487–498.e8. doi:10.1016/j.ajog.2021.09.015.
- [4] Magalhaes RS, Williams JK, Yoo KW, et al. A tissue-engineered uterus supports live births in rabbits. *Nat Biotechnol*. 2020;38(11):1280–1287. doi:10.1038/s41587-020-0547-7.
- [5] Yi X, Liu F, Gao K, et al. Reconstructable uterus-derived materials for uterus recovery toward efficient live births. *Adv Mater (Deerfield Beach, Fla)*. 2022;34(8):e2106510. doi:10.1002/adma.202106510.
- [6] Hanstede MMF, van der Meij E, Veersema S, et al. Live births after Asherman syndrome treatment. *Fertil Steril*. 2021;116(4):1181–1187. doi:10.1016/j.fertnstert.2021.05.099.

- [7] Carbonnel M, Pirtea P, de Ziegler D, et al. Uterine factors in recurrent pregnancy losses. *Fertil Steril*. 2021;115(3): 538–545. doi:10.1016/j.fertnstert.2020.12.003.
- [8] Chang Y, Duan H, Shen X, et al. Controversy in the management of oestrogen therapy before hysteroscopic adhesiolysis: a systematic review and meta-analysis. *Reprod Biomed Online*. 2020;41(4):715–723. doi:10.1016/j.rbmo.2020.06.012.
- [9] Hanstede MM, van der Meij E, Goedemans L, et al. Results of centralized Asherman surgery, 2003–2013. *Fertil Steril*. 2015;104(6):1561–1568.e1. doi:10.1016/j.fertnstert.2015.08.039.
- [10] Bosteels J, van Wessel S, Weyers S, et al. Hysteroscopy for treating subfertility associated with suspected major uterine cavity abnormalities. *Cochrane Database Syst Rev*. 2018;12(12):Cd009461. doi:10.1002/14651858.CD009461.pub4.
- [11] Deng K, Song X-H, Han X-M, et al. Optimal waiting period for fresh embryo transfer after hysteroscopic adhesiolysis: a retrospective cohort study. *Chin Med J (Engl)*. 2019;132(19):2333–2339. doi:10.1097/CM9.0000000000000456.
- [12] Chambers GM, Sullivan EA, Ishihara O, et al. The economic impact of assisted reproductive technology: a review of selected developed countries. *Fertil Steril*. 2009;91(6):2281–2294. doi:10.1016/j.fertnstert.2009.04.029.
- [13] Craciunas L, Gallos I, Chu J, et al. Conventional and modern markers of endometrial receptivity: a systematic review and meta-analysis. *Hum Reprod Update*. 2019;25(2):202–223. doi:10.1093/humupd/dmy044.
- [14] Li Y, Duan H, Wang S. An XGBoost predictive model of ongoing pregnancy in patients following hysteroscopic adhesiolysis. *Reprod Biomed Online*. 2023;46(6):965–972. doi:10.1016/j.rbmo.2023.01.019.
- [15] Zhu R, Duan H, Xu W, et al. Decision tree model predicts live birth after surgery for moderate-to-severe intrauterine adhesions. *BMC Pregnancy Childbirth*. 2022;22(1):78. doi:10.1186/s12884-022-04375-x.
- [16] Laganà AS, Ciancimino L, Mancuso A, et al. 3D sonohysterography vs hysteroscopy: a cross-sectional study for the evaluation of endouterine diseases. *Arch Gynecol Obstet*. 2014;290(6):1173–1178. doi:10.1007/s00404-014-3366-6.
- [17] Sun D, Yi S, Zeng F, et al. Developing and validating a prediction model of live birth in patients with moderate-to-severe intrauterine adhesions: a new approach with endometrial morphology measurement by 3D transvaginal ultrasound. *Quant Imaging Med Surg*. 2024;14(1):995–1009. doi:10.21037/qims-23-1014.
- [18] Smit JG, Kasius JC, Eijkemans MJC, et al. Hysteroscopy before in-vitro fertilisation (inSIGHT): a multicentre, randomised controlled trial. *Lancet*. 2016;387(10038):2622–2629. doi:10.1016/S0140-6736(16)00231-2.
- [19] Feng L, Liu Z, Li C, et al. Development and validation of a radiopathomics model to predict pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer: a multicentre observational study. *Lancet Digit Health*. 2022;4(1):e8–e17. doi:10.1016/S2589-7500(21)00215-6.
- [20] Zhao K, Li Z, Yao S, et al. Artificial intelligence quantified tumour-stroma ratio is an independent predictor for overall survival in resectable colorectal cancer. *EBioMedicine*. 2020;61:103054. doi:10.1016/j.ebiom.2020.103054.
- [21] Hosny A, Parmar C, Quackenbush J, et al. Artificial intelligence in radiology. *Nat Rev Cancer*. 2018;18(8):500–510. doi:10.1038/s41568-018-0016-5.
- [22] Gulbrandsen P. Shared decision making: improving doctor-patient communication. *BMJ*. 2020;368:m97. doi:10.1136/bmj.m97.
- [23] She Y, Jin Z, Wu J, et al. Development and validation of a deep learning model for non-small cell lung cancer survival. *JAMA Netw Open*. 2020;3(6):e205842. doi:10.1001/jamanetworkopen.2020.5842.
- [24] Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ*. 2015;351:h5527. doi:10.1136/bmj.h5527.
- [25] Li B, Chen H, Duan H. Development of a one-step office hysteroscopic lightweight artificial intelligence application for subfertility risk stratification of intrauterine adhesions. *Fundam Res*. 2024;
- [26] Li B, Chen H, Duan H. Artificial intelligence-driven prognostic system for conception prediction and management in intrauterine adhesions following hysteroscopic adhesiolysis: a diagnostic study using hysteroscopic images. *Front Bioeng Biotechnol*. 2024;12:1327207. doi:10.3389/fbioe.2024.1327207.
- [27] Zhu R, Duan H, Gan L, et al. Comparison of intrauterine suitable balloon and foley balloon in the prevention of adhesion after hysteroscopic adhesiolysis. *Biomed Res Int*. 2018;2018:9494101–9494106. doi:10.1155/2018/9494101.
- [28] Katzman JL, Shaham U, Cloninger A, et al. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Med Res Methodol*. 2018;18(1):24. doi:10.1186/s12874-018-0482-1.
- [29] Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making*. 2006;26(6):565–574. doi:10.1177/0272989X06295361.
- [30] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9(1):559. doi:10.1186/1471-2105-9-559.
- [31] Blockwise Network Analysis of Large Data. Peter Langfelder; 2018.
- [32] Li B, Wang S, Duan H, et al. Discovery of gene module acting on ubiquitin-mediated proteolysis pathway by co-expression network analysis for endometriosis. *Reprod Biomed Online*. 2021;42(2):429–441. doi:10.1016/j.rbmo.2020.10.005.
- [33] GhoshRoy D, Alvi PA, Santosh K. Explainable AI to predict male fertility using extreme gradient boosting algorithm with SMOTE. *Electronics*. 2022;12(1):15. doi:10.3390/electronics12010015.
- [34] GhoshRoy D, Alvi PA, Santosh KC. Unboxing industry-standard AI models for male fertility prediction with SHAP. *Healthcare (Basel, Switzerland)*. 2023;(7):11.
- [35] Li B, Duan H, Wang S, et al. Gradient boosting machine learning model for defective endometrial receptivity prediction by macrophage-endometrium interaction modules. *Front Immunol*. 2022;13:842607. doi:10.3389/fimmu.2022.842607.
- [36] GhoshRoy D, Alvi PA, Santosh KC. AI tools for assessing human fertility using risk factors: a state-of-the-art review. *J Med Syst*. 2023;47(1):91. doi:10.1007/s10916-023-01983-8.
- [37] El Beze J, Mazeaud C, Daul C, et al. Evaluation and understanding of automated urinary stone recognition methods. *BJU Int*. 2022;130(6):786–798. doi:10.1111/bju.15767.

- [38] Li B, Chen H, Lin X, et al. Multimodal learning system integrating electronic medical records and hysteroscopic images for reproductive outcome prediction and risk stratification of endometrial injury: a multicenter diagnostic study. *Int J Surg*. 2024;110(6):3237–3248. doi:[10.1097/J59.0000000000001241](https://doi.org/10.1097/J59.0000000000001241).
- [39] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z, editors. Rethinking the inception architecture for computer vision. *Proceedings of the IEEE Conference on Computer Vision and Pattern recognition*; 2016.
- [40] AAGL practice report: practice guidelines for management of intrauterine synechiae. *J Minim Invasive Gynecol*. 2010;17(1):1–7.
- [41] Cao M, Pan Y, Zhang Q, et al. Predictive value of live birth rate based on different intrauterine adhesion evaluation systems following TCRA. *Reprod Biol Endocrinol*. 2021;19(1):13. doi:[10.1186/s12958-021-00697-1](https://doi.org/10.1186/s12958-021-00697-1).
- [42] Stein HS, Guevarra D, Newhouse PF, et al. Machine learning of optical properties of materials – predicting spectra from images and images from spectra. *Chem Sci*. 2019;10(1):47–55. doi:[10.1039/c8sc03077d](https://doi.org/10.1039/c8sc03077d).
- [43] Zilberberg E, Smith R, Nayot D, et al. Endometrial compaction before frozen euploid embryo transfer improves ongoing pregnancy rates. *Fertil Steril*. 2020;113(5):990–995. doi:[10.1016/j.fertnstert.2019.12.030](https://doi.org/10.1016/j.fertnstert.2019.12.030).
- [44] Huang LN, Tan J, Hitkari J, et al. Should IVF be used as first-line treatment or as a last resort? A debate presented at the 2013 Canadian Fertility and Andrology Society meeting. *Reprod Biomed Online*. 2015;30(2):128–136. doi:[10.1016/j.rbmo.2014.10.004](https://doi.org/10.1016/j.rbmo.2014.10.004).
- [45] Evers JL. Female subfertility. *Lancet*. 2002;360(9327):151–159. doi:[10.1016/S0140-6736\(02\)09417-5](https://doi.org/10.1016/S0140-6736(02)09417-5).
- [46] van Eekelen R, Tjon-Kon-Fat RI, Bossuyt PMM, et al. Natural conception rates in couples with unexplained or mild male subfertility scheduled for fertility treatment: a secondary analysis of a randomized controlled trial. *Hum Reprod*. 2018;33(5):919–923. doi:[10.1093/humrep/dey051](https://doi.org/10.1093/humrep/dey051).
- [47] Shen M, Duan H, Lv R, et al. Efficacy of autologous platelet-rich plasma in preventing adhesion reformation following hysteroscopic adhesiolysis: a randomized controlled trial. *Reprod Biomed Online*. 2022;45(6):1189–1196. doi:[10.1016/j.rbmo.2022.07.003](https://doi.org/10.1016/j.rbmo.2022.07.003).
- [48] Guo J, Chaemsaitong P, Huang J, et al. Comparison of uterine artery Doppler measurements at 6 weeks of pregnancy after IVF between pregnancies that resulted in miscarriage and ongoing pregnancies. *Int J Gynaecol Obstet*. 2021;152(2):249–255. doi:[10.1002/ijgo.13371](https://doi.org/10.1002/ijgo.13371).