BMC Pregnancy and Childbirth

RESEARCH

Open Access

# Effectiveness and clinical impact of using deep learning for first-trimester fetal ultrasound image quality auditing

Xiaoyan Cao[1†], Binghan Li[2†], Yongsong Zhou[2], Yan Cao[3], Xin Yang[2], Xindi Hu[3], Chaoyu Chen[2], Shaokao Zhu[1], Hengli Lin[1], Tao Wang[1], Yuling Yan[1], Tao Tan[4], Lin Wang[1*] and Dong Ni[2]

## Abstract

**Background** Regular auditing of ultrasound images is required to maintain quality; however, manual auditing is time-consuming and can be inconsistent. We therefore aimed to develop and validate an artificial intelligence-based image quality audit (AI-IQA) system to audit images from the four key planes used in first-trimester scanning.

**Methods** The AI-IQA system was developed based on the YOLOv7 structure detection network and a multi-branch image quality regression network using a large multicenter internal dataset. Clinical validation was performed using 567 cases scanned by four radiologists with different experience levels, of which 349 were performed without AI-IQA feedback (clinical test set 1) and 218 were performed after 2–3 rounds of AI-IQA feedback (clinical test set 2). The proportion of standard images obtained and detailed expert audit results were compared to verify whether AI-IQA could objectively and accurately provide feedback on deficiencies in nonstandard images to assist radiologists at different experience levels in improving image quality.

**Results** In the internal test set, the AI-IQA system achieved high average accuracy precision, recall and F1-score in auditing the overall plane quality (0.881, 0.833, 0.842 and 0.837, respectively) and structure quality (0.906, 0.861, 0.857 and 0.859, respectively). In clinical test sets 1 and 2, AI-IQA results showed strong consistency with expert assessment results, with the average Cohen's Kappa coefficient exceeding 0.8 for all four planes. In addition, following AI-IQA feedback, the proportion of standard images obtained by junior and mid-level radiologists increased by 7.7% and 5.1%, respectively. AI-IQA takes only 0.05 s to assess each image, while experts require more than 20 s ($p < 0.001$).

**Conclusions** The proposed AI-IQA system proved to be a highly accurate and efficient method of automatically auditing first-trimester scanning image quality, providing precise and rapid key plane quality control. This tool can also assist radiologists with different levels of experience to improve the image quality.

**Keywords** Prenatal ultrasound, First-trimester scanning, Image quality control, Artificial intelligence, Deep learning

[†] Xiaoyan Cao and Binghan Li contributed equally to this work and are co-first authors.

*Correspondence:
Lin Wang
13728893488@163.com
[1] Ultrasound Department, Shenzhen Futian District Maternity & Child Healthcare Hospital, Shenzhen, Guangdong 518016, China
[2] National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen, Guangdong 518073, China
[3] Shenzhen RayShape Medical Technology Co., Ltd., Shenzhen, Guangdong 518071, China
[4] Faculty of Applied Sciences, Macao Polytechnic University, Macao SAR, Taipa Island 999078, China

## Background

Ultrasound (US) has become an important diagnostic tool for initial gestational screening, owing to its instantaneous results, low cost and noninvasive nature [1]. Although prenatal ultrasonography mainly focuses on the second trimester [2], technological advancements have made first-trimester scanning (FTS) increasingly important in modern medicine. Complete early screening requires standard planes from multiple views covering the fetal head, brain, neck, heart, abdomen, limbs, placenta and biometric anatomical regions [3]. It provides vital early fetal information such as fetal size and gestational age (GA), aiding in timely decision-making for subsequent care and interventions [4].

In FTS, there are four key planes closely related to fetal screening for chromosomal abnormalities, structural malformations and biometric measurements: the nuchal translucency (NT) plane (NTP), the midsagittal view of the fetus (MSF), the axial view of the fetal abdomen (AFA) and the axial view of the fetal head in the transventricular plane (AFTP) [5]. NTP allows for the observation of increased NT thickness, predicting the risk of chromosomal abnormalities and structural malformations [6–9]. MSF is used to measure crown-rump length (CRL) to calculate GA and detect facial malformations such as cleft lip and palate [10, 11]. AFA primarily visualizes the connection between the umbilical cord and abdominal wall as well as detects omphalocele and gastroschisis [12]. AFTP assists in identifying central nervous system malformations such as brain deformity [13, 14]. Obtaining good quality images in these four planes is vital for the reliability of FTS results, although a complete examination involves additional planes as outlined in International Society of Ultrasound in Obstetrics and Gynecology (ISUOG) guidelines [15].

Obtaining standard plane is fundamental for accurate measurements and diagnoses. Therefore, quality control of acquired US images is a critical task in clinical practice. Traditionally, this process relies on manual auditing, where experienced professionals visually assess the image quality. Although this approach can be reliable to some extent, it has several limitations. First, it imposes an additional burden on experts, requiring them to step away from actual clinical diagnostic work, which hampers the efficient use of medical resources. Second, manual auditing is inherently subjective, making it difficult to eliminate inter-observer variability [16]. Finally, this method is inefficient and struggles to meet the demands of large-scale screening or provide timely feedback [17]. Facing these limitations, rapidly advancing deep learning (DL) technologies offer a promising solution for US image quality control.

DL has become a vital tool for image quality analysis. Classification-based methods offer advantages in speed and reliability [18, 19], while contrastive learning and anomaly detection have shown potential in feature extraction [20–22]. For example, Qu et al. employed a differential convolutional neural network (DCNN) to accurately detect specific fetal brain planes [23]. However, such simple classification methods lack granular feedback. Additionally, some studies have evaluated image quality by detecting key anatomical structures [24, 25], but the presence of a structure does not fully reflect image quality. To address this, Dong et al. proposed a general deep learning framework that incorporates image gain and scaling analysis to identify the standard four-chamber heart view [26]. However, the multi-step nature of the process may lead to error accumulation. Despite significant advancements in US image quality analysis, limitations remain, particularly the lack of specific improvement feedback and the research gap in FTS applications.

To address these issues, we propose an artificial intelligent (AI)-based image quality audit (AI-IQA) system that innovatively integrates YOLOv7 for object detection and a multi-branch quality regression network for quality assessment. This integrated design not only improves detection accuracy but also provides granular feedback. We validated the system's reliability on four key planes in FTS, compared its efficacy with manual expert assessment and analyzed its utility for radiologists with different levels of experience.

## Methods

### Study design

The study design is illustrated in Fig. 1. This study collected data from multiple centers to form an internal dataset with which to develop the AI-IQA system. Clinical validation was performed using two clinical test sets: clinical test set 1 from four radiologists at different experience levels without AI-IQA feedback and clinical test set 2 from the same four radiologists after 2–3 rounds of AI-IQA feedback. An expert panel (each member with ≥ 15 years of clinical experience) independently evaluated image quality in the two clinical test sets.

The AI-IQA system's audits were then compared against the expert panel's assessments to evaluate its accuracy and consistency. In addition, we compared the proportion of standard images obtained in the two clinical test sets to verify whether AI-IQA could objectively and accurately provide feedback on the reasons for nonstandard images to improve the quality of images obtained by radiologists at different experience levels. This study also assessed the speed and efficiency of the AI-IQA system.

The study was approved by the Ethics Committee of Shenzhen Futian District Maternity & Child Healthcare Hospital (protocol number: K-2023-04-01) and

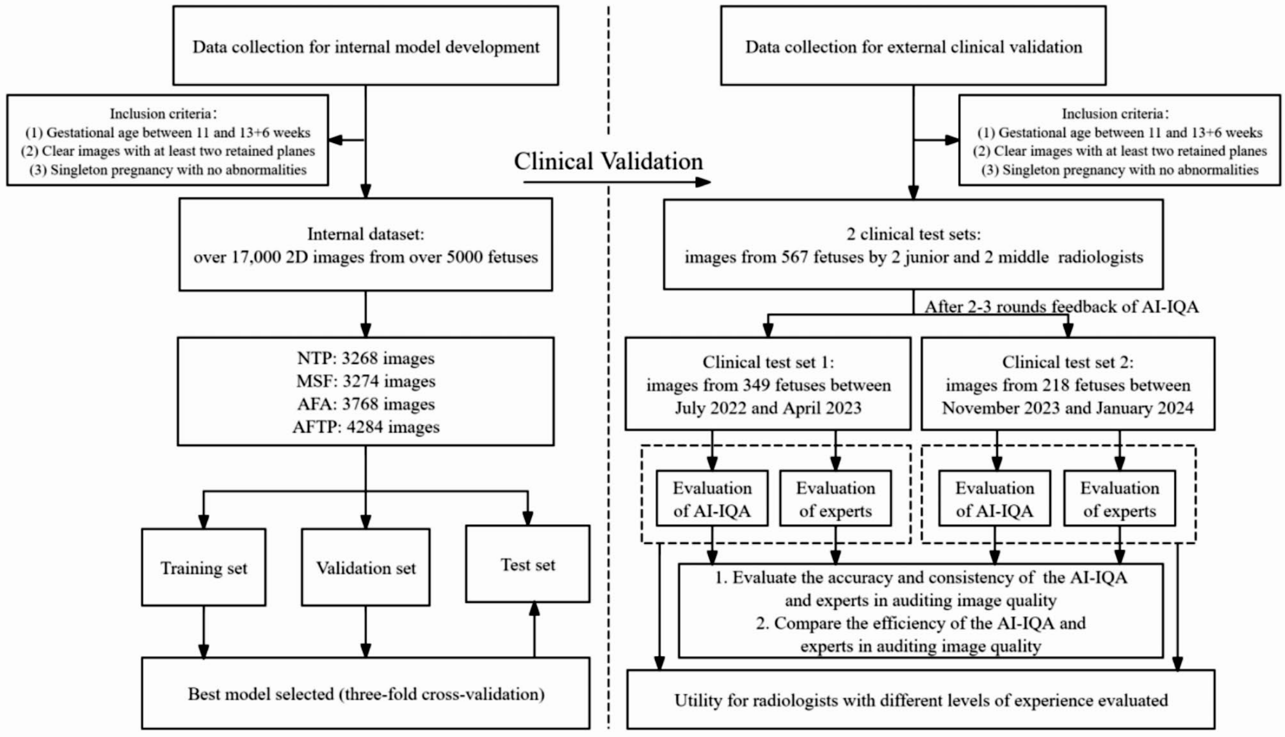

Data collection for internal model development
Inclusion criteria:
(1) Gestational age between 11 and 13+6 weeks
(2) Clear images with at least two retained planes
(3) Singleton pregnancy with no abnormalities
Internal dataset:
over 17,000 2D images from over 5000 fetuses
NTP: 3268 images
MSF: 3274 images
AFA: 3768 images
AFTP: 4284 images
Training set
Validation set
Test set
Best model selected (three-fold cross-validation)
Clinical Validation
Data collection for external clinical validation
Inclusion criteria:
(1) Gestational age between 11 and 13+6 weeks
(2) Clear images with at least two retained planes
(3) Singleton pregnancy with no abnormalities
2 clinical test sets:
images from 567 fetuses by 2 junior and 2 middle radiologists
After 2-3 rounds feedback of AI-IQA
Clinical test set 1:
images from 349 fetuses between
July 2022 and April 2023
Clinical test set 2:
images from 218 fetuses between
November 2023 and January 2024
Evaluation of AI-IQA
Evaluation of experts
Evaluation of AI-IQA
Evaluation of experts
1. Evaluate the accuracy and consistency of the AI-IQA and experts in auditing image quality
2. Compare the efficiency of the AI-IQA and experts in auditing image quality
Utility for radiologists with different levels of experience evaluated

**Fig. 1** Flowchart summarizing the study design. NTP: nuchal translucency plane, MSF: midsagittal view of the fetus, AFTP: axial view of fetal head in the transventricular plane, AFA: axial view of fetal abdomen, AI-IQA: artificial intelligence-based image quality audit

conducted in accordance with the principles outlined in the Declaration of Helsinki.

## Data collection

The inclusion criteria for images in this study were as follows: (1) GA between 11 and 13$^{+6}$ weeks; (2) clear fetal images with at least two planes retained on NTP, MSF, AFA and AFTP; (3) singleton pregnancy with no abnormalities. Adhering to the inclusion criteria, a total of more than 17,000 2D images from 5000 fetuses obtained during FTS were collected from multiple centers to form an internal dataset for model development, including 3268 NTP, 3274 MSF, 3768 AFA and 4284 AFTP images. The images were acquired using US machines of various brands, including GE, Mindray, Samsung, Philips and Siemens. The internal dataset was split into training set, validation set, and internal test set in a 7:2:1 ratio based on patient data, using three-fold cross-validation. The validation set was used for hyperparameter tuning, while the internal test set was used for model performance evaluation.

Clinical validation was performed at the Department of Ultrasound in Shenzhen Futian District Maternity & Child Health Hospital, with all participants providing written informed consent. Images of 349 normal fetuses were collected and examined by two junior (<5 years of experience) and two mid-level (5–10 years of experience) radiologists as clinical test set 1, comprising 388 NTP, 397 MSF, 377 AFA and 380 AFTP images, between July 2022 and April 2023. Following 2–3 rounds of AI-IQA feedback, the same four radiologists obtained images of 218 fetuses as clinical test set 2, comprising 249 NTP, 265 MSF, 264 AFA and 244 AFTP images, between November 2023 and January 2024. The distribution of each plane image, along with the corresponding dataset, is illustrated in Table 1.

**Table 1** Distribution of images in clinical test sets 1 and 2

| Dataset | Plane | Junior group | Mid-level group |
|---|---|---|---|
| Clinical test set 1 | NTP (*n* = 388) | 140 (36.08%) | 248 (63.92%) |
| | MSF (*n* = 397) | 146 (36.78%) | 251 (63.22%) |
| | AFA (*n* = 377) | 144 (38.20%) | 233 (61.80%) |
| | AFTP (*n* = 380) | 142 (37.37%) | 238 (62.63%) |
| Clinical test set 2 | NTP (*n* = 249) | 115 (46.18%) | 134 (53.82%) |
| | MSF (*n* = 265) | 117 (44.15%) | 148 (55.85%) |
| | AFA (*n* = 264) | 137 (51.89%) | 127 (48.11%) |
| | AFTP (*n* = 244) | 119 (48.77%) | 125 (51.23%) |

Data are expressed as number (percentage). NTP: nuchal translucency plane, MSF: midsagittal view of the fetus, AFA: axial view of fetal abdomen, AFTP: axial view of fetal head in the transventricular plane

## Markers and annotations

In this study, image annotation was conducted using the medical imaging intelligent software Pair [27] (version 2.6; Shenzhen, China), developed by Shenzhen RayShape

Medical Technology Co., Ltd. The annotation process was divided into three stages:

(1) Image Quality Categorization: Six mid- and senior-level radiologists strictly followed ISUOG guidelines [3] and Fetal Medicine Foundation standards to audit the four-plane image quality. Images were categorized as standard or nonstandard.

(2) Anatomical Structure Annotation: Ten experienced radiologists annotated the main anatomical structures within the planes using bounding boxes, which amplified the AI-IQA system's understanding of local anatomical structures.

(3) Detailed Audit Annotation: Six mid- and senior-level radiologists comprehensively annotated audit details across multiple dimensions as listed in Table 2. Main structures with clear, well-defined boundaries were labeled as good; otherwise, as bad. Mutually exclusive structures were classified as visible or invisible based on their presence, while overall image quality, including image clarity, image zoom and fetal position, was categorized as fit or unfit. These criteria, aligned with ISUOG guidelines and refined by experts with over 15 years of experience, trained the AI-IQA system to identify and provide feedback on nonstandard images.

Before formal annotation, all annotators underwent comprehensive training, including detailed criteria explanations, case studies and practice sessions. Only those who passed a qualification assessment by senior experts proceeded with formal annotation. Discrepancies during annotation were resolved through consensus by two additional experts, who also reviewed all annotations to ensure accuracy. Examples of annotated NTP, MSF, AFA and AFTP images are shown in Fig. 2.

### Development of the AI-IQA system

To achieve a comprehensive evaluation of both global image features (e.g., fetal position, image clarity) and local anatomical structures, the AI-IQA system integrates YOLOv7 and a ResNet50-based regression network for object detection and quality assessment [28, 29], respectively, as shown in Fig. 3. In the initial detection phase, YOLOv7 identifies regions of interest (ROIs) and key anatomical structures (Table 2), crucial for image quality. These features are then input into the ResNet50-based regression network, which uses residual connections and weight-sharing to extract complex features. The network is split into a main branch and multiple structural branches, each consisting of a global average pooling layer and two fully connected layers. The structural branches learn and output specific structural scores, while the main branch combines these weighted features to output the overall quality score. The final score is calculated by weighted averaging of the component scores. Each score is categorized into binary results based on a 60-point threshold. If mutually exclusive structures are detected, the final score is forced below the threshold, indicating a nonstandard image.

Our experiments were conducted using PyTorch 1.12.1, on a workstation equipped with NVIDIA GeForce RTX 2080 Ti. During data preprocessing, input images were resized to 416 × 416 pixels and normalized with mean= (0.485, 0.456, 0.406) and std= (0.229, 0.224, 0.225). For online data augmentation, YOLOv7 applied default augmentations such as geometric transformations (scaling: 0.5x-1.5x, translation: 20–40 pixels, rotation: ±10 degrees), color adjustments and noise injection, while ResNet50 applied more conservative augmentations (scaling: 0.9x-1.1x, translation: 10–20 pixels, rotation: ±5 degrees) to preserve fetal anatomy. To avoid gradient conflicts and simplify the process, a staged training strategy was employed. In training phase, YOLOv7 used

**Table 2** Evaluation criteria used by radiologists to annotate specific plane details

| Plane | Main structure<br>(Good, Bad) | Mutually exclusive structure<br>(Visible, Invisible) | Overall image evaluation<br>(Fit, Unfit) |
|---|---|---|---|
| NTP | NT, IT, Palate, Calvaria,<br>Interbrain, Midbrain,<br>Nasal tip, Nasal bone,<br>Mandible, Upper chest | LV, Orbit | Fetal position,<br>Image clarity,<br>Image zoom (> 2/3) |
| MSF | GN, NT, IT, SSS, Pygal, Calvaria, Interbrain,<br>Midbrain, Nasal tip,<br>Abdomen | LV, Orbit, Ilium,<br>Thigh | Fetal position,<br>Image clarity,<br>Image zoom (> 2/3) |
| AFTP | CBO, Falx cerebri, CP | / | Image clarity,<br>Image zoom (> 1/3) |
| AFA | UCI, Spine, Abdomen | Bladder | Image clarity,<br>Image zoom (> 1/3) |

The main structures were categorized as good or bad, the mutually exclusive structures as visible or invisible, and the fetal position, image clarity and image zoom as fit or unfit. NTP: nuchal translucency plane, MSF: midsagittal view of the fetus, AFTP: axial view of fetal head in the transventricular plane, AFA: axial view of fetal abdomen, NT: nuchal translucency, IT: intracranial translucency, LV: lateral ventricle, GN: gonadal node, SSS: spinal sagittal section, CBO: cranial bone ossification, CP: choroid plexus, UCI: umbilical cord insertion

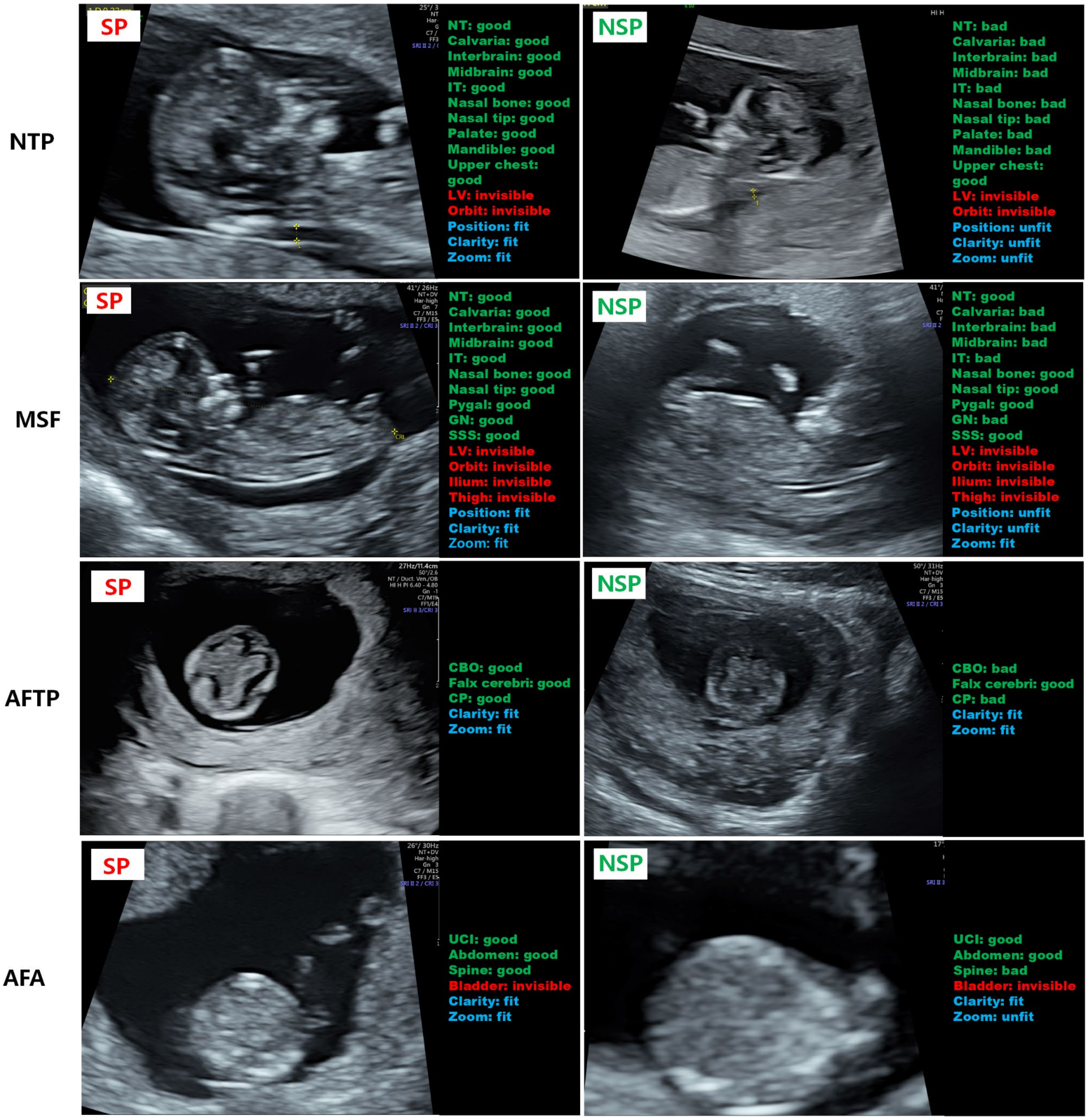


**Fig. 2** Examples of standard and nonstandard images and the reasons manually annotated by radiologists for these categorizations. Green text, main structure assessment; red text, mutually exclusive structure assessment; blue text, overall image assessment. SP: standard plane, NSP: nonstandard plane, NTP: nuchal translucency plane, MSF: midsagittal view of the fetus, AFTP: axial view of fetal head in the transventricular plane, AFA: axial view of fetal abdomen, NT: nuchal translucency, IT: intracranial translucency, LV: lateral ventricle, GN: gonadal node, SSS: spinal sagittal section, CBO: cranial bone ossification, UCI: umbilical cord insertion, CP: choroid plexus

the SGD optimizer (Momentum: 0.937) with Complete Intersection over Union (IoU) and Binary Cross-Entropy (BCE) Loss, whereas ResNet50 employed the Adam optimizer (betas = (0.9, 0.999)) and Mean Squared Error (MSE) Loss. Both networks were trained separately with a batch size of 128 and a learning rate of 1e-3. Early stopping was not employed, as cross-validation indicated no overfitting within 100 epochs.

### Statistical analysis

Statistical analysis was conducted using SPSS software version 22.0 (IBM Corp., Armonk, NY, USA). Accuracy (ACC), precision, recall and F1-score were used as

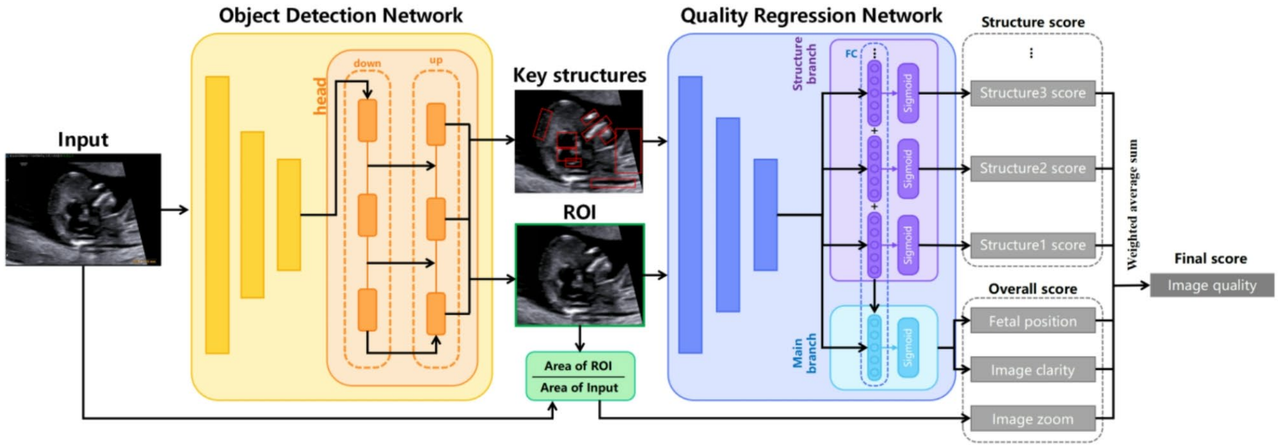


**Fig. 3** Flowchart illustrating the artificial intelligence-based image quality audit process. ROI: region of interest, FC: fully connected layer

metrics to evaluate the performance of the AI-IQA system. These metrics were calculated by comparing the binary qualitative results derived from the model's predicted quality scores with the gold standard. The specific calculation formulas are as follows:

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

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

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

$$F1\ score = \frac{2\times\ \mathrm{Precison}\times\ \mathrm{Recall}}{\mathrm{Precison}+\mathrm{Recall}} \tag{4}$$

Where TP represents the number of samples correctly classified as positive, FP represents the number of samples incorrectly classified as positive, FN represents the number of samples incorrectly classified as negative and TN represents the number of samples correctly classified as negative.

Consistency between AI-IQA system and expert audit results was assessed using the Cohen's Kappa analysis, with the coefficient interpreted as follows: 0.81–1.00 (strong), 0.61–0.80 (moderate to strong), 0.41–0.60 (moderate), 0.21–0.40 (fair), and <0.2 (poor). For comparing the time consumption of AI-IQA and expert audits, we used the Wilcoxon signed-rank test, as the paired differences were symmetrically distributed. McNemar's test was used to evaluate differences between paired binary data (AI-IQA vs. expert assessments), appropriate for categorical data with dependent observations. The chi-square test, which requires expected frequencies greater than 5 in each cell, was employed to evaluate the significance of improvements in the proportion of standard images among radiologists. $P<0.05$ was considered statistically significant.

**Table 3** Characteristics of pregnant women undergoing routine prenatal screening in the first-trimester

| **Characteristic** | **Clinical test set 1 (*n* = 349)** | **Clinical test set 2 (*n* = 218)** |
|---|---|---|
| Mean age (years) | 29.92 | 29.96 |
| Mean BMI (kg/m$^2$) | 23.9 | 24.1 |
| GA (weeks) | 12W5d ± 4d | 12W4d ± 3d |
| **Machine used (n [%])** | | |
| GE Voluson E8 | 45 (12.9%) | 29 (13.3%) |
| GE Voluson E10 | 174 (49.9%) | 98 (45.0%) |
| Samsung WS 80 A | 130 (37.2%) | 91 (41.7%) |

No significant differences were observed in maternal age, maternal BMI and gestational age between clinical test sets 1 and 2 ($p>0.05$). GA is presented with the mean ± SD format. BMI: body mass index, GA: gestational age

## Results

### Characteristics of the study

The mean age and body mass index of pregnant women and the GA of the fetuses are summarized in Table 3. There were no significant differences ($p>0.05$) in these characteristics between clinical test sets 1 and 2. The US devices used by the four radiologists in the clinical validation, including the GE Voluson E8, GE Voluson E10 and Samsung WS 80 A, are also listed in Table 3.

### Performance of the AI-IQA system in the internal test set

Using the internal test set, we assessed the capacity of the AI-IQA system to appraise the holistic quality of plane images and the integrity of the key structures.

The detection model achieved a precision of 0.83, recall of 0.85 and mean Average Precision 50 (mAP50) of 0.85 across the four types of standard plane images. The auditing results of the AI-IQA system, including ACC precision, recall and F1-score, consistently exceeded 0.8, as summarized in Table 4. Notably, for the appraisal of image quality, AI exhibited a remarkable average ACC of 0.881. The individual ACCs for NTP, MSF, AFTP and AFA were 0.879, 0.857, 0.914 and 0.875, respectively.

**Table 4** Performance of the artificial intelligence-based image quality audit system in the internal test set

| Plane | Image quality | | | | Structure quality | | | |
|---|---|---|---|---|---|---|---|---|
| | ACC | Precision | Recall | F1-score | ACC | Precision | Recall | F1-score |
| NTP | 0.879 | 0.870 | 0.847 | 0.859 | 0.905 | 0.869 | 0.825 | 0.848 |
| MSF | 0.857 | 0.826 | 0.793 | 0.808 | 0.886 | 0.863 | 0.826 | 0.844 |
| AFTP | 0.914 | 0.839 | 0.892 | 0.864 | 0.936 | 0.894 | 0.910 | 0.902 |
| AFA | 0.875 | 0.796 | 0.834 | 0.815 | 0.895 | 0.817 | 0.866 | 0.840 |
| Average | 0.881 | 0.833 | 0.842 | 0.837 | 0.906 | 0.861 | 0.857 | 0.859 |

ACC: accuracy, NTP: nuchal translucency plane, MSF: midsagittal view of the fetus, AFTP: axial view of fetal head in the transventricular plane, AFA: axial view of fetal abdomen

These results suggest that the AI-IQA system demonstrates effectiveness and reliability in evaluating both the image quality and the structural integrity of major components during internal testing, reflecting solid performance.

To better visualize AI-IQA predicted scores and expert annotation results, examples of four plane images are shown in Fig. 4. In the table on the right side of Fig. 4, the second column is the AI-IQA predicted scores and the third column is the expert annotated results which served as gold standards.

### Performance of the AI-IQA system in clinical test sets

For a more in-depth analysis of the performance of the AI-IQA system in the clinical test sets, two experienced experts independently assessed the image quality based on the same criteria. In cases of disagreement, a consensus was reached following discussion. Table 5 shows the number of clinical test sets images classified as standard and nonstandard by the AI-IQA system and experts. The AI-IQA identified 93.6% (596/637), 85.2% (564/662), 88.6% (553/624) and 75.7% (485/641) of NTP, MSF, AFTP and AFA plane images, respectively, as standard; for the experts, these values were 94.2%, 86.4%, 89.7% and 76.9%, respectively. There was therefore consistency between the AI-IQA and expert results, with the Cohen's Kappa coefficient exceeding 0.8 for all four planes.

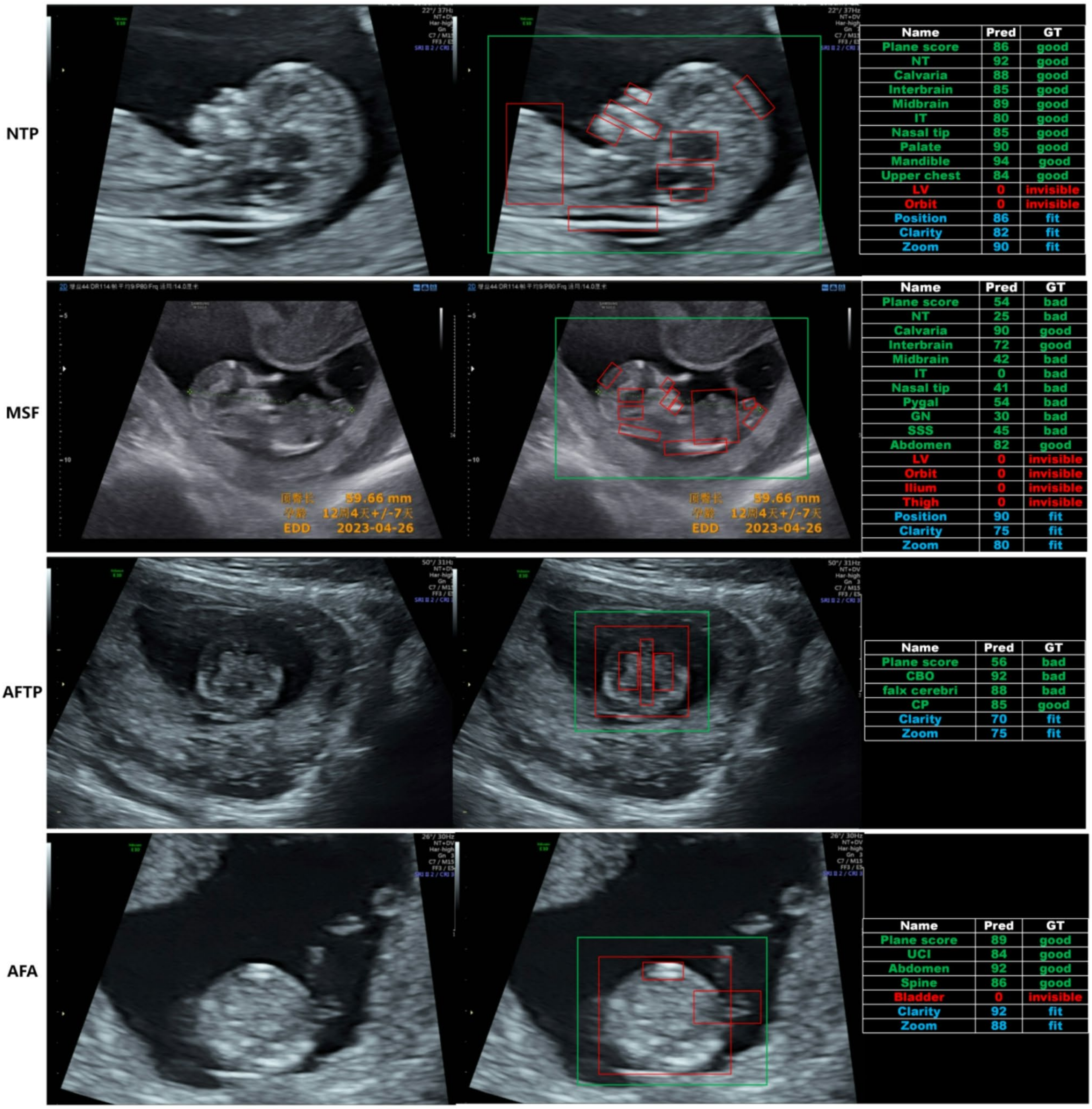
Table 6 summarizes the average time taken by the AI-IQA system and experts to assess the quality of each clinical test set image. The AI-IQA system demonstrated an average evaluation time of 0.05 s per image, over 100 times faster than the experts (9.8–36.7 s). The difference between the two was statistically significant ($p < 0.001$), as determined by the Wilcoxon signed-rank test.

For a more systematic validation of whether AI-IQA results comply with clinical practice standards, we used expert audit results as the gold standard and calculated the ACC of AI-IQA in the four planes, as shown in Table 7. In clinical test set 1, the ACCs in the four planes were 0.884, 0.841, 0.896 and 0.825, with an average ACC of 0.862. In clinical test set 2, the ACCs were 0.932, 0.964, 0.881 and 0.847, with an average ACC of 0.906. Additionally, we separately compared the percentage of standard images between the two radiologist groups and the AI-IQA and expert audit results. All the results showed close proximity, with McNemar's test indicating no statistical difference between the AI-IQA and expert results ($p > 0.05$).

### Image quality before and after AI-IQA feedback

To assess the effect of the AI-IQA feedback on image quality, we compared the image quality for the two groups of radiologists before and after receiving AI-IQA feedback. In each feedback round, US images acquired by the radiologists within a recent period were collected and audited by the AI-IQA system. Feedback was delivered as a visual report with quality scores for each item, guiding radiologists on improvements. Expert assessment results were used as gold standards for comparison. As presented in Table 8, both junior and mid-level radiologists exhibited an improvement in the quality of obtained images after AI-IQA feedback training. Specifically, the proportion of standard images obtained by junior radiologists increased from 80.8% to 88.5% and that obtained by mid-level radiologists increased from 86.7% to 91.8%, improvements of 7.7% and 5.1%, respectively. The chi-square test indicated that the differences in pass rates were statistically significant ($P < 0.05$). Analysis of the various imaging planes demonstrated that both groups achieved a standard image rate exceeding 90% on NTP and MSF post-feedback images. AFA image quality was initially subpar, particularly among junior radiologists, who achieved a standard image rate of 62.5%. However, this value improved to 76.6% after AI-IQA feedback.

To further analyze the performance of radiologists, the error proportions of different attributes in all nonstandard images were calculated, as shown in Fig. 5. Both junior and mid-level radiologists made the most errors in the representation of the main structures, with error rates of 58.76% and 52.64%, respectively. This was followed by errors in overall image evaluation, with rates of 35.33% and 38.53%, respectively. Over 90% of nonstandard NTP and MSF images did not clearly display the brain structures. Additionally, the fetal position and display of NT error rates, which can significantly affect CRL and NT measurements, exceeded 70% in both groups.



| Name | Pred | GT |
|---|---|---|
| Plane score | 86 | good |
| NT | 92 | good |
| Calvaria | 88 | good |
| Interbrain | 85 | good |
| Midbrain | 89 | good |
| IT | 80 | good |
| Nasal tip | 85 | good |
| Palate | 90 | good |
| Mandible | 94 | good |
| Upper chest | 84 | good |
| LV | 0 | invisible |
| Orbit | 0 | invisible |
| Position | 86 | fit |
| Clarity | 82 | fit |
| Zoom | 90 | fit |

| Name | Pred | GT |
|---|---|---|
| Plane score | 54 | bad |
| NT | 25 | bad |
| Calvaria | 90 | good |
| Interbrain | 72 | good |
| Midbrain | 42 | bad |
| IT | 0 | bad |
| Nasal tip | 41 | bad |
| Pygal | 54 | bad |
| GN | 30 | bad |
| SSS | 45 | bad |
| Abdomen | 82 | good |
| LV | 0 | invisible |
| Orbit | 0 | invisible |
| Ilium | 0 | invisible |
| Thigh | 0 | invisible |
| Position | 90 | fit |
| Clarity | 75 | fit |
| Zoom | 80 | fit |

| Name | Pred | GT |
|---|---|---|
| Plane score | 56 | bad |
| CBO | 92 | bad |
| falx cerebri | 88 | bad |
| CP | 85 | good |
| Clarity | 70 | fit |
| Zoom | 75 | fit |

| Name | Pred | GT |
|---|---|---|
| Plane score | 89 | good |
| UCI | 84 | good |
| Abdomen | 92 | good |
| Spine | 86 | good |
| Bladder | 0 | invisible |
| Clarity | 92 | fit |
| Zoom | 88 | fit |

**Fig. 4** Comparison of artificial intelligence-based image quality audit predicted scores with expert annotated results. The green rectangles represent the region of interest for the plane. The red rectangles delineate the anatomical structures appearing in the plane image, including both main and mutually exclusive structures. NTP: nuchal translucency plane, MSF: midsagittal view of the fetus, AFTP: axial view of fetal head in the transventricular plane, AFA: axial view of fetal abdomen, NT: nuchal translucency, IT: intracranial translucency, LV: lateral ventricle, GN: gonadal node, SSS: spinal sagittal section, CBO: cranial bone ossification, UCI: umbilical cord insertion, CP: choroid plexus

Falx cerebri and choroid plexus were the two structures most likely to be imaged incorrectly in AFTP images, whereas in the worst-performing AFA plane, the lack of clear display of the spine and umbilical cord insertion (UCI) were the two main reasons for images being classified as nonstandard.

## Discussion

### Principal findings

Many previous studies have emphasized the importance of quality control in fetal ultrasonography, with regular auditing of images and feedback on identified issues contributing to improvements in image quality. However, manual assessments are associated with low

**Table 5** Comparison of clinical test set images analyzed by the AI-IQA system and experts

| Plane | AI-IQA | | Experts | | Cohen' Kappa [95%CI] |
|---|---|---|---|---|---|
| | Standard | Nonstandard | Standard | Nonstandard | |
| NTP (*n* = 637) | 596 (93.6) | 41 (6.4) | 600 (94.2) | 37 (5.8) | 0.892 [0.829,0.955] |
| MSF (*n* = 662) | 564 (85.2) | 98 (14.8) | 572 (86.4) | 90(13.6) | 0.879 [0.800,0.957] |
| AFTP (*n* = 624) | 553 (88.6) | 71 (11.4) | 560 (89.7) | 64(10.3) | 0.850 [0.783,0.917] |
| AFA (*n* = 641) | 485 (75.7) | 156 (24.3) | 493 (76.9) | 148 (23.1) | 0.857 [0.761,0.953] |

Data are expressed as number (percentage). AI-IQA: artificial intelligence-based image quality audit, NTP: nuchal translucency plane, MSF: midsagittal view of the fetus, AFTP: axial view of fetal head in the transventricular plane, AFA: axial view of fetal abdomen

**Table 6** A summary of the time-consuming taken to assess each image

| Plane | Clinical test set 1 | | Clinical test set 2 | |
|---|---|---|---|---|
| | AI-IQA | Experts | AI-IQA | Experts |
| NTP | 0.05 | 36.7 | 0.05 | 32.8 |
| MSF | 0.05 | 35.4 | 0.05 | 33.9 |
| AFTP | 0.05 | 12.5 | 0.05 | 11.3 |
| AFA | 0.05 | 10.3 | 0.05 | 9.8 |
| Average | 0.05 | 23.7 | 0.05 | 22.0 |

Data are expressed in s. The time required for image assessment differed significantly between the system and the experts ($p < 0.001$). AI-IQA: artificial intelligence-based image quality audit, NTP: nuchal translucency plane, MSF: midsagittal view of the fetus, AFTP: axial view of fetal head in the transventricular plane, AFA: axial view of fetal abdomen

**Table 7** Accuracy of the AI-IQA system and a comparison of junior and mid-level radiologists, with tests indicating consistency between AI-IQA and expert results ($p > 0.05$)

| Plane | AI-IQA vs. experts | Junior vs. experts | Junior vs. AI-IQA | Mid-level vs. experts | Mid-level vs. AI-IQA |
|---|---|---|---|---|---|
| **Clinical test set 1** | | | | | |
| NTP | 0.884 | 0.914 | 0.879 | 0.927 | 0.952 |
| MSF | 0.841 | 0.808 | 0.788 | 0.817 | 0.809 |
| AFTP | 0.896 | 0.887 | 0.866 | 0.899 | 0.887 |
| AFA | 0.825 | 0.625 | 0.597 | 0.824 | 0.815 |
| Average | 0.862 | 0.808 | 0.783 | 0.867 | 0.891 |
| **Clinical test set 2** | | | | | |
| NTP | 0.932 | 0.983 | 0.974 | 0.963 | 0.933 |
| MSF | 0.964 | 0.940 | 0.932 | 0.939 | 0.926 |
| AFTP | 0.881 | 0.874 | 0.882 | 0.928 | 0.912 |
| AFA | 0.847 | 0.766 | 0.744 | 0.835 | 0.843 |
| Average | 0.906 | 0.885 | 0.883 | 0.918 | 0.904 |

AI-IQA: artificial intelligence-based image quality audit, NTP: nuchal translucency plane, MSF: midsagittal view of the fetus, AFTP: axial view of fetal head in the transventricular plane, AFA: axial view of fetal abdomen

**Table 8** Proportion of standard images obtained by junior and mid-level radiologists

| Plane | Junior group | | Mid-level group | |
|---|---|---|---|---|
| | Pre-feedback | Post-feedback | Pre-feedback | Post-feedback |
| NTP | 128/140 (91.4%) | 113/115 (98.3%) | 230/248 (92.7%) | 129/134 (96.3%) |
| MSF | 118/146 (80.8%) | 110/117 (94.0%) | 205/251 (81.7%) | 139/148 (93.9%) |
| AFTP | 126/142 (88.7%) | 104/119 (87.4%) | 214/238 (89.9%) | 116/125 (92.8%) |
| AFA | 90/144 (62.5%) | 105/137 (76.6%) | 192/233 (82.4%) | 106/127 (83.5%) |
| Average | 462/572 (80.8%) | 432/488 (88.5%) | 841/970 (86.7%) | 490/534 (91.8%) |

Data are expressed as number/total (percentage). NTP: nuchal translucency plane, MSF: midsagittal view of the fetus, AFTP: axial view of fetal head in the transventricular plane, AFA: axial view of fetal abdomen

consistency and inefficiency, limiting their potential to significantly enhance examination quality. To overcome the deficiencies of manual auditing, we developed an AI model, AI-IQA, which utilizes plane structure detection and a quality regression network to intelligently audit the image quality of the four key planes of FTS.

Recent advancements in DL have demonstrated its potential in fetal US imaging. Chen et al. pioneered the use of CNNs to locate the fetal abdominal standard plane (FASP) in US videos [30], while Zhang et al. focused on recognizing key anatomical structures in the fetal abdomen, head and heart [31]. Other studies have also achieved real-time detection in 2D and 3D US videos [32, 33]. However, these approaches either focus on mid-to-late pregnancy images or are limited to a single standard plane in early pregnancy, lacking comprehensive quality assessments. Zhen et al.'s work, the most comparable to ours, developing a quality control system for early pregnancy images based on expert scoring [34]. It is important to note that, unlike these studies primarily focused on assisting radiologists in detecting standard planes during US screening, our objective is to evaluate the quality of acquired images. This represents a subtle difference from other tasks and research in this area remains limited. Our work aims to provide a quality auditing tool, offering feedback to assist radiologists in improving image quality.

To validate performance, we verified the AI-IQA system in an internal test set and two clinical test sets, with results indicating that the AI-IQA system meets clinical practice standards and is consistent with expert audits. Following AI-IQA feedback, both junior and mid-level radiologists exhibited substantial improvements in obtained image quality.

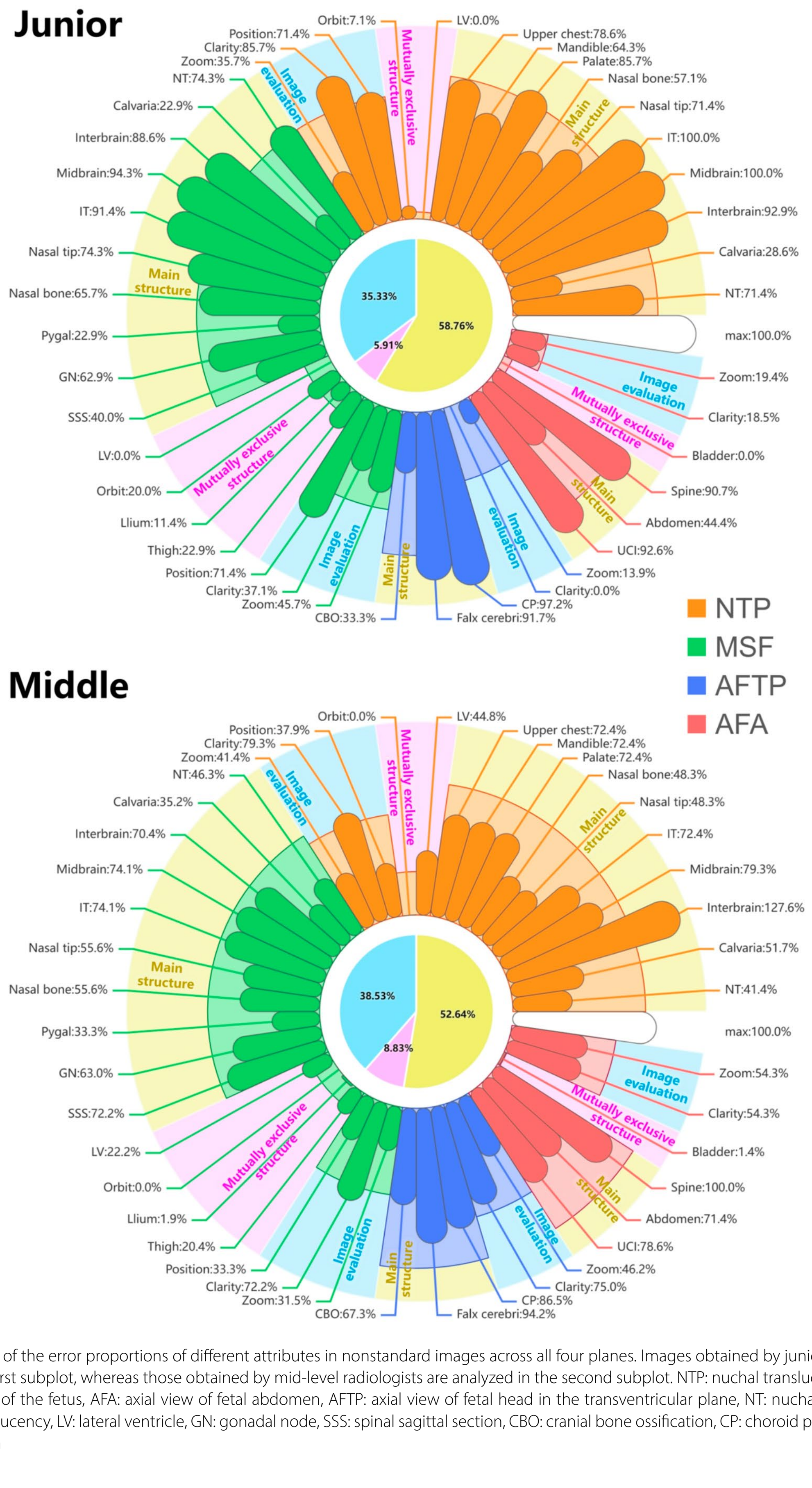


**Fig. 5** Summary of the error proportions of different attributes in nonstandard images across all four planes. Images obtained by junior radiologists are analyzed in the first subplot, whereas those obtained by mid-level radiologists are analyzed in the second subplot. NTP: nuchal translucency plane, MSF: midsagittal view of the fetus, AFA: axial view of fetal abdomen, AFTP: axial view of fetal head in the transventricular plane, NT: nuchal translucency, IT: intracranial translucency, LV: lateral ventricle, GN: gonadal node, SSS: spinal sagittal section, CBO: cranial bone ossification, CP: choroid plexus, UCI: umbilical cord insertion

### Clinical implications

Further investigations have showed that comprehensive image quality audits can improve the completeness and quality of US scans conducted by radiologists [35]. However, manual image auditing is a labor-intensive task and requires experienced experts to ensure the effectiveness of the assessment results. Therefore, hospitals can only perform a limited number of selective audits, which may not fully assess the performance of radiologists or provide timely feedback. In less developed regions, resource shortages may further limit the performance of effective audits, exacerbating healthcare quality disparities.

In theory, an AI-based model can serve as a comprehensive and convenient approach for US image quality control to tackle the above issues. Compared to manual assessments, the AI-IQA system provides a uniform and objective evaluation, eliminating intra-observer variability and addressing human resource disparities between hospitals. The AI-IQA system is also 100 times quicker than manual assessment, significantly enhancing efficiency. Without the constraints of observer experience and time, the utilization of AI-IQA should enable comprehensive quality audits of all cases, thereby substantially elevating US screening in hospitals.

### Research implications

Standard US plane images are fundamental for assessing anatomical structures and performing biometric measurements in maternal-fetal screening, enabling accurate diagnoses that guide clinical decision-making and pregnancy care. Therefore, this study explored the value of AI-IQA feedback in improving image quality among radiologists with different levels of experience.

NT thickness and CRL are two key biometric measurements of the fetus performed in the first trimester using NTP and MSF images, respectively. These measurements are crucial for chromosomal abnormality screening and calculating GA. In this study, we found that nonstandard images in these planes were mostly due to radiologists failing to position the fetus in the neutral position and incorrectly believing that showing only a small portion of NT fulfilled the requirement. However, as positional flexion or extension can alter the distance from the fetal head to the buttock, NT measurement requires a clear view of the entire cross section of the fetal neck, including the skin boundaries on both sides of the neck and head.

These issues may reflect a lack of operational experience or thorough understanding of standard planes. Even more experienced mid-level radiologists can sometimes overlook these critical details. In the AFA plane, both junior and mid-level radiologists failed to obtain a clear view of the UCI and spine, resulting in a lower standard rate of AFA images. These findings reveal some common issues that affect the quality of hospital US examinations. Therefore, it is essential that junior and mid-level radiologists receive continuous education and training to enhance the overall quality of US examinations.

The AI-IQA system objectively and clearly identified these issues and the quality of NTP, MSF and AFA images obtained by both junior and mid-level radiologists improved significantly following AI-IQA feedback. The proportion of standard images obtained by junior radiologists after AI-IQA feedback approached or even surpassed that of mid-level radiologists before feedback, suggesting that the AI-IQA system could shorten the hospital training cycle. Especially in the AFA plane, there is a significant difference between the junior and mid-level group in pre-feedback standard rate, indicating that the scanning may be challenging and experience may be especially important for this plane. AI-IQA assists radiologists in reducing the accumulation period of experience, enabling junior radiologists to approach the level of senior radiologists in a short period of time. The intervention of the AI system deepened the understanding of standard planes among radiologists, corrected previous misconceptions and provided continuous education. This study indicates that AI-IQA not only performs accurate quality audits but also has value in improving the quality of images obtained by radiologists with varying levels of expertise.

### Strengths and limitations

This investigation represents a pioneering endeavor to introduce an AI-IQA system meticulously tailored to fetal plane images during FTS. Our work has substantiated the viability of this methodology, which may streamline intelligent oversight of quality control protocols in obstetric imaging. It is important to note that, while the AI-IQA system showed a good level of agreement with expert evaluations, it is better suited as a tool for audits and learning rather than as a direct replacement for human clinical assessment.

This study has several limitations. There is no evidence in this study suggesting that the improvement in image quality depends more on the feedback from AI-IQA than on the accumulation of radiologists' own experience and skill enhancement over time. However, since our mid-level radiologists have 5–10 years of clinical experience, it is unlikely that the significant improvement in their image quality can be solely attributed to a few months of additional experience. Additionally, the study was limited to fetal images with normative anatomical structures, which may restrict the generalizability of the AI-IQA system to more complex populations and real-world clinical settings, where a broader and more heterogeneous patient demographic is typically encountered. Finally, this single-center validation involved a relatively homogeneous sample obtained from only three US machines.

The performance of the AI-IQA system may vary with different US equipment, as each machine has unique imaging characteristics and signal processing methods. Future work will involve multi-center validation to more comprehensively assess the performance of the AI-IQA system. We also plan to address sample size and imbalance issues using generative methods and to include more heterogeneous samples to enhance the model's generalizability [36]. Furthermore, we aim to conduct a large-scale controlled study to evaluate the impact of the system on improving hospital training cycles.

## Conclusion

In this study, we developed an AI-IQA system for automatically auditing image quality during FTS, which showed good consistency with expert assessments. Our results indicate that Al-QA is a useful tool for conducting comprehensive quality audits and has the potential to support radiologists with varying levels of expertise in improving image quality, contributing to maternal and fetal health screening. In clinical practice, the AI-IQA system could significantly improve efficiency and optimize resource utilization, particularly in resource-limited settings. Further multi-center validation and the development of supporting software will be essential to facilitate the clinical adoption and integration of this technology.

### Abbreviations

| | |
|---|---|
| ACC | Accuracy |
| AFA | Axial view of fetal abdomen |
| AFTP | Axial view of fetal head in the transventricular plane |
| AI | Artificial intelligence |
| AI-IQA | Artificial intelligence-based image quality audit |
| CBO | Cranial bone ossification |
| CRL | Crown-rump length |
| DL | Deep learning |
| GA | Gestational age |
| GN | Gonadal node |
| ISUOG | International Society of Ultrasound in Obstetrics and Gynecology |
| IT | Intracranial transparency |
| LV | Lateral ventricle |
| MSF | Midsagittal view of the fetus |
| NT | Nuchal translucency |
| NTP | Nuchal translucency plane |
| ROI | Region of interest |
| SSS | Spinal sagittal section |
| UCI | Umbilical cord insertion |
| US | Ultrasound |
| IoU | Intersection over union |
| BCE | Binary cross-entropy |
| MSE | Mean squared error |
| mAP50 | Mean average precision 50 |
| CNN | Convolutional neural network |

## Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12884-025-07485-4.

Supplementary Material 1: Detailed Description of Development of the AI-IQA System: This supplementary file provides a detailed methodology of the AI-IQA system, including the workflow of the YOLOv7 object detection network and the ResNet50-based quality regression network, along with data processing details. The document supports the model development and experimental setup for image quality assessment in the study

### Acknowledgements

We thank Editage for providing professional English language editing.

### Author contributions

XYC, XY: Conceptualization; XYC, YC, SKZ, YLY, HLL, TW: Data collection, Annotations, Quality control; XYC, BHL, YSZ, XY, YC, XDH, CYC: Methods, Results analysis; XYC, YC, BHL, YSZ: Manuscript writing; XYC, XY, TT, XDH, CYC: Funding acquisition, Project administration; XYC, XY, BHL, YSZ, LW, DN: Supervision; XYC, XY, BHL, YSZ, YC, XDH, CYC, SKZ, HLL, TW, YLY, TT, LW, DN: Manuscript revision. All authors read and approved the final manuscript.

### Funding

This work was supported by the Research Fund of the Shenzhen Health Economics Association (2023149), Futian District Health and Health System Research Fund (FTWS2023013), Science and Technology Planning Project of Guangdong Province (2023A0505020002), Science and Technology Development Fund of Macao (0021/2022/AGJ) and National Natural Science Foundation of China (Nos.62101343, 62171290).

### Data availability

The datasets and codes are not publicly available because of hospital policy and privacy considerations but are available from the corresponding author upon reasonable request.

## Declarations

### Ethics approval and consent to participate

The study was approved by the Ethics Committee of Shenzhen Futian District Maternity & Child Healthcare Hospital (protocol number: K-2023-04-01). Informed consent was obtained from all pregnant women. This study was conducted in accordance with the principles outlined in the Declaration of Helsinki.

### Consent for publication

Not applicable.

### Competing interests

The authors declare no competing interests.

Received: 12 December 2024 / Accepted: 17 March 2025
Published online: 31 March 2025

### References

1. Zaffino P, Moccia S, De Momi E, Spadea MF. A review on advances in Intra-operative imaging for surgery and therapy: imagining the operating room of the future. Ann Biomed Eng. 2020;48:2171–91.
2. Salomon LJ, Alfirevic Z, Berghella V, Bilardo CM, Chalouhi GE, Da Silva Costa F, et al. ISUOG practice guidelines (updated): performance of the routine mid-trimester fetal ultrasound scan. Ultrasound Obstet Gynecol. 2022;59:840–56.
3. International Society of Ultrasound in Obstetrics and Gynecology, Bilardo CM, Chaoui R, Hyett JA, Kagan KO, Karim JN, et al. ISUOG practice guidelines (updated): performance of 11–14-week ultrasound scan. Ultrasound Obstet Gynecol. 2023;61:127–43.
4. Liao Y, Wen H, Ouyang S, Yuan Y, Bi J, Guan Y, et al. Routine first-trimester ultrasound screening using a standardized anatomical protocol. Am J Obstet Gynecol. 2021;224:e3961–39615.
5. Chinese Society of Ultrasound in Medicine Obstetric Ultrasound Group, National Health Commission Maternal and Child Health Division National Expert Group on Prenatal Diagnosis Medical Imaging Group. Guidelines for prenatal ultrasound screening. Chin J Ultrasonography. 2022;31:1–12.
6. Sun Y, Zhang L, Dong D, Li X, Wang J, Yin C, et al. Application of an individualized nomogram in first-trimester screening for trisomy 21. Ultrasound Obstet Gynecol. 2021;58:56–66.

7. Charasson T, Ko-Kivok-Yun P, Martin F, Sarramon MF. Screening for trisomy 21 by measuring nuchal translucency during the first trimester of pregnancy. J Gynecol Obstet Biol Reprod (Paris). 1997;26:671–8.
8. Chen M, Xue S, Chen J, Chen D, Liu Y, Yan H. OP05.09: correlation between increased nuchal translucency and chromosomal abnormalities. Ultrasound Obstet Gynecol. 2019;54:101–101.
9. Almeida A, Moura C v., Alves Tm, Braga J, Martins LG, Cunha A. VP28.05: predicting chromosomal abnormalities through first trimester screening and nuchal translucency: experience of a tertiary centre. Ultrasound Obstet Gynecol. 2021;58:213–213.
10. Napolitano R, Dhami J, Ohuma E, Ioannou C, Conde-Agudelo A, Kennedy S, et al. Pregnancy dating by fetal crown–rump length: a systematic review of charts. BJOG: Int J Obstet Gynecol. 2014;121:556–65.
11. Chaoui R, Orosz G, Heling KS, Sarut-Lopez A, Nicolaides KH. Maxillary gap at 11–13 weeks' gestation: marker of cleft lip and palate. Ultrasound Obstet Gynecol. 2015;46:665–9.
12. Verla MA, Style CC, Olutoye OO. Prenatal diagnosis and management of omphalocele. Semin Pediatr Surg. 2019;28:84–8.
13. Volpe N, Dall'Asta A, Di Pasquo E, Frusca T, Ghi T. First-trimester fetal neurosonography: technique and diagnostic potential. Ultrasound Obstet Gynecol. 2021;57:204–14.
14. Zhang N, Dong H, Wang P, Wang Z, Wang Y, Guo Z. The value of obstetric ultrasound in screening fetal nervous system malformation. World Neurosurg. 2020;138:645–53.
15. Salomon L-J, Bernard J-P, Ville Y. Quality control of prenatal ultrasound. A role for biometry. Gynecol Obstet Fertil. 2006;34:683–91.
16. Chinese Physicians Association Ultrasound Physicians Branch. Expert consensus on standardized training and assessment criteria for obstetric ultrasound (2022 edition). Chin J Ultrasonography. 2022;31:369–78.
17. Wu L, Cheng J-Z, Li S, Lei B, Wang T, Ni D. FUIQA: fetal ultrasound image quality assessment with deep convolutional networks. IEEE Trans Cybern. 2017;47:1336–49.
18. He S, Lin Z, Yang X, Chen C, Wang J, Shuang X et al. Statistical Dependency Guided Contrastive Learning for Multiple Labeling in Prenatal Ultrasound. 2022.
19. Burgos-Artizzu XP, Coronado-Gutiérrez D, Valenzuela-Alcaraz B, Bonet-Carne E, Eixarch E, Crispi F, et al. Evaluation of deep convolutional neural networks for automatic classification of common maternal fetal ultrasound planes. Sci Rep. 2020;10:10200.
20. Fu Z, Jiao J, Yasrab R, Drukker L, Papageorghiou AT, Noble JA. Anatomy-Aware contrastive representation learning for fetal ultrasound. Comput Vis ECCV. 2022;2022:422–36.
21. Migliorelli G, Fiorentino MC, Di Cosmo M, Villani FP, Mancini A, Moccia S. On the use of contrastive learning for standard-plane classification in fetal ultrasound imaging. Comput Biol Med. 2024;174:108430.
22. Zhao H, Zheng Q, Teng C, Yasrab R, Drukker L, Papageorghiou AT, et al. Memory-based unsupervised video clinical quality assessment with multimodality data in fetal ultrasound. Med Image Anal. 2023;90:102977.
23. Qu R, Xu G, Ding C, Jia W, Sun M. Standard plane identification in fetal brain ultrasound scans using a differential convolutional neural network. IEEE Access. 2020;8:83821–30.
24. Lin Z, Li S, Ni D, Liao Y, Wen H, Du J, et al. Multi-task learning for quality assessment of fetal head ultrasound images. Med Image Anal. 2019;58:101548.
25. Fiorentino MC. A review on deep-learning algorithms for fetal ultrasound-image analysis. Med Image Anal. 2023;83:102629.
26. Dong J, Liu S, Liao Y, Wen H, Lei B, Li S, et al. A generic quality control framework for fetal ultrasound cardiac Four-Chamber planes. IEEE J Biomedical Health Inf. 2020;24:931–42.
27. Liang J, Yang X, Huang Y, Li H, He S, Hu X, et al. Sketch guided and progressive growing GAN for realistic and editable ultrasound image synthesis. Med Image Anal. 2022;79:102461.
28. Wang C-Y, Bochkovskiy A, Liao H-YM. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2023:7464–75.
29. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016:770–8.
30. Chen H, Ni D, Qin J, Li S, Yang X, Wang T, et al. Standard plane localization in fetal ultrasound via domain transferred deep neural networks. IEEE J Biomed Health Inf. 2015;19:1627–36.
31. Zhang B, Liu H, Luo H, Li K. Automatic quality assessment for 2D fetal sonographic standard plane based on multitask learning. Medicine. 2021;100:e24427.
32. Baumgartner CF, Kamnitsas K, Matthew J, Fletcher TP, Smith S, Koch LM, et al. SonoNet: Real-Time detection and localisation of fetal standard scan planes in freehand ultrasound. IEEE Trans Med Imaging. 2017;36:2204–15.
33. Ramirez Zegarra R, Ghi T. Use of artificial intelligence and deep learning in fetal ultrasound imaging. Ultrasound Obstet Gynecol. 2023;62:185–94.
34. Zhen C, Wang H, Cheng J, Yang X, Chen C, Hu X, et al. Locating multiple standard planes in First-Trimester ultrasound videos via the detection and scoring of key anatomical structures. Ultrasound Med Biol. 2023;49:2006–16.
35. Yaqub M, Kelly B, Stobart H, Napolitano R, Noble JA, Papageorghiou AT. Quality-improvement program for ultrasound-based fetal anatomy screening using large-scale clinical audit. Ultrasound Obstet Gynecol. 2019;54:239–45.
36. Lasala A, Fiorentino MC, Bandini A, Moccia S, FetalBrainAwareNet. Bridging GANs with anatomical insight for fetal ultrasound brain plane synthesis. Comput Med Imaging Graph. 2024;116:102405.

## Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.