Radiology: Artificial Intelligence

ORIGINAL RESEARCH

Deep Learning Model for Real-Time Nuchal Translucency Assessment at Prenatal US

Yuanji Zhang, MS[1,2,3] • *Xin Yang, PhD*[1,2,3] • *Chunya Ji, MS*[4] • *Xindi Hu, MS*[5] • *Yan Cao, MS*[5] • *Chaoyu Chen, MS*[1,2,3] • *He Sui, BS*[6] • *Binghan Li, BS*[1,2,3] • *Chaojiong Zhen, MD*[7] • *Weijun Huang, MD*[7] • *Xuedong Deng, MD*[4] • *Linliang Yin, MD*[4] • *Dong Ni, PhD*[1,2,3]

Author affiliations, funding, and conflicts of interest are listed at the end of this article.

See also commentary by Horii in this issue.

Radiology: Artificial Intelligence 2025; 7(4):e240498 • https://doi.org/10.1148/ryai.240498 • Content codes: AI US OB

Purpose: To develop and evaluate an artificial intelligence–based model for real-time nuchal translucency (NT) plane identification and measurement in prenatal US assessments.

Materials and Methods: In this retrospective multicenter study conducted from January 2022 to October 2023, the Automated Identification and Measurement of NT (AIM-NT) model was developed and evaluated using internal and external datasets. NT plane assessment, including identification of the NT plane and measurement of NT thickness, was independently conducted by AIM-NT and experienced radiologists, with the results subsequently audited by radiology specialists and accuracy compared between groups. To assess alignment of artificial intelligence with radiologist workflow, discrepancies between the AIM-NT model and radiologists in NT plane identification time and thickness measurements were evaluated.

Results: The internal dataset included a total of 3959 NT images from 3153 fetuses, and the external dataset included 267 US videos from 267 fetuses. On the internal testing dataset, AIM-NT achieved an area under the receiver operating characteristic curve of 0.92 for NT plane identification. On the external testing dataset, there was no evidence of differences between AIM-NT and radiologists in NT plane identification accuracy (88.8% vs 87.6%, P = .69) or NT thickness measurements on standard and nonstandard NT planes (P = .29 and .59). AIM-NT demonstrated high consistency with radiologists in NT plane identification time, with 1-minute discrepancies observed in 77.9% of cases, and NT thickness measurements, with a mean difference of 0.03 mm and mean absolute error of 0.22 mm (95% CI: 0.19, 0.25).

Conclusion: AIM-NT demonstrated high accuracy in identifying the NT plane and measuring NT thickness on prenatal US images, showing minimal discrepancies with radiologist workflow.

Supplemental material is available for this article.

US imaging is performed during the first trimester of pregnancy to detect and measure the thickness of fetal nuchal translucency (NT) (1–4). Increased NT thickness can indicate up to 80% of trisomy 21 (5–7) cases, cardiac defects (8–10), and genetic syndromes (11,12). NT thickness typically ranges from 0 to 3 mm, and slight deviations can substantially affect diagnosis (13–15).

Identifying the standard NT plane reliably is challenging due to substantial variation in NT plane assessment, requiring highly skilled observers, extensive training, and rigorous quality control (11,16–20). Although traditional NT assessment protocols have been established to ensure accurate and standardized techniques for the NT plane (11), they are limited by subjective, labor-intensive, and non–real-time analysis (21–23), which can adversely affect diagnostic accuracy.

Fetal NT thickness is closely associated with the risk of chromosomal defects and other abnormalities, necessitating adherence to a standardized measurement technique to ensure uniformity across operators (24). Previous studies have explored both intra- and interoperator variability, demonstrating that quality control measures can effectively reduce measurement discrepancies (25). However, proper training remains critical for sonographers to accurately obtain the correct midsagittal section, adjust the calipers, and measure the maximum NT—tasks that are time-consuming and require experienced radiologists.

Artificial intelligence (AI), particularly deep learning, has demonstrated remarkable accuracy in medical tasks (26–28). When standardized criteria for US image quality are applied using established protocols, it becomes feasible for deep learning to outperform human analysis in certain cases. The integration of deep learning into routine NT US examinations could substantially enhance prenatal screening, particularly for NT plane identification and thickness measurement. However, the lack of detailed evaluations of AI performance and repeatability, especially in alignment with senior radiologists, has reduced clinicians' confidence, thus limiting broader clinical adoption of AI (29).

To address these challenges, we developed an AI-based model called the Automated Identification and Measurement of NT (AIM-NT). This model was designed to provide standardized NT images that adhere to standards and offer real-time prenatal evaluation. The primary objective of this study was to evaluate the AIM-NT model in real-time prenatal NT assessments, focusing on accuracy and alignment with radiologists' workflow patterns.

Abbreviations

AI = artificial intelligence, AIM-NT = Automated Identification and Measurement of NT, FMF = Fetal Medicine Foundation, NT = nuchal translucency

Summary

The Automated Identification and Measurement of Nuchal Translucency model for nuchal translucency assessment demonstrated high diagnostic accuracy and closely matched radiologist workflow patterns in real-time US diagnostics.

Key Points

- The Automated Identification and Measurement of Nuchal Translucency (AIM-NT) model is an artificial intelligence–based tool developed using the You Only Look Once version 7 framework to automatically identify the nuchal translucency (NT) plane and measure NT thickness in prenatal US imaging.
- The AIM-NT model demonstrated high accuracy comparable with that of experienced radiologists in NT plane identification (88.8% vs 87.6%; $P = .69$), and its NT plane identification time closely matched radiologists' timelines.
- The AIM-NT model showed minimal discrepancies compared with radiologists for measurements of NT thickness, with a mean difference of 0.03 mm and mean absolute error of 0.22 mm (95% CI: 0.19, 0.25) for a standard NT plane.

Keywords

Ultrasound, Fetus, Segmentation, Feature Detection, Diagnosis, Convolutional Neural Network (CNN)

Materials and Methods

Study Design and Data Source

Study design.

This study was conducted under the approval of the Medical Research Ethics Committee (project K-2023-067-H01 and no. 2021-006). Due to its retrospective nature, the requirement for written informed consent was waived by the ethics committee. Data management and analysis were independently conducted by one author (Y.Z.) to ensure objectivity and eliminate any potential conflicts of interest.

In this multicenter, retrospective study, data were collected from 3153 fetuses with 3959 US images and 267 fetuses with 267 videos during NT screenings conducted between 11 and 13 weeks and 6 days of gestation. The study was carried out at the Affiliated Suzhou Hospital of Nanjing Medical University and First People's Hospital of Foshan from January 2022 to October 2023. The inclusion criteria for the dataset were healthy singleton fetuses without chromosomal abnormalities, as determined by noninvasive prenatal testing or follow-up results, and a crown-rump length between 45 and 84 mm. The study design is shown in Figure 1A.

US imaging.

US imaging was performed (GE Voluson E8/E10, GE HealthCare; Philips EPIQ7/Affiniti 70, Philips Healthcare; Samsung UGEO WS80A, Samsung Medison), and data such as maternal age, body mass index, gestational age, and machine type were recorded. The US images and videos contained NT planes, with measurements that adhered to the standards set by the International Society of Ultrasound in Obstetrics and Gynecology and the Fetal Medicine Foundation (FMF). The standard NT plane followed six criteria: magnification, standard midsagittal plane, neutral fetal position, calipers "on-to-on," maximum lucency, and thin nuchal membrane. Magnification was required only when the fetal head and thorax occupied the entire screen. Standard midsagittal plane requires the presence of the tip of the nose and the rectangular shape of the palate anteriorly, translucent diencephalon and nuchal membrane posteriorly, and no visibility of the frontal process of the maxilla. The neutral fetal position requires the head aligned with the spine without hyperextension or flexion. The maximum lucency is required to measure the widest part of the translucency. Calipers on-to-on requires placing the calipers correctly on the border of the translucency. A thin nuchal membrane requires a sharp edge of the line to place the caliper for nuchal measurement.

US examinations were performed by experienced radiologists (C.Z.), each with 15–20 years of expertise, and all radiologists had completed a comprehensive training program established by the FMF. A total of 3959 images comprised the internal dataset for the development phase, which includes training and testing of the AIM-NT model, and 267 videos constituted the external testing dataset to assess the model's performance. Quality control for all data was rigorously conducted by radiology specialists (X.D., L.Y., W.H.) with over 30 years of experience in fetal examination to ensure the presence of a standard NT plane in each fetal video met the FMF guideline.

Development of AIM-NT Model Using the Internal Dataset

Data annotation.

The development of AIM-NT model is shown in Figure 1B. To effectively use image information, three annotation strategies were implemented, and a US expert performed strict quality control for all three categories.

1. Standardization classification for overall plane quality: Experienced radiologists (with 10–20 years of experience in fetal examination) conducted subjective evaluations of image quality independently. Each image was assigned a binary qualification score (1 for standard and 0 for nonstandard). These labels trained the AIM-NT model to recognize standard NT planes globally.
2. Key structural boxes: Experienced radiologists annotated key anatomic structures using bounding boxes that highlighted critical areas like the head, nasal tip, palate anteriorly, diencephalon, nuchal membrane posteriorly, and thorax. This enhanced the AIM-NT model's understanding of local anatomic details.
3. NT thickness calipers: Experienced radiologists who had annotated key structural boxes also marked NT thickness calipers, teaching the AIM-NT model the spatial coordinates of the measurement line end to aid in NT value prediction.

Data distribution and preprocessing.

A total of 3959 cases were collected, with each case contributing a single NT image to the internal dataset for model training and testing. The dataset was divided into internal training and validation datasets in a 7:3 ratio. The images underwent

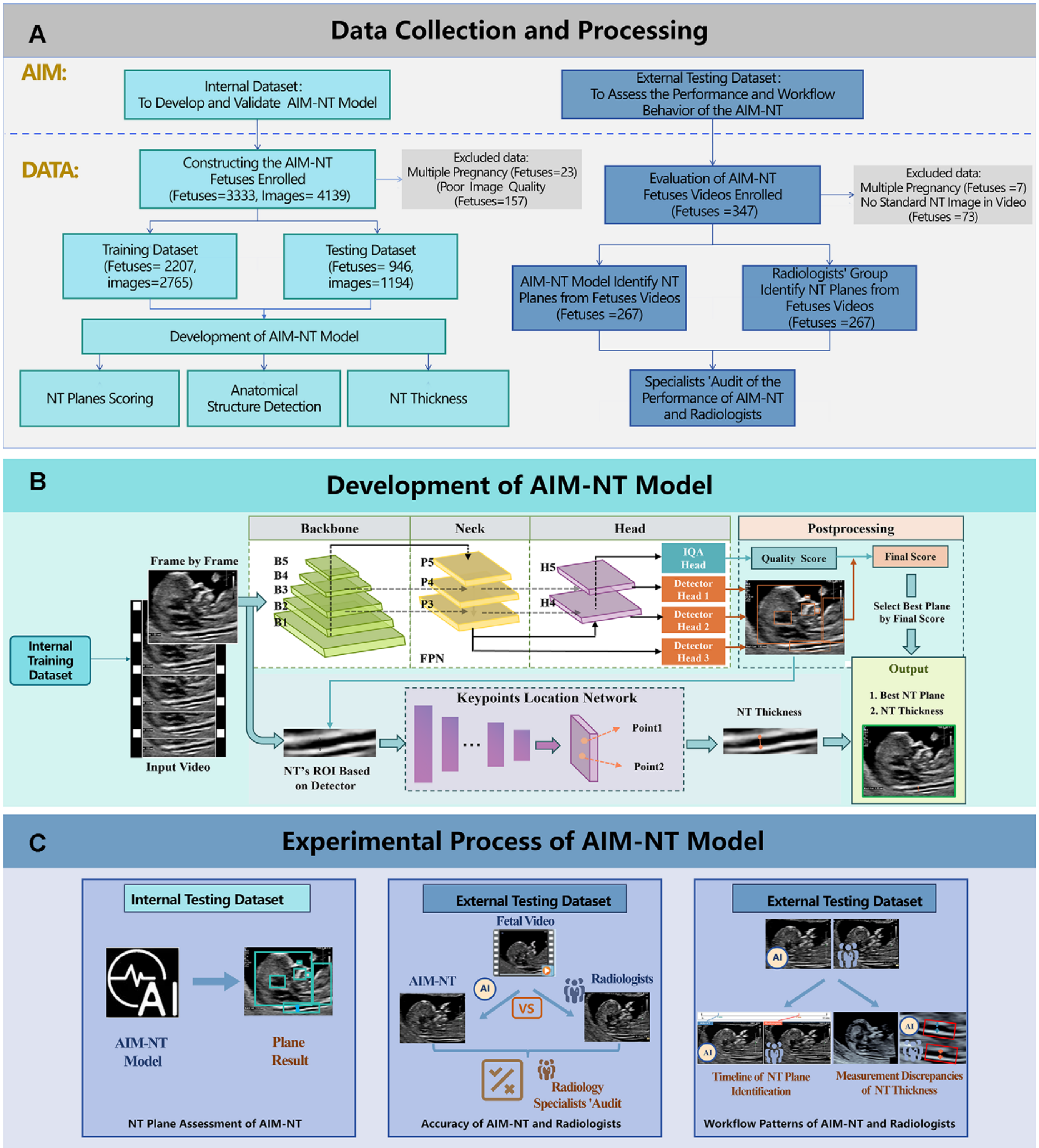

Figure 1: Overview of the Automated Identification and Measurement of Nuchal Translucency (AIM-NT) study. **(A)** Data collection and processing. **(B)** Development of the AIM-NT model. **(C)** Processes for model evaluation on the internal and external testing dataset. AI = artificial intelligence, IQA= image quality assessment, NT = nuchal translucency, ROI = region of interest.

preprocessing that included augmentation techniques such as scaling, cropping, rotation, and noise addition before being resized to 512 × 512 pixels for network training.

The framework of AIM-NT.

The AIM-NT model integrates three specialized modules, each meticulously designed to address key challenges in NT plane analysis. The selection of residual neural network (ResNet) (30) and You Only Look Once version 7 (31) was guided by their proven effectiveness in handling the distinct requirements of feature extraction, detection, and measurement tasks.

1. Plane-scoring module: This module leverages the ResNet architecture to evaluate NT plane quality by learning the relationship between plane features and labeled scores. It outputs a standardized score between

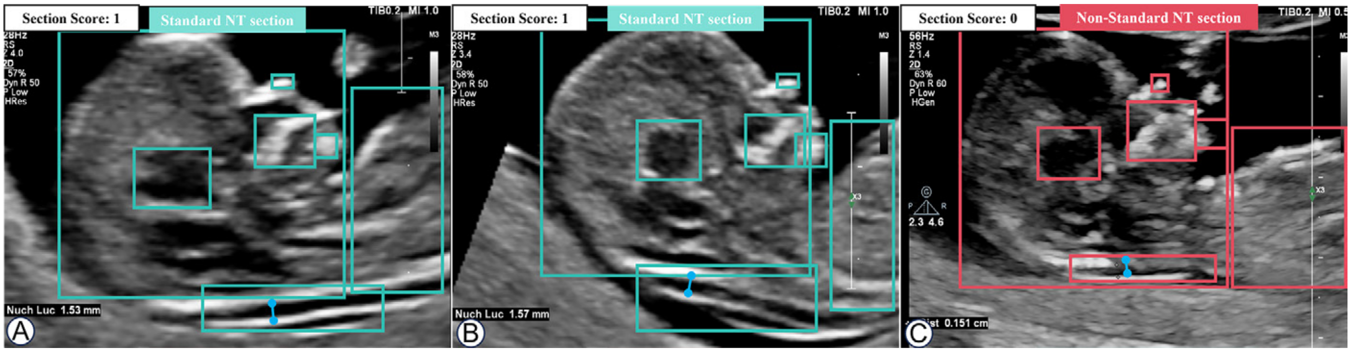

Figure 2: Predictions generated by the Automated Identification and Measurement of Nuchal Translucency (AIM-NT) model on the external testing dataset. **(A, B)** Results for the standard nuchal translucency (NT) plane. **(C)** Results for a nonstandard NT plane predicted by the AIM-NT model.

0 to 1, classifying a plane scoring above the 0.5 threshold as qualified. ResNet's deep residual learning ensures robust feature representation and reliable plane quality assessment across diverse input conditions.

2. Anatomic structure detection module: Powered by You Only Look Once version 7, this module efficiently identifies anatomic structures within NT planes, outputting their positions, categories, and confidence scores. You Only Look Once version 7's advanced anchor-free design and enhanced feature fusion make it particularly adept at detecting small, intricate structures characteristic of NT regions. Postprocessing corrections, aligned with FMF regulations, further enhance accuracy and consistency.

3. NT measurement module: Using ResNet, this module refines the detection network's output by isolating the NT box and focusing on relevant regions for precise measurement. It calculated end point coordinates and pixel distances, which are subsequently converted to physical measurements using a calibrated pixel-per-centimeter ratio. ResNet's robust multiscale feature extraction supports high precision and reliability in measurement tasks. Figure 2 provides a visual representation of the predictions generated by the AIM-NT model, highlighting the seamless integration and performance of its modules.

Assessment of AIM-NT Model Performance on Internal and External Testing Datasets

The process for evaluation of model performance is shown in Figure 1C and described below.

Plane assessment of AIM-NT.

To evaluate NT plane assessment in the internal testing dataset, the AIM-NT model underwent validation across three modules on the internal testing dataset: plane-scoring, anatomic structure detection, and NT measurement, comparing results with reference standard data labeled by experienced radiologists.

Accuracy of AIM-NT and radiologists.

To assess the accuracy of the AIM-NT model, eight experienced radiologists (C.Z., with 15–20 years of experience in fetal examination), who were not involved in the annotation, conducted an independent assessment to identify the standard NT plane and measure NT thickness on the external testing dataset. Using the Pair version 3.0 (RayShape) (32) annotation software in compliance with standards, the radiologists independently selected and measured the optimal NT plane. Simultaneously, the AIM-NT model automatically processed videos to identify and measure the best NT plane. In addition, three specialists (X.D., L.Y., and W.H., with 30 years of experience in fetal examination) audited the results from both AIM-NT and the radiologist groups to ensure impartial and objective assessments. They evaluated the images using a protocol that mirrors the FMF's review process, classifying each as pass (standard NT plane) or fail (nonstandard NT plane) based on six defined criteria and documenting reasons for any fail ratings. Only NT planes rated as standard by two or more radiology specialists would be considered pass; otherwise, they would be rated fail. The assessment was double-blind to guarantee impartiality, preventing identification of the groups from the images alone.

Workflow patterns of AIM-NT and radiologists.

To ensure alignment in workflow patterns between the AIM-NT model and the same eight experienced radiologists on the external testing dataset, comparisons were made regarding the timeline of NT plane identification and the measurement discrepancies in NT thickness. To assess the relationship between NT measurement discrepancies and gestational age, we employed a linear mixed-effects model.

Statistical Analysis

To evaluate the AIM-NT model's performance, metrics such as the area under the receiver operating characteristic curve, accuracy, 95% CI, sensitivity, specificity, mean average precision, mean absolute error, SD, and Pearson correlation coefficient were used on the internal testing dataset. To compare the accuracy and workflow patterns of the AIM-NT model with those of eight experienced radiologists in the external test dataset, analyses included NT plane identification metrics like pass and failure rates, NT measurement metrics such as means and SDs, t tests, the consistency of workflow patterns including timelines of NT identification, Bland-Altman plots, and measurement errors of NT thickness. All analyses were conducted using Python version 3.11 (Python Software Foundation), with a P value less than .05 deemed statistically significant.

Table 1: Demographic and Clinical Characteristics of Training, Internal Testing, and External Testing Datasets Used to Develop and Evaluate AIM-NT Model

Characteristics	Internal Training Dataset (Fetuses = 2207, Images = 2765)	Internal Testing Dataset (Fetuses = 946, Images = 1194)	External Testing Dataset (Fetuses = 267, Videos = 267)
Mean MA (y)	32 ± 3	33 ± 2	31 ± 4
BMI	23 ± 0.1	23 ± 5.2	24 ± 3.2
Mean GA (weeks ± days)	12 ± 5	12 ± 3	12 ± 5
Mean CRL (mm)	63 ± 6.18	60 ± 5.01	64 ± 6.84
Duration of the video (min)	NA	NA	15
Machine type (images)			
GE E8	630	279	45
GE E10	989	424	175
Philips EPIQ7	523	224	17
Philips Affiniti 70	491	210	30
Samsung UGEO WS80A	83	36	0
Samsung Medison	49	21	0

Note.—Data are presented as means ± SDs or frequency. AIM-NT = Automated Identification and Measurement of Nuchal Translucency, BMI = body mass index (measured by dividing the weight in kilograms by the height in centimeters squared), CRL = crown-rump length, GA = gestational age, MA = maternal age, NA = not applicable.

Results

Study Sample Characteristics

The study flowchart is shown in Figure 1, and clinical characteristics of the internal and external testing datasets used in this study are detailed in Table 1. The internal training dataset included 2207 fetuses with 2765 US images, with a mean ± SD maternal age of 32 years ± 3, and a mean gestational age of 12 weeks ± 5. The internal testing dataset included 946 fetuses with 1194 US images, with a mean maternal age of 33 years ± 2 and a mean gestational age of 12 weeks ± 3. The external test dataset included 267 US videos of fetuses, with a mean maternal age of 31 years ± 4 and a mean gestational age of 12 weeks ± 5.

Performance of AIM-NT Model on the Internal Testing Dataset

The plane scoring module of the AIM-NT model achieved an area under the receiver operating characteristic curve of 0.92 (95% CI: 0.88, 0.95) on the internal testing dataset, with both sensitivity and specificity at 90.4% (95% CI: 88.2, 93.1). The anatomic structure detection module had a mean average precision of 0.80. The NT measurement module displayed a mean absolute error of 0.02 mm ± 0.02 and a Pearson correlation coefficient of 0.91, demonstrating the model's precision and reliability in NT assessment.

Accuracy of AIM-NT Model on External Testing Dataset

For NT plane identification, the comparative results from the specialists' evaluations of the AIM-NT model and radiologists are outlined in Table 2. The table categorizes results as pass for standard images and fail for nonstandard images. The AIM-NT group achieved an accuracy of 88.8% with 237 planes, and radiologists achieved an accuracy of 87.6% with 234 planes, with no evidence of a difference between groups ($\chi^2 = 0.16$; $P = .69$). The primary reason for image failure in the AIM-NT group was the neutral fetal position, whereas image failure in the radiologist group was most commonly due to midsagittal plane issues. Figure 3 displays example diagrams illustrating the reasons for an image to pass and fail based on the specialists' audits.

For NT thickness measurement, there was no evidence of differences between the two groups in terms of accuracy on standard NT planes (analysis of variance, $P = .29$) and nonstandard planes (analysis of variance, $P = .59$). The means and SDs of the NT thickness conducted from the AIM-NT and radiologist groups were 1.49 mm ± 0.42 and 1.53 mm ± 0.42, respectively, as detailed in Table S1.

The Alignment of Workflow Behavior of AIM-NT Model on the External Testing Dataset

In the analysis of workflow patterns related to the timeline of NT plane identification, the AIM-NT model and radiologists demonstrated high consistency. Time discrepancies between the two groups were most commonly within 1 minute (207 of 267, 77.9%), indicating that both typically acquire images simultaneously. Specifically, 67.9% (146 of 215) of planes that achieved a pass result and 65.4% (34 of 52) of planes in the fail category showed no time discrepancies (Table 3). Figure 4 illustrates these findings, presenting paired images of NT planes from the same video, thereby emphasizing the consistency between the two methods.

In analyzing the measurement discrepancies of NT thickness, both the AIM-NT model and radiologists showed high consistency when evaluating standard NT images that achieved a pass result. The Bland-Altman plot showed a minimal mean difference of 0.03 mm in NT thickness measurement between the two groups (Fig 5). Additionally, across the paired NT images, the measurement discrepancies between the AIM-NT model and radiologists showed a total mean absolute error of 0.22 mm (95%

Table 2: Accuracy of NT Plane Identifications by the AIM-NT Model and Radiologists on the External Testing Dataset

	No. of Cases	
Image Result	AIM-NT	Radiologists
Pass	237 (88.8%)	234 (87.6%)
Fail	30 (11.2%)	33 (12.4%)
Standard MSP	12	19
Neutral fetal position	11	7
Maximum lucency	4	1
Thin nuchal membrane	3	3
Calipers on-to-on	1	1
Magnification	4	3

Note.—Accuracy of US images was assessed by radiology specialists, who classified each image as pass for standard or fail for nonstandard NT planes based on six criteria defined by the Fetal Medicine Foundation. A pass image met all six criteria, whereas a fail image was assigned specific reasons. Because specialists may list multiple reasons for a single nonstandard image, the total number of fail reasons exceeds the count of fail images. AIM-NT = Automated Identification and Measurement of Nuchal Translucency, MSP = midsagittal plane, NT = nuchal translucency.

CI: 0.19, 0.25). The analysis indicated that gestational age did not significantly affect NT measurement discrepancies, with a gestational age coefficient of 0.007 (95% CI: −0.002, 0.016; P = .14). More details regarding gestational age–level mean absolute error, median error, and SD are presented in Figure 6 and Table 4. The relationship between NT thickness and crown-rump length, based on the guideline model curve by Wright et al (33) recommended by the FMF, revealed that NT thickness measurements from both the AIM-NT model and radiologists consistently fell below the recommended values (Fig S1).

Discussion

The development of the AIM-NT model, leveraging the You Only Look Once version 7 framework (31), represents a substantial advance in integrating AI with prenatal NT screening protocols. The model demonstrated high performance, with no evidence of differences compared with experienced radiologists in NT plane identification accuracy (88.8% vs 87.6%; P = .69) or NT thickness measurements on both standard (P = .29) and nonstandard planes (P = .59). The model also showed minimal discrepancies compared with radiologists in NT plane identification time and NT thickness measurement errors. These findings underscore AIM-NT's practical applicability and reliability in clinical settings.

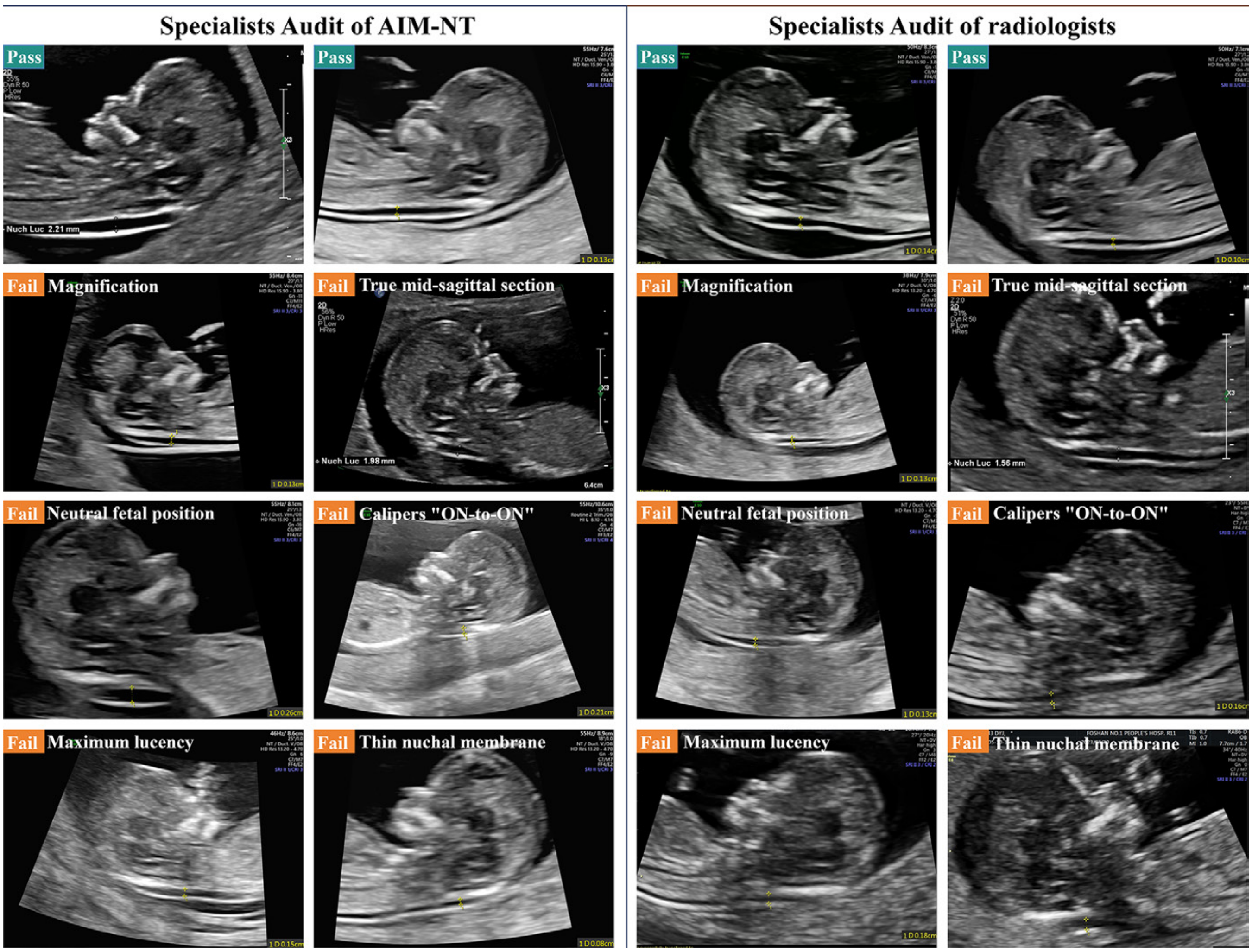

Figure 3: Example US images with pass and fail results for the nuchal translucency (NT) plane based on radiology specialist audits in both Automated Identification and Measurement of Nuchal Translucency (AIM-NT) and radiologist groups. Reasons for each fail result are shown.

Table 3: Time Discrepancies between AIM-NT and Radiologists in NT Plane Identification according to Image Results of Pass or Fail on the External Testing Dataset

	Time Discrepancies (min)											
Cases	0	1	2	3	4	5	6	7	8	9	10	>10
No. of pass	148	20	15	7	9	7	4	0	1	1	0	3
No. of fail	33	7	4	2	2	1	2	1	0	0	0	0
Total	181	27	19	9	11	8	6	1	1	1	0	3

Note.—AIM-NT = Automated Identification and Measurement of Nuchal Translucency, NT = nuchal translucency.

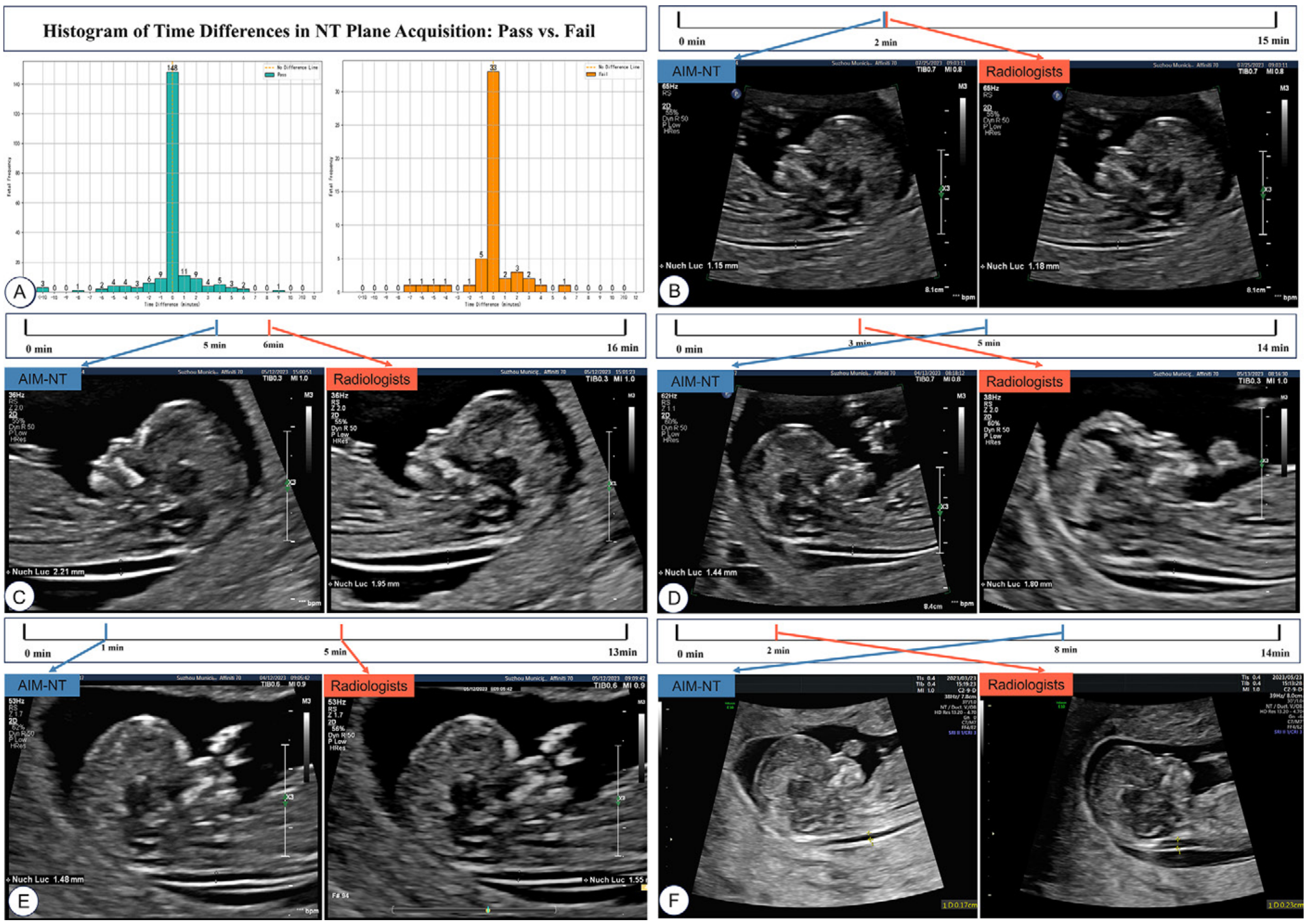

Figure 4: Timelines of nuchal translucency (NT) plane identification on the external testing dataset. Different timelines of paired NT plane image acquisition from the same video by the Automated Identification and Measurement of Nuchal Translucency (AIM-NT) model and radiologist. **(A)** Histogram of time differences between the two groups in NT plane (green: pass; orange: fail). **(B–F)** Identification time differences of between AIM-NT (blue) and radiologists (red): 0 minutes **(B)**, 1 minute **(C)**, 2 minutes **(D)**, 4 minutes **(E)**, and 6 minutes **(F)**.

Previous research on NT plane analysis has primarily focused on enhancing accuracy and implementing stringent quality control measures (21–23,29), such as the NT assessment criteria provided by the FMF (1), but has lacked real-time feedback during fetal screenings. Therefore, providing a real-time AI-based model that adheres to quality control standards could substantially enhance the accuracy of NT plane identification. Additionally, because US is a real-time method for NT plane identification, its crucial to consider radiologists' workflow patterns (34), which is often overlooked in research. This consideration highlights a preference among radiologists for manual results over AI-assisted outcomes, primarily due to asynchronous workflow patterns.

Our study concentrated on real-time US scanning for NT planes, demonstrating high performance, compared with traditional assessments by radiologists. Regarding the alignment of workflow patterns in the NT identification timeline, our study demonstrates that the AIM-NT model's ability to synchronize with radiologists' identification timelines facilitates its seamless integration into existing clinical operations without disrupting the pace of workflows. However, two notable scenarios emerged concerning the identification of standard NT planes. First, AIM-NT sometimes identified a standard NT image faster than radiologists, especially when fetal movements disrupted manual scanning, by analyzing each frame. In the second scenario, our study observed that despite minor differences in identification rates and measurements, radiologists often selected NT images that did not align with AIM-NT standards due to multiple scans during fetal screening. This highlighted differing approaches to image selection: the AIM-NT model focuses on anatomic details, showing greater accuracy in midsagittal plane identification, whereas radiologists prioritize overall image recognition, excelling more in identifying neutral fetal positions. This distinction underlines the complementary strengths of AI and human expertise in prenatal imaging. In summary, AIM-NT's ability to accurately capture NT results and align closely with radiologist-like workflow patterns support its potential for real-time NT screening.

Another challenge in past research is minimizing measurement errors in NT assessments and establishing an acceptable range of error, due to the close association of NT thickness with anomalies and chromosomal disorders (10,11). Because NT is measured in millimeters, even slight deviations can lead to different diagnoses, potentially resulting in misdiagnoses (22,35,36). According to a study by Pandya et al (24), the intraobserver difference in NT measurement was 0.54 mm, whereas the interobserver difference was 0.62 mm. Such discrepancies could lead to significantly different diagnoses. For instance, a measurement of 3 mm could result in a false-negative diagnosis if the true NT measurement is 3.5 mm (37). In another study (25), a comparison of manual and semiautomated NT measurements showed that semiautomated measurements tended to be slightly higher, with a mean difference of 0.04 mm and a mean SD of 0.109 mm. Despite this, semiautomated methods still require senior clinicians' involvement for final decision-making. The accuracy of NT measurements depends on various factors, such as the measurement technique employed, user training, and implementation of quality control measures (38,39). Furthermore, the scarcity of specific guidelines on acceptable NT measurement error ranges underscores the urgent need for detailed standards that can ensure more accurate prenatal diagnostics and reduce the risk of misdiagnosis.

In our study, we focused on the consistency of NT measurements and their error margins. We achieved high measurement consistency with minimal discrepancies compared with physicians. Notably, although the precision of a caliper's adjustment is typically 0.1 mm, we managed to control our mean difference to within an impressive 0.03 mm using the Bland-Altman plot. Additionally, our research established that the acceptable measurement discrepancies for a standard NT plane should have a total mean absolute error of 0.22 mm (95% CI: 0.19, 0.25). Furthermore, we found that the distribution of NT measurement discrepancies did not show a clear relationship with gestational age. Although our findings regarding the mean difference were similar to those of previous studies (25), this study was the first to compare manual and automated methods rather than focusing on observer comparisons. This insight has the potential to serve as a criterion for evaluating NT measurements in AI-based measurement systems, providing a standardized benchmark for future developments.

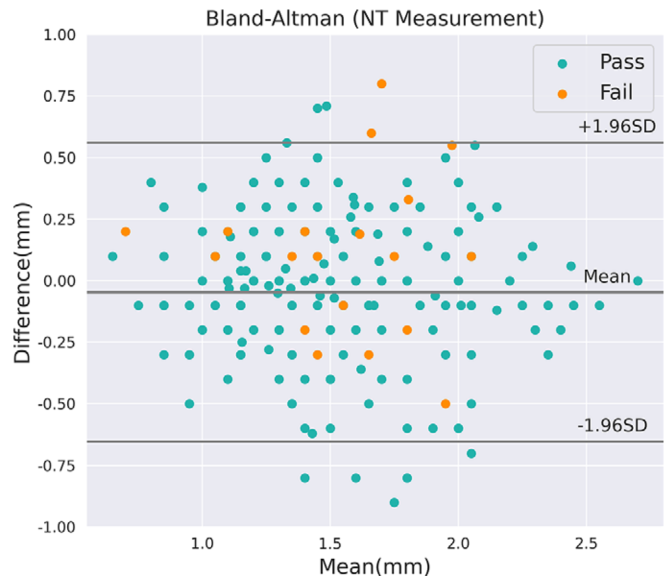

Figure 5: Bland-Altman plot of the consistency between the Automated Identification and Measurement of Nuchal Translucency (AIM-NT) model and radiologists for nuchal translucency (NT) measurements in the standard NT plane (green dots: pass) and nonstandard NT plane (orange dots: fail) in the external testing dataset.

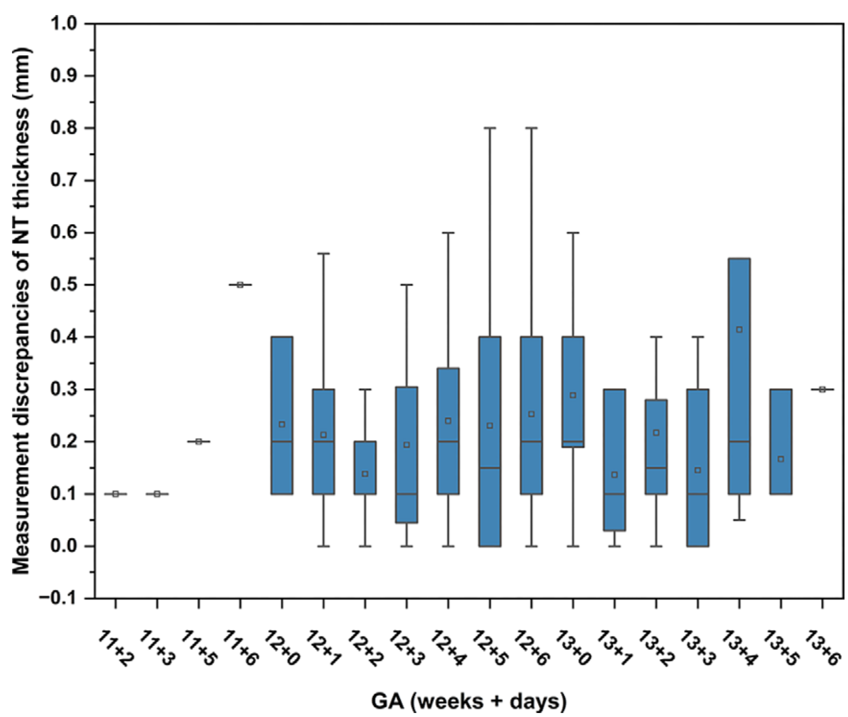

Figure 6: Box and whiskers plots of measurement discrepancies between the Automated Identification and Measurement of Nuchal Translucency (AIM-NT) model and radiologists in nuchal translucency (NT() thickness at different gestational ages (GAs) on the external testing dataset. Blue boxes represent SD, the boxes represents mean absolute error, and the lines represent median error.

Table 4: Measurement Discrepancies of NT Thickness of the Standard NT Plane Acquired by AIM-NT and the Radiologists at Different Gestational Ages on the External Testing Dataset

GA (Weeks + Days)	No. of Fetuses	NT MAE (mm)	NT Median Error (mm)	SD (mm)
11 + 2	1	0.10	0.10	0.00
11 + 3	1	0.10	0.10	0.00
11 + 5	1	0.20	0.20	0.00
11 + 6	1	0.50	0.50	0.00
12 + 0	3	0.23	0.20	0.15
12 + 1	15	0.22	0.21	0.15
12 + 2	19	0.14	0.13	0.10
12 + 3	24	0.19	0.14	0.19
12 + 4	25	0.24	0.21	0.20
12 + 5	23	0.23	0.16	0.24
12 + 6	34	0.25	0.19	0.23
13 + 0	19	0.29	0.23	0.21
13 + 1	10	0.14	0.10	0.13
13 + 2	16	0.22	0.17	0.20
13 + 3	11	0.15	0.13	0.14
13 + 4	7	0.41	0.20	0.55
13 + 5	3	0.17	0.17	0.12
13 + 6	1	0.30	0.30	0.00
Total	214	0.22	0.20	0.21

Note.—AIM-NT = Automated Identification and Measurement of Nuchal Translucency, GA = gestational ages, MAE = mean absolute error, NT = nuchal translucency.

With its high performance and alignment with the workflow of senior radiologists, the AIM-NT model has the potential to be an essential assistive tool in real-world clinical scenarios. It can be employed for real-time NT scan screening, automatically identifying standard NT planes and measuring NT thickness, thereby enhancing consistency and efficiency in line with the practices of senior radiologists. Furthermore, it may serve as a training tool for junior radiologists by replicating senior radiologist workflows to select optimal planes and ensure accurate measurements. It can also serve as an educational tool for medical students, providing detailed structural visualization in accordance with the strict criteria of International Society of Ultrasound in Obstetrics and Gynecology and FMF.

Although this study presents important advances, it is important to acknowledge its limitations. First, the lack of positive cases to validate the AIM-NT model's performance and its alignment with senior radiologists in diagnosing abnormal NT ranges is a substantial limitation for detecting chromosomal abnormalities in clinical practice (37). Second, it is necessary to establish a clinical experimental environment to validate the effectiveness of AIM-NT in real time. Third, because AI models are intended to serve as assistive tools, the responsibility for the final diagnosis ultimately lies with radiologists. Therefore, the acceptance of AI's diagnostic and measurement results by radiologists, and the extent to which these results need adjustment to align with radiologists' expectations, were not fully explored in this study. To address these limitations, a validation process involving positive cases, including those with NT greater than 3.5 mm and noninvasive prenatal testing–positive cases, will be added to enhance the robustness of the AIM-NT model. The model will be validated on both normal and abnormal NT data to assess its performance, alignment with the workflow of senior radiologists—including plane identification, measurement, and final diagnosis—and effectiveness as an assistive tool for junior radiologists. Additionally, we plan to implement the model in a hospital setting and conduct a prospective trial to evaluate its real-time performance during NT scans in alignment with radiologists, thereby improving its applicability in real-world clinical scenarios. This validation process aims to improve the model's reliability and applicability.

In conclusion, our AI-based model, AIM-NT, meets the requirements for NT assessment by offering high accuracy in NT plane identification and thickness measurement and aligning with radiologist workflow. It may serve as an effective assistive tool to radiologists during real-time NT assessments. Future studies should focus on validating AIM-NT in clinical settings, expanding sample sizes, and further enhancing its diagnostic accuracy to improve prenatal US screening.

Author affiliations:

[1] National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, School of Biomedical Engineering, Health Science Center, Shenzhen University, No. 3688 Nanhai Avenue, Nanshan District, Shenzhen 518060, PR China

[2] Medical Ultrasound Image Computing (MUSIC) Lab, Shenzhen University, Shenzhen, PR China

[3] Marshall Laboratory of Biomedical Engineering, Shenzhen University, Shenzhen, PR China

[4] Center for Medical Ultrasound, The Affiliated Suzhou Hospital of Nanjing Medical University, Suzhou Municipal Hospital, Suzhou, PR China

[5] Shenzhen RayShape Medical Technology Co, Ltd, Shenzhen, PR China

[6] Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, PR China

[7] Department of Medical Ultrasonics, First People's Hospital of Foshan, Foshan, PR China

Received August 12, 2024; revision requested October 3; revision received February 13, 2025; accepted April 14.

Address correspondence to: D.N. (email: nidong@szu.edu.cn).

Funding: Supported by Science and Technology Planning Project of Guangdong Province (2023A0505020002), National Natural Science Foundation of China (no. 62101343, 62171290, 12326619), Science and Technology Development Fund of Macao (grant no. 0021/2022/AGJ), National Key Research and Development Program of China (2019YFC0118300), Multi-center Clinical Study of Intelligent Prenatal Ultrasound (ChiCTR2300071300).

Author contributions: Guarantors of integrity of entire study, **Y.Z., X.Y., C.J., Y.C., C.C., H.S., C.Z., X.D., L.Y., D.N.**; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, **Y.Z., X.Y., C.J., Y.C., C.Z., W.H., L.Y., D.N.**; clinical studies, **Y.Z., C.J., B.L., C.Z., W.H., X.D., L.Y., D.N.**; experimental studies, **Y.Z., C.J., X.H., Y.C., C.C., H.S., W.H., X.D., D.N.**; statistical analysis, **Y.Z., X.Y., C.J., X.H., Y.C., C.C., H.S., B.L., C.Z., W.H., X.D.**; and manuscript editing, **Y.Z., X.Y., C.J., Y.C., B.L., W.H., L.Y., D.N.**

Acknowledgments: The authors express their gratitude to the FMF and the International Society of Ultrasound in Obstetrics and Gynecology for providing the standards used to evaluate the NT plane. Additional thanks go to the doctors and nurses at the Center for Medical Ultrasound, Affiliated Suzhou Hospital of Nanjing Medical University, and the Department of Medical Ultrasonics at the First People's Hospital of Foshan for their invaluable assistance in data collection. We also appreciate the efforts of our colleagues from Shenzhen University in developing the AI model. Special acknowledgment is given to the women who participated in this study, whose cooperation was crucial in providing the data used for this research.

Data sharing: The datasets analyzed during the current study are not publicly available due to patient privacy restrictions.

Disclosures of conflicts of interest: **Y.Z.** No relevant relationships. **X.Y.** No relevant relationships. **C.J.** No relevant relationships. **X.H.** No relevant relationships. **Y.C.** No relevant relationships. **C.C.** No relevant relationships. **H.S.** No relevant relationships. **B.L.** No relevant relationships. **C.Z.** No relevant relationships. **W.H.** No relevant relationships. **X.D.** No relevant relationships. **L.Y.** No relevant relationships. **D.N.** No relevant relationships.

References

1. Merz E, Thode C, Hackelöer BJ, et al. Quality Assurance of Ultrasound Examinations during First Trimester Screening: the Fetal Medicine Foundation (FMF) Germany after 20 Years. Ultraschall Med 2022;43(2):115–119.
2. Bilardo CM, Chaoui R, Hyett JA, et al; International Society of Ultrasound in Obstetrics and Gynecology. ISUOG Practice Guidelines (updated): performance of 11–14-week Ultrasound Scan. Ultrasound Obstet Gynecol 2023;61(1):127–143.
3. Rawat V, Jain A, Shrimali V. Automated Techniques for the Interpretation of Fetal Abnormalities: A Review. Appl Bionics Biomech 2018;2018:6452050.
4. Sun Y, Zhang L, Dong D, et al. Application of an Individualized Nomogram in First-Trimester Screening for Trisomy 21. Ultrasound Obstet Gynecol 2021;58(1):56–66.
5. Borrell A, Robinson JN, Santolaya-Forgas J. Report on the 11- to 13+6-week Ultrasound Evaluation as a Screening Test for Trisomy 21 in Singleton Pregnancies. Am J Perinatol 2009;26(10):703–710.
6. 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(7):671–678.
7. Chen Y, Wang X, Li L, Lu S, Zhang Z. New Cut-off Values for Screening of Trisomy 21, 18 and Open Neural Tube Defects (ONTD) during the Second Trimester in Pregnant Women with Advanced Maternal Age. BMC Pregnancy Childbirth 2020;20(1):776.
8. Chervenak FA, McCullough LB, Chasen ST. Clinical Implications of the Ethics of Informed Consent for First-Trimester Risk Assessment for Trisomy 21. Semin Perinatol 2005;29(4):277–279.
9. Mersy E, Smits LJM, van Winden LAAP, et al; South-East Netherlands NIPT Consortium. Noninvasive Detection of Fetal Trisomy 21: systematic Review and Report of Quality and Outcomes of Diagnostic Accuracy Studies Performed between 1997 and 2012. Hum Reprod Update 2013;19(4):318–329.
10. Kagan KO, Wright D, Etchegaray A, Zhou Y, Nicolaides KH. Effect of Deviation of Nuchal Translucency Measurements on the Performance of Screening for Trisomy 21. Ultrasound Obstet Gynecol 2009;33(6):657–664.
11. Snijders RJM, Thom EA, Zachary JM, et al. First-Trimester Trisomy Screening: nuchal Translucency Measurement Training and Quality Assurance to Correct and Unify Technique. Ultrasound Obstet Gynecol 2002;19(4):353–359.
12. Allan L. Prenatal Diagnosis of Structural Cardiac Defects. Am J Med Genet C Semin Med Genet 2007;145C(1):73–76.
13. Geipel A, Gembruch U. Screening Performance of First Trimester Nuchal Translucency, Ductus Venosus Blood Flow and Tricuspid Regurgitation for Cardiac Defects. Z Geburtshilfe Neonatol 2012;216(4):157–161.
14. Maiz N, Nicolaides KH. Ductus Venosus in the First Trimester: contribution to Screening of Chromosomal, Cardiac Defects and Monochorionic Twin Complications. Fetal Diagn Ther 2010;28(2):65–71.
15. Papatheodorou SI, Evangelou E, Makrydimas G, Ioannidis JPA. First-trimester ductus venosus screening for cardiac defects: a meta-analysis. BJOG 2011;118(12):1438–1445.
16. Allan LD. Cardiac Anatomy Screening: what is the Best Time for Screening in Pregnancy? Curr Opin Obstet Gynecol 2003;15(2):143–146.
17. Burger NB, Bekker MN, de Groot CJM, Christoffels VM, Haak MC. Why increased nuchal translucency is associated with congenital heart disease: a systematic review on genetic mechanisms. Prenat Diagn 2015;35(6):517–528.
18. Hyett J. Does Nuchal Translucency Have a Role in Fetal Cardiac Screening? Prenat Diagn 2004;24(13):1130–1135.
19. Pajkrt E, Weisz B, Firth HV, Chitty LS. Fetal Cardiac Anomalies and Genetic Syndromes. Prenat Diagn 2004;24(13):1104–1115.
20. Burgin KB. Prenatal Genetic Screening. J Midwifery Womens Health 2008;53(4):391–392.
21. Chen H, Wu L, Dou Q, et al. Ultrasound Standard Plane Detection Using a Composite Neural Network Framework. IEEE Trans Cybern 2017;47(6):1576–1586.
22. Park J, Sofka M, Lee S, Kim D, Zhou SK. Automatic Nuchal Translucency Measurement from Ultrasonography. In: Mori K, Sakuma I, Sato Y, Barillot C, Navab N, eds. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013. MICCAI 2013. Lecture Notes in Computer Science, vol 8151. Springer, 2013; 243–250.
23. Axell RG, Gillett A, Pasupathy D, et al. Accuracy of Nuchal Translucency Measurement Depends on Equipment Used and Its Calibration. Ultrasound Obstet Gynecol 2014;44(1):31–37.
24. Pandya PP, Altman DG, Brizot ML, Pettersen H, Nicolaides KH. Repeatability of measurement of fetal nuchal translucency thickness. Ultrasound Obstet Gynecol 1995;5(5):334–337.
25. Moratalla J, Pintoffl K, Minekawa R, et al. Semi-automated system for measurement of nuchal translucency thickness. Ultrasound Obstet Gynecol 2010;36(4):412–416.
26. Walker MC, Willner I, Miguel OX, et al. Using deep-learning in fetal ultrasound analysis for diagnosis of cystic hygroma in the first trimester. PLoS One 2022;17(6):e0269323.
27. Zhang L, Dong D, Sun Y, et al. Development and Validation of a Deep Learning Model to Screen for Trisomy 21 During the First Trimester From Nuchal Ultrasonographic Images. JAMA Netw Open 2022;5(6):e2217854.
28. Wataganara T. Deep Learning for Nuchal Translucency-Based Screening for Fetal Trisomy 21. JAMA Netw Open 2022;5(6):e2217869.
29. Clementschitsch G, Hasenöhrl G, Schaffer H, Steiner H. Comparison Between Two- and Three-Dimensional Ultrasound Measurements of Nuchal Translucency. Ultrasound Obstet Gynecol 2001;18(5):475–480.
30. 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). IEEE, 2016; 770–778.
31. Wang CY, Bochkovskiy A, Liao HYM. 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). IEEE, 2023; 7464–7475.
32. Liang J, Yang X, Huang Y, et al. Sketch Guided and Progressive Growing GAN for Realistic and Editable Ultrasound Image Synthesis. Med Image Anal 2022;79:102461.
33. Wright D, Kagan KO, Molina FS, Gazzoni A, Nicolaides KH. A Mixture Model of Nuchal Translucency Thickness in Screening for Chromosomal Defects. Ultrasound Obstet Gynecol 2008;31(4):376–383.
34. Drukker L, Sharma H, Karim JN, Droste R, Noble JA, Papageorghiou AT. Clinical workflow of sonographers performing fetal anomaly ultrasound scans: a deep-learning-based analysis. Ultrasound Obstet Gynecol 2022;60(6):759–765.
35. Nicolaides KH, Wegrzyn P. Increased nuchal translucency with normal karyotype [in Polish]. Ginekol Pol 2005;76(8):593–601.
36. Deng Y, Wang Y, Chen P. Automated detection of fetal nuchal translucency based on hierarchical structural model. In: 2010 IEEE 23rd International Symposium on Computer-Based Medical Systems (CBMS). IEEE, 2010; 78–84.
37. De Domenico R, Faraci M, Hyseni E, et al. Increased nuchal translucency in normal karyotype fetuses. J Prenat Med 2011;5(2):23–26.
38. Anzalone A, Fusco G, Isgrò F, et al. A system for the automatic measurement of the nuchal translucency thickness from ultrasound video stream of the foetus. In: Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems. IEEE, 2013; 239–244.
39. Snijders RJ, Noble P, Sebire N, Souka A, Nicolaides KH. UK Multicentre Project on Assessment of Risk of Trisomy 21 by Maternal Age and Fetal Nuchal-Translucency Thickness at 10-14 Weeks of Gestation. Lancet 1998;352(9125):343–346.