



Original Investigation | Obstetrics and Gynecology

Development and Validation of a Deep Learning Model to Screen for Trisomy 21 During the First Trimester From Nuchal Ultrasonographic Images

Liwen Zhang, PhD; Di Dong, PhD; Yongqing Sun, MD; Chaoen Hu, MS; Congxin Sun, BD; Qingqing Wu, MD; Jie Tian, PhD

Abstract

IMPORTANCE Accurate screening of trisomy 21 in the first trimester can provide an early opportunity for decision-making regarding reproductive choices.

OBJECTIVE To develop and validate a deep learning model for screening fetuses with trisomy 21 based on ultrasonographic images.

DESIGN, SETTING, AND PARTICIPANTS This diagnostic study used data from all available cases and controls enrolled at 2 hospitals in China between January 2009 and September 2020. Two-dimensional images of the midsagittal plane of the fetal face in singleton pregnancies with gestational age more than 11 weeks and less than 14 weeks were examined. Observers were blinded to subjective fetus nuchal translucency (NT) marker measurements. A convolutional neural network was developed to construct a deep learning model. Data augmentation was applied to generate more data. Different groups were randomly selected as training and validation sets to assess the robustness of the deep learning model. The fetal NT was shown and measured. Each detection of trisomy 21 was confirmed by chorionic villus sampling or amniocentesis. Data were analyzed from March 1, 2021, to January 3, 2022.

MAIN OUTCOMES AND MEASURES The primary outcome was detection of fetuses with trisomy 21. The receiver operating characteristic curve, metrics of accuracy, area under the curve (AUC), sensitivity, and specificity were used for model performance evaluation.

RESULTS A total of 822 case and control participants (mean [SD] age, 31.9 [4.6] years) were enrolled in the study, including 550 participants (mean [SD] age, 31.7 [4.7] years) in the training set and 272 participants (mean [SD] age, 32.3 [4.7] years) in the validation set. The deep learning model showed good performance for trisomy 21 screening in the training (AUC, 0.98; 95% CI, 0.97-0.99) and validation (AUC, 0.95; 95% CI, 0.93-0.98) sets. The deep learning model had better detective performance for fetuses with trisomy 21 than the model with NT marker and maternal age (training: AUC, 0.82; 95% CI, 0.77-0.86; validation: AUC, 0.73; 95% CI, 0.66-0.80).

CONCLUSIONS AND RELEVANCE These findings suggest that this deep learning model accurately screened fetuses with trisomy 21, which indicates that the model is a potential tool to facilitate universal primary screening for trisomy 21.

JAMA Network Open. 2022;5(6):e2217854. doi:10.1001/jamanetworkopen.2022.17854

Key Points

Question Can a noninvasive deep learning model identify fetuses with trisomy 21 based on ultrasonographic images?

Findings In this diagnostic study including 822 cases and controls, the deep learning model showed good performance for identifying fetuses with trisomy 21 in the training and validation sets. The deep learning model had higher area under the curve than the model developed by nuchal translucency marker and maternal age.

Meaning These findings suggest that this deep learning model is a potential tool to facilitate the universal primary screening for fetuses with trisomy 21.

+ [Invited Commentary](#)

+ [Supplemental content](#)

Author affiliations and article information are listed at the end of this article.

Open Access. This is an open access article distributed under the terms of the CC-BY License.

JAMA Network Open. 2022;5(6):e2217854. doi:10.1001/jamanetworkopen.2022.17854

June 21, 2022 1/11

Introduction

Trisomy 21 is the most prevalent chromosomal anomaly disorder that causes children's developmental delay and intellectual disability.¹ Accurate screening for trisomy 21, which can provide an early opportunity for decision-making regarding reproductive choices in the first trimester of pregnancy, has been widely investigated in the past few decades.² Currently, analyses of cell-free fetal DNA validation show high accuracy (up to 99%) in screening for trisomy 21.^{3,4} However, some studies have suggested that further cost-saving approaches should be explored, considering the potentially high cost of cell-free fetal DNA testing.^{4,5}

For decades, ultrasonographic images have been widely used for screening fetuses for trisomy 21, owing to the advantages of safety, convenience, and low cost.^{6,7} Fetal nuchal translucency (NT) thickness, measured in ultrasonographic images, has been used to screen fetuses with trisomy 21.^{7,8} Moreover, a 2021 study⁷ also found that some measured markers (eg, prenatal skin thickness, nasal bone length) were significant to trisomy 21 screening. However, these markers need sonographers' elaborate annotations and measurements in ultrasonographic low-resolution images. Therefore, a better artificial intelligence (AI) approach should be explored to screen for trisomy 21 accurately.

In the past decade, AI based on machine learning has captured much attention in the field of medical image analysis owing to its encouraging findings in cancer prediction and screening.^{9,10} Recent advances for quantitative medical image analysis using deep learning (DL) methods, particularly convolutional neural networks (CNNs), have shown remarkable performance, such as classifications based on computed tomography images^{11,12} and prognoses based on magnetic resonance imaging¹³⁻¹⁵ and computed tomography images.¹⁶⁻¹⁸ In the field of ultrasonographic image analysis, previous studies have reported remarkable breakthroughs using CNNs, such as the diagnoses and classifications for breast and liver cancers.¹⁹⁻²¹ However, whether an end-to-end DL network model can capture discriminative features automatically to accurately facilitate the screening of fetuses for trisomy 21 remains unknown.

In this study, we focused on the challenge to investigate whether a noninvasive DL model could screen fetuses for trisomy 21 based on ultrasonographic images. Therefore, we hypothesized that DL model would be able to screen fetuses for trisomy 21 accurately.

Methods

This diagnostic study received ethical approval from each participating institution's institutional review board. The requirement of informed consent from patients was waived because it was deemed urgent to collect clinical data for our study. This study is reported following the Standards for Reporting of Diagnostic Accuracy (STARD) reporting guideline.

This diagnostic study used retrospective data. The primary outcome was detection of fetuses with trisomy 21. We proposed a shallow CNN, named Trisomy21Net, to develop a DL model. A flowchart of the DL model for ultrasonographic images is shown in eFigure 1 in the [Supplement](#). We assessed the performance of model by receiver operating characteristic (ROC) curves.

Collection and Enrollment of Multicenter Data Set

We enrolled all available cases and controls at the Department of Ultrasound at the Beijing Obstetrics and Gynecology Hospital between January 2009 and February 2019 and Shijiazhuang Obstetrics and Gynecology Hospital between April 2018 and September 2020. We validated controls for euploidy by documented neonatal examination. All ultrasonographic images were digitally stored in the hospital information systems. Voluson E8 (GE), Voluson E10 (GE), WS80A (Samsung), and HS70A (Samsung) ultrasonography machines were used for the acquisition of ultrasonographic images. We formulated 4 criteria for inclusion. First, we selected 2-dimensional ultrasonographic images of the midsagittal plane of the fetal face in the first trimester at more than 11 weeks and less than 14 weeks of gestation. Second, the fetus NT was shown to and measured by 3 certified

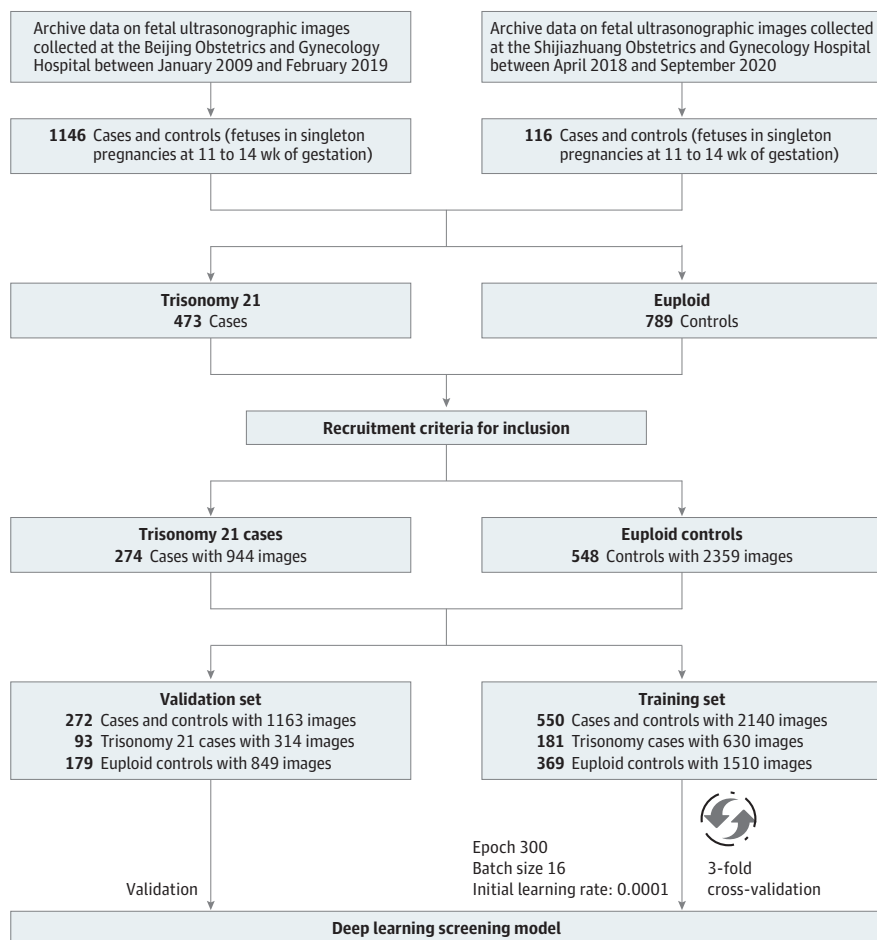
sonographers (Y.S., C.S., and Q.W.). Two sonographers annotated the NT markers. If there was disagreement for annotation, they discussed with the third sonographer for a final agreement. Third, the criterion standard of fetal karyotypes for each case in this study was confirmed. Fourth, we included only fetuses with complete clinical data (maternal age and the measurement of NT marker). Finally, 822 cases with 3303 ultrasonographic images (548 euploid fetuses with 2359 images and 274 fetuses with trisomy 21 with 944 images) were selected according to the recruitments. The flowchart of data collection is shown in **Figure 1**.

NT measurements were performed by trained and certified sonographers based on the International Society of Ultrasound in Obstetrics and Gynecology practice guidelines.⁸ The measurement precision of ultrasonographic machines was 0.1 mm. A fetal sagittal section was obtained first, then it was magnified to show only fetal head and upper thorax. For a standard ultrasonographic image, the echogenic tip of the nose and rectangular shape of the palate were in anterior. The translucent diencephalon was in the center, and the nuchal membrane was in posterior. Calipers were placed correctly (on-on) to measure NT as the maximum distance between the nuchal membrane and the edge of the soft tissue overlying the cervical spine. If more than 1 measurement meeting all the criteria were obtained, the maximum measurement was recorded.

Image Segmentation

For original ultrasonographic images, some basic information is recorded at the edge of the image, which may cause poor performance of the screening model. Moreover, some studies have

Figure 1. Flowchart of Multicenter Data Sets Collection for Construction of Deep Learning Model



Two ultrasonographic image data sets were collected from Beijing Obstetrics and Gynecology Hospital and Shijiazhuang Obstetrics and Gynecology Hospital. Strict recruitment criteria for inclusion were formulated. Finally, 550 cases and controls with 2140 images were selected for training set, and 272 cases and controls were selected as validation set.

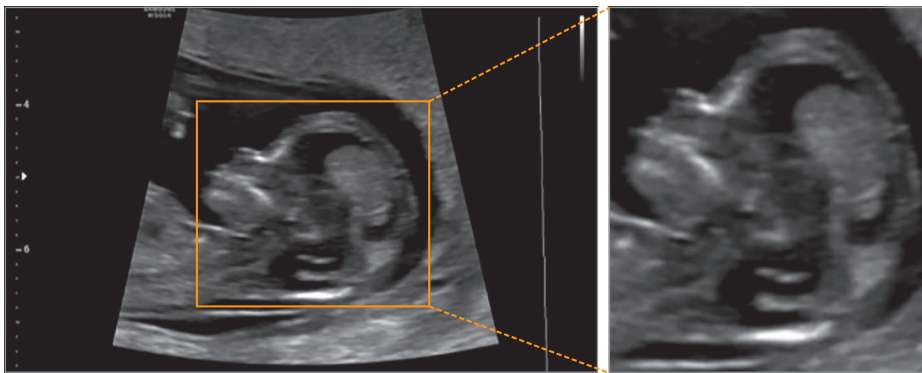
demonstrated that many markers observed from fetal head position are significant in screening for trisomy 21.⁷ Therefore, we only focused on the fetal head region to train our model. Each original ultrasonographic image was segmented by a bounding box instead of delineating the boundaries by sonographers. Examples of segmented sonographic images used are shown in **Figure 2**. We double-checked all the tailored images to ensure complete heads were included. A data augmentation strategy was applied to generate more training data with different forms of transformation from the existing images.¹⁸ We used Keras DL software for data augmentation, including flipping, transformation, rotation, scaling, and cropping.

Model Construction

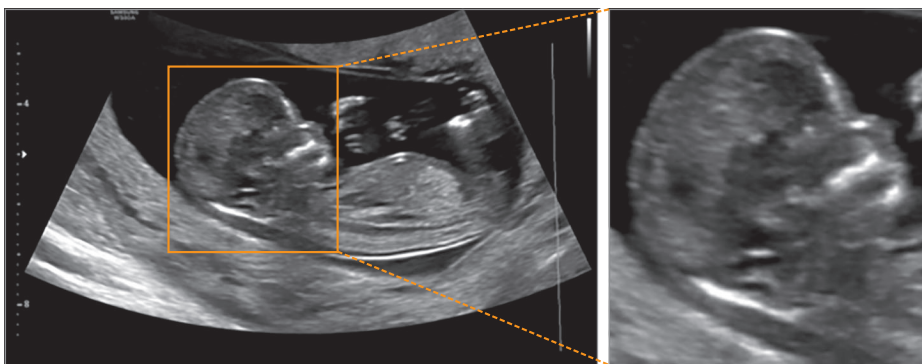
We propose a shallow Trisomy21Net DL model with 11 layers. The input of our model was a predefined size of 224 224 pixels. As shown in eFigure 1 in the **Supplement**, our model consisted of a self-defined residual block to extract different levels of features.²²⁻²⁴ The feature maps were visualized to show discriminative information that the model focused on (eFigure 1 in the **Supplement**). The drop out method was applied to avoid overfitting. The random initialization method was adopted. For the hyperparameter setting, we implemented the Keras DL library, with the TensorFlow machine learning library (Google) as the backend. We also used an initial learning rate of 0.0001 and batch size of 16 for each iteration of 300 epochs. The Adam optimizer based on the Keras library, with a default parameter was applied. We defined the probability value of our model as the risk score for each fetus. The range of risk scores was from 0 to 1. A higher risk score represented that a fetus was at higher risk of having trisomy 21. Meanwhile, binary cross entropy was used to train our model. In the process of training, we exploited the strategy of dynamically adjusting learning rate to get the best trained model.

Figure 2. Illustration of the 2-Dimensional Ultrasonographic Images Segmentation of Fetal Heads

A First of 2 randomly selected 2-dimensional ultrasonographic images with segmentation of fetal head



B Second of 2 randomly selected 2-dimensional ultrasonographic images with segmentation of fetal head



We randomly selected 2 fetuses to show the process of segmentation. The fetal heads were our regions of interest. The orange bounding boxes on ultrasonographic images were regions of fetal heads, which were defined as regions of interest as inputs for the deep learning model.

Model Comparison and Visualization

To show the superiority of our DL model, we constructed a model with just NT markers, $Model_{NT}$, and a model with NT markers and maternal age, $Model_{NT+age}$, in the same training set of 550 cases and validation set of 272 cases. These cases and controls were divided into training and validation data sets in a 2:1 ratio using an approach of simple random sampling. We also constructed a combined model ($Model_{DL+age}$) using DL risk scores in combination with maternal age to investigate whether maternal age could be a covariable to boost the performance for screening. The 4 models were evaluated using AUC, accuracy, sensitivity, and specificity.

To further interpret the DL model in a human-readable form, we used a class activation map (CAM) technique to shed light on what the model focused on and how it explicitly enabled the CNN to learn discriminative features for risk scores.^{25,26} Therefore, we visualized response regions of our model to produce different localization maps visualized by CAM in images from 2 perspectives: (1) randomly selecting several cases, visualizing the region of interest (ROI), and (2) visualizing the self-learned multilevel (6 levels) features from shallow to deep layers. Our code for proposed Trisomy21Net has been made available in the Github repository.²⁷

Statistical Analysis

We used R software version 4.0.4 (R Project for Statistical Computing) for the statistical analysis. ROC curves were depicted to evaluate the performance of the model. We also calculated the 95% CIs for numerical results.²⁸ The result was considered statistically significant when a 2-sided *P* value was less than .05. We used the *t* test or Mann-Whitney *U* test for continuous variables and the χ^2 test or Fisher test for categorical variables, as appropriate. Data were analyzed from March 1, 2021, to January 3, 2022.

Results

Clinical Characteristics

A total of 822 case and control participants (mean [SD] age, 31.9 [4.6] years) were enrolled in the study. There were 550 participants (mean [SD] age, 31.7 [4.7] years) in the training set and 272 participants (mean [SD] age, 32.3 [4.7] years) in the validation set.

ROC Curve Analysis of the Image-Based Model

We trained our DL model based on 550 participants with 2140 ultrasonographic images (hereafter, $Model_{Image}$). The training set had an AUC of 0.97 (95% CI, 0.97-0.98), and the validation set had an AUC of 0.94 (95% CI, 0.92-0.95) (eFigure 2 in the Supplement). In the training set, the accuracy was 0.92 (95% CI, 0.91-0.93), the sensitivity was 0.90 (95% CI, 0.88-0.92), and the specificity was 0.93 (95% CI, 0.92-0.94). In the validation set, the accuracy was 0.89 (95% CI, 0.87-0.91), the sensitivity was 0.92 (95% CI, 0.94-0.95), and the specificity was 0.76 (95% CI, 0.71-0.80).

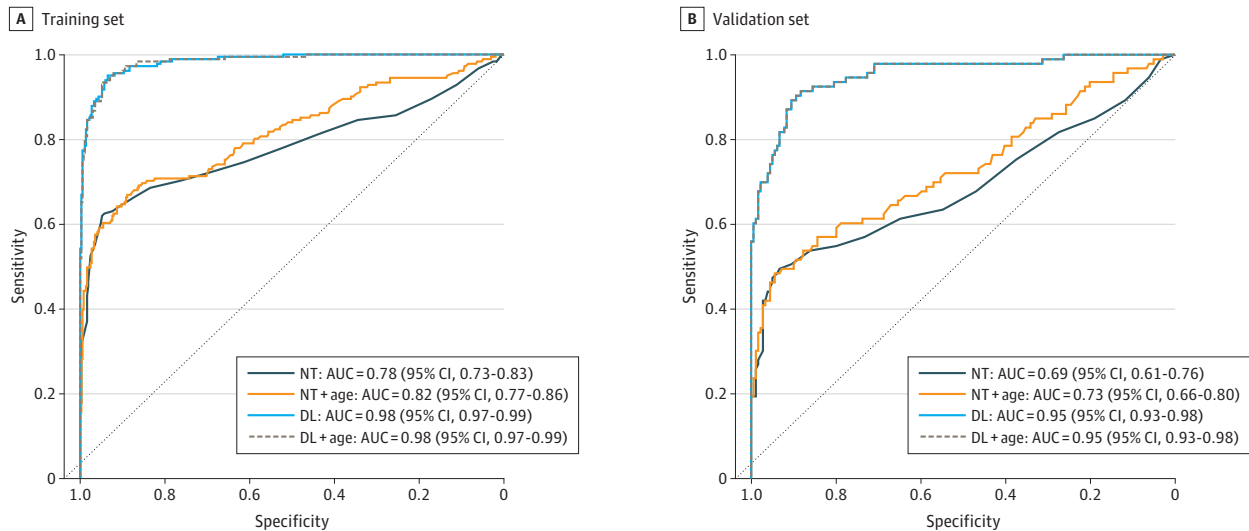
Model Assessment and Comparison for Trisomy 21 Screening

We also constructed the patient-level DL model called $Model_{DL}$ to compare with the fetal NT marker. $Model_{DL}$ was obtained by calculating the mean risk scores of all ultrasonographic images of each fetus estimated by $Model_{Image}$. For comparison, we plotted ROC curves for $Model_{NT}$, $Model_{NT+age}$, $Model_{DL+age}$ in the same training set (Figure 3). As is shown in the Table, we found the consistent results that the NT marker was a significant indicator for trisomy 21 screening in the training set (AUC = 0.78; 95% CI, 0.73-0.83).⁷ However, the performance for NT in the validation set was poor (AUC = 0.69; 95% CI, 0.61-0.76). $Model_{NT+age}$ showed better performance in the training (AUC = 0.82; 95% CI, 0.77-0.86) and validation (AUC = 0.73; 95% CI, 0.66-0.80) sets. $Model_{DL}$ showed the best performance compared with other models in the training (AUC = 0.98; 95% CI, 0.97-0.99) and validation (AUC = 0.95; 95% CI, 0.93-0.98) sets. There were significant differences between $Model_{DL}$ and $Model_{NT+age}$ ($P < .001$).

Interpretation and Visualization for the DL Model

As shown in **Figure 4**, we randomly selected 4 representative cases and controls to show the ROIs that the deep learning focused on. One fetus (ID 44 in Figure 4) was diagnosed by DL model and NT marker as negative for trisomy 21, but positive results were found in the karyotype analysis. Although some cases had a short NT marker that made it difficult to screen fetuses for trisomy 21 by clinical characteristics and visual observation on the ultrasonographic images, the DL model was able to focus on highlights based on the ultrasonographic images. We used the CAM technique in different levels of feature maps to show how the deep learning model focused on the ROIs to screen fetuses with trisomy 21 in (eFigure 3 in the Supplement).²⁶ In the first 5 levels, the visualized localization maps were generated by the DL model with the operation of convolution, which showed where the DL model focused. In the final level, level 6, the most responsive areas were activated by our model. The high-risk scores were estimated by 4 models (Model_{DL}: 0.74; Model_{DL+age}: 0.89; Model_{NT}: 0.72; Model_{NT+age}: 0.67). Inversely, compared with fetuses with trisomy 21, the example of a euploid fetus showed that the highly responsive areas were activated on the region of the forehead. The lower risk

Figure 3. Model Performance Comparisons of Area Under the Receiver Operating Characteristic (AUC) Curves



DL indicates deep learning; NT, nuchal translucency.

Table. Comparisons of Model Screening Performance in Training and Validation Sets

Model performance	Measure (95% CI)			
	AUC	Accuracy	Sensitivity	Specificity
NT^a				
Training	0.78 (0.73-0.83)	0.84 (0.80-0.87)	0.94 (0.91-0.96)	0.62 (0.55-0.70)
Validation	0.69 (0.61-0.76)	0.79 (0.73-0.83)	0.47 (0.37-0.58)	0.95 (0.91-0.98)
NT + age^b				
Training	0.82 (0.77-0.86)	0.82 (0.78-0.85)	0.67 (0.59-0.74)	0.89 (0.85-0.92)
Validation	0.73 (0.66-0.80)	0.76 (0.71-0.81)	0.52 (0.41-0.62)	0.89 (0.83-0.93)
DL				
Training	0.98 (0.97-0.99)	0.94 (0.92-0.96)	0.95 (0.91-0.98)	0.93 (0.90-0.96)
Validation	0.95 (0.93-0.98)	0.88 (0.84-0.92)	0.76 (0.66-0.85)	0.94 (0.90-0.97)
DL + age^c				
Training	0.98 (0.97-0.99)	0.94 (0.91-0.96)	0.95 (0.91-0.98)	0.93 (0.90-0.95)
Validation	0.95 (0.93-0.98)	0.89 (0.84-0.92)	0.78 (0.69-0.86)	0.94 (0.89-0.97)

Abbreviations: AUC, area under the receiver operating characteristic curve; DL, deep learning; NT, nuchal translucency.

^a Model constructed based on fetal NT.

^b Model constructed based on fetal NT and maternal age.

^c Model constructed by DL integrating maternal age.

scores were obtained ($\text{Model}_{\text{DL}}: 0.27; \text{Model}_{\text{DL+age}}: 0.01; \text{Model}_{\text{NT}}: 0.24; \text{Model}_{\text{NT+age}}: 0.18$) (eFigure 3 in the Supplement).

Assessment of Screening Stability and Robustness of the DL Model

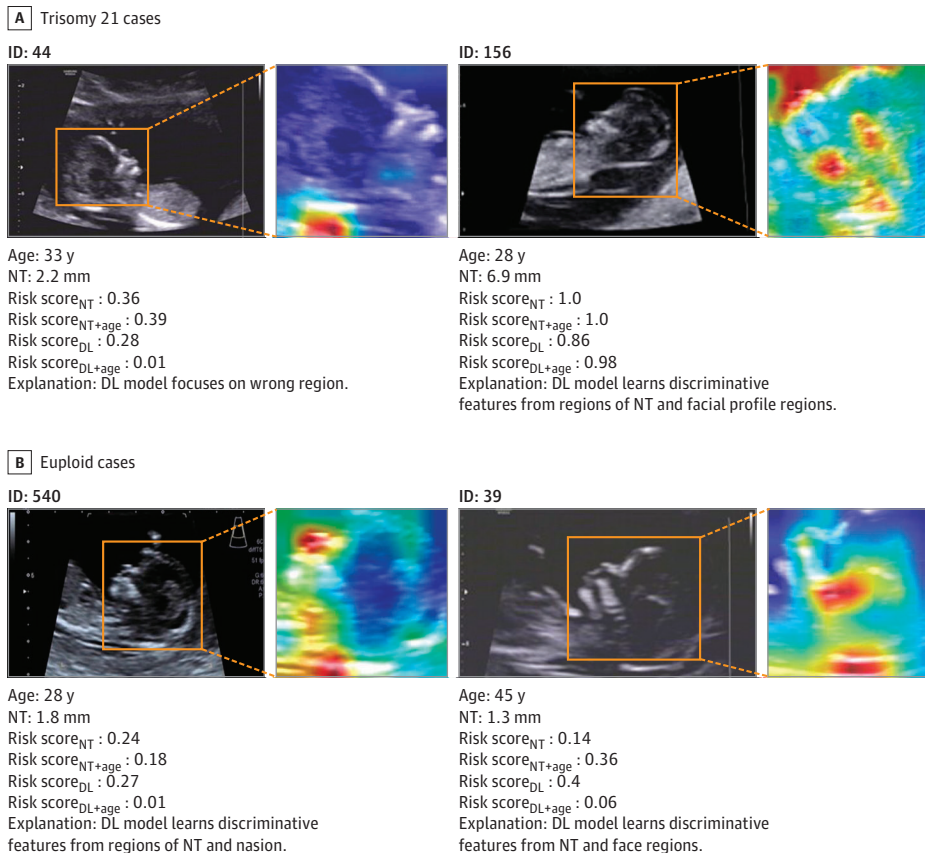
To investigate the potential influence of selection of different training sets in all collected images, we randomly divided the data into 3 equal parts (randomly selected 2 parts as training set and the rest as the validation set) to evaluate the robustness of our model. eFigure 4 in the Supplement presents the model performance in different training and validation sets. For each DL model in either the training or validation set, the AUCs were greater than 0.90. Our results demonstrated that the DL model was robust regardless of the partition of training and validation sets.

Discussion

In this diagnostic study, we constructed a DL model for automatic and accurate screening for trisomy 21. We experimentally demonstrated that the DL model was associated with improved accuracy in screening for fetuses with trisomy 21 compared with existing screening methods based on NT and maternal age. Meanwhile, the assessments of model robustness also demonstrated that our model was robust in first trimester screening for trisomy 21.

Our previous study found that most fetuses with trisomy 21 had thicker NT marker thickness whereas euploid fetuses had thinner NT.⁷ The results of this study indicated that the performance of $\text{Model}_{\text{NT+age}}$ was easily subjected to measurement distribution of fetal NT thickness in training and validation sets. However, the DL model achieved a greater accuracy in the validation set than the $\text{Model}_{\text{NT+age}}$ achieved in the training set, while the $\text{Model}_{\text{NT+age}}$ showed poor performance in the

Figure 4. Visualization of Representative Cases and Controls to Show Focus of the Deep Learning (DL) Model



training and validation sets. The main reason was that the DL model might learn richer information (not limited to the feature of NT thickness) from the fetus's head by operations of pooling and convolution. Additionally, although the unbalanced distribution of sample size existed owing to the scarcity of fetuses with trisomy 21, the results for assessment of screening stability and robustness of the DL model revealed that Model_{DL} was robust to the distribution of unbalanced sample size in training and validation sets. Hence, our method is a potential tool for primary screening to reduce the burden of marker annotation and evaluation on sonographers for all the pregnant people who need screening.

Most studies on trisomy 21 screening mainly analyzed the diagnostic value of 1 or several annotated facial quantitative indicators. Our previous study found that some fetal facial markers, facial angles, and the facial profile line were significant markers in screening for trisomy 21.⁷ However, the manual markers were time-consuming for annotations and influenced by subjective experience for evaluation of these observations. Previous studies have found that a small nose was a common characteristic in fetuses with trisomy 21, and some studies demonstrated that fetal nasal bone length and prefrontal space ratio were the representative markers for trisomy 21 screening.^{29,30} However, these markers were too small to accurately measure them. Our DL model exploited the operations of convolution, pooling, and nonlinear transformations to focus on representative features on the fetal heads. We used CAM to further disclose local focused regions, which revealed that the model might also learn the soft markers, such as choroid plexus cysts or ventriculomegaly. The first 5-level feature maps visualized by CAM could vividly show the process for learning representative features. The CAM applied in the final layers (level 6) could show the visualized response regions for model's decision-making. Our model was able to localize the discriminative regions of nasal bone length and prefrontal space. The visualized examples showed that our DL model focused on different regions to screen euploid fetuses and fetuses with trisomy 21, which is reasonable for a powerful model to recognize markers in a flexible way, like humans can.

Limitations

Our study has several limitations. We used a rough box for cropping to release the burden of segmentation. Further work should be done to investigate more labor-saving and accurate methods of automatic segmentation for screening fetuses for trisomy 21. Our study only focused on fetuses with trisomy 21; future work should investigate multitask learning of the CNN network for fetuses with trisomy 18 and 13 simultaneously. The sample size was limited, so the screening model was dependent on eligible ultrasonographic images. Therefore, further work should be done to train a robust and universally applicable DL model for a large-scale screening. Furthermore, future work also should consider and address the practical issues for real-time recognition of standardized landmarks (eg, round shaped diencephalon, relatively bright occipital bone) prior to archiving of the images. Although our study used pixel normalization to reduce the potential influence of image color difference for model robustness, further work should be done to investigate the potential impact of differences of the color gradient in various ultrasound images from different ultrasonographic machines.

Although we visualized the high response regions with the DL model, we were still unable to quantitatively assess the importance of different features for the decision-making of the model. Further studies will be conducted for visualization techniques to quantitatively evaluate the weight of different features on the results of the DL model.

In clinical practice, some fetuses with trisomy 21 show an NT marker within reference range (0-2.5 mm) and facial profiles that have no obvious anomaly, which may result in incorrect screening outcomes. For examples, a case (ID 44 in Figure 4) was diagnosed by DL model and NT marker as negative for trisomy 21, but positive results were found in karyotypes analysis. Therefore, our method should be further investigated to design a multitask model to explore the performance for prediction of NT marker length, which is meaningful to reduce false-negative rate.

Conclusions

This diagnostic study presents a DL model for screening fetuses with trisomy 21. Our model is a potential tool to improve the primary trisomy 21 screening based on ultrasonographic images for universal clinical application.

ARTICLE INFORMATION

Accepted for Publication: April 19, 2022.

Published: June 21, 2022. doi:[10.1001/jamanetworkopen.2022.17854](https://doi.org/10.1001/jamanetworkopen.2022.17854)

Open Access: This is an open access article distributed under the terms of the [CC-BY License](https://creativecommons.org/licenses/by/4.0/). © 2022 Zhang L et al. *JAMA Network Open*.

Corresponding Author: Qingqing Wu, MD, Beijing Obstetrics and Gynecology Hospital, Capital Medical University, Beijing Maternal and Child Health Care Hospital, No. 251 Yaojia Yuan Rd, Chaoyang District, Beijing 100026, China (qingqingwu@ccmu.edu.cn); Jie Tian, PhD, CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, No. 95 Zhongguancun E Rd, Beijing 100190, China (jie.tian@ia.ac.cn).

Author Affiliations: CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, Beijing, China (Zhang, Dong, Hu, Tian); Zhuhai Precision Medical Center, Zhuhai People's Hospital, Jinan University, Zhuhai, China (Dong); Beijing Obstetrics and Gynecology Hospital, Capital Medical University, Beijing Maternal and Child Health Care Hospital, China (Y. Sun, Wu); Department of Ultrasound, Shijiazhuang Obstetrics and Gynecology Hospital, Shijiazhuang, China (C. Sun); Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, School of Medicine and Engineering, Beihang University, Beijing, China (Tian).

Author Contributions: Dr Tian had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Drs Zhang, Dong, Y. Sun, and C. Sun contributed to this work equally.

Concept and design: Zhang, Dong, Y. Sun, C. Sun, Wu, Tian.

Acquisition, analysis, or interpretation of data: Zhang, Dong, Y. Sun, Hu, Wu, Tian.

Drafting of the manuscript: Zhang, Dong, Y. Sun, Hu, Tian.

Critical revision of the manuscript for important intellectual content: Zhang, Dong, C. Sun, Wu, Tian.

Statistical analysis: Zhang, Hu.

Obtained funding: Dong, Tian.

Administrative, technical, or material support: Zhang, Y. Sun, Hu, C. Sun, Wu.

Supervision: Dong, C. Sun, Wu, Tian.

Conflict of Interest Disclosures: None reported.

Funding/Support: This work was supported by the Ministry of Science and Technology of China (grant No. 2017YFA0205200), National Key R&D Program of China (grant No. 2017YFC1309100), National Natural Science Foundation of China (grant No. 82022036, 91959130, 81971776, 81771924, 6202790004, 81930053, and 8197070119), Beijing Natural Science Foundation (grant No. L182061), Strategic Priority Research Program of Chinese Academy of Sciences (grant No. XDB 38040200), Chinese Academy of Sciences (grant No. GJJSTD20170004 and QYZDJ-SSW-JSC005), Project of High-Level Talents Team Introduction in Zhuhai City (grant No. Zhuhai HLHPTP201703), Special Foundation of State Key Laboratory of Complex Systems Management and Control (grant No. 2022QN03), and Youth Innovation Promotion Association CAS (grant No. 2017175).

Role of the Funder/Sponsor: The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Additional Contributions: We acknowledge the instrumental and technical support of Multi-modal Biomedical Imaging Experimental Platform, Institute of Automation, Chinese Academy of Sciences.

REFERENCES

1. Valetto A, Orsini A, Bertini V, et al. Molecular cytogenetic characterization of an interstitial deletion of chromosome 21 (21q22.13q22.3) in a patient with dysmorphic features, intellectual disability and severe generalized epilepsy. *Eur J Med Genet*. 2012;55(5):362-366. doi:[10.1016/j.ejmg.2012.03.011](https://doi.org/10.1016/j.ejmg.2012.03.011)

2. Kagan KO, Wright D, Baker A, Sahota D, Nicolaides KH. Screening for trisomy 21 by maternal age, fetal nuchal translucency thickness, free beta-human chorionic gonadotropin and pregnancy-associated plasma protein-A. *Ultrasound Obstet Gynecol*. 2008;31(6):618-624. doi:10.1002/uog.5331
3. Norton ME, Jacobsson B, Swamy GK, et al. Cell-free DNA analysis for noninvasive examination of trisomy. *N Engl J Med*. 2015;372(17):1589-1597. doi:10.1056/NEJMoa1407349
4. Snyder MW, Kircher M, Hill AJ, Daza RM, Shendure J. Cell-free DNA comprises an in vivo nucleosome footprint that informs its tissues-of-origin. *Cell*. 2016;164(1-2):57-68. doi:10.1016/j.cell.2015.11.050
5. Kagan KO, Maier V, Sonek J, et al. False-positive rate in first-trimester screening based on ultrasound and cell-free DNA versus first-trimester combined screening with additional ultrasound markers. *Fetal Diagn Ther*. 2019;45(5):317-324. doi:10.1159/000489121
6. Abele H, Wagner P, Sonek J, et al. First trimester ultrasound screening for Down syndrome based on maternal age, fetal nuchal translucency and different combinations of the additional markers nasal bone, tricuspid and ductus venosus flow. *Prenat Diagn*. 2015;35(12):1182-1186. doi:10.1002/pd.4664
7. 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. doi:10.1002/uog.22087
8. Salomon LJ, Alfirevic Z, Bilardo CM, et al. ISUOG practice guidelines: performance of first-trimester fetal ultrasound scan. *Ultrasound Obstet Gynecol*. 2013;41(1):102-113. doi:10.1002/uog.12342
9. Brattain LJ, Telfer BA, Dhyani M, Grajo JR, Samir AE. Machine learning for medical ultrasound: status, methods, and future opportunities. *Abdom Radiol (NY)*. 2018;43(4):786-799. doi:10.1007/s00261-018-1517-0
10. Bi WL, Hosny A, Schabath MB, et al. Artificial intelligence in cancer imaging: clinical challenges and applications. *CA Cancer J Clin*. 2019;69(2):127-157. doi:10.3322/caac.21552
11. Jiang Y, Liang X, Wang W, et al. Noninvasive prediction of occult peritoneal metastasis in gastric cancer using deep learning. *JAMA Netw Open*. 2021;4(1):e2032269-e2032269. doi:10.1001/jamanetworkopen.2020.32269
12. Wang S, Zhou M, Liu Z, et al. Central focused convolutional neural networks: developing a data-driven model for lung nodule segmentation. *Med Image Anal*. 2017;40:172-183. doi:10.1016/j.media.2017.06.014
13. Wu Q, Wang S, Zhang S, et al. Development of a deep learning model to identify lymph node metastasis on magnetic resonance imaging in patients with cervical cancer. *JAMA Netw Open*. 2020;3(7):e2011625-e2011625. doi:10.1001/jamanetworkopen.2020.11625
14. Dong D, Tang L, Li ZY, et al. Development and validation of an individualized nomogram to identify occult peritoneal metastasis in patients with advanced gastric cancer. *Ann Oncol*. 2019;30(3):431-438. doi:10.1093/annonc/mdz001
15. Zhong LZ, Fang XL, Dong D, et al. A deep learning MR-based radiomic nomogram may predict survival for nasopharyngeal carcinoma patients with stage T3N1M0. *Radiother Oncol*. 2020;151:1-9. doi:10.1016/j.radonc.2020.06.050
16. Wang S, Shi J, Ye Z, et al. Predicting EGFR mutation status in lung adenocarcinoma on computed tomography image using deep learning. *Eur Respir J*. 2019;53(3):1800986. doi:10.1183/13993003.00986-2018
17. Wang S, Zha Y, Li W, et al. A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis. *Eur Respir J*. 2020;56(2):2000775. doi:10.1183/13993003.00775-2020
18. Zhang L, Dong D, Zhang W, et al. A deep learning risk prediction model for overall survival in patients with gastric cancer: a multicenter study. *Radiother Oncol*. 2020;150:73-80. doi:10.1016/j.radonc.2020.06.010
19. Ciritis A, Rossi C, Eberhard M, Marcon M, Becker AS, Boss A. Automatic classification of ultrasound breast lesions using a deep convolutional neural network mimicking human decision-making. *Eur Radiol*. 2019;29(10):5458-5468. doi:10.1007/s00330-019-06118-7
20. Zhou H, Wang K, Tian J. Online transfer learning for differential diagnosis of benign and malignant thyroid nodules with ultrasound images. *IEEE Trans Biomed Eng*. 2020;67(10):2773-2780. doi:10.1109/TBME.2020.2971065
21. Wang K, Lu X, Zhou H, et al. Deep learning radiomics of shear wave elastography significantly improved diagnostic performance for assessing liver fibrosis in chronic hepatitis B: a prospective multicentre study. *Gut*. 2019;68(4):729-741. doi:10.1136/gutjnl-2018-316204
22. Lin T-Y, Dollár P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. *arXiv*. Preprint posted online April 19, 2017. doi:10.48550/arXiv.1612.03144
23. Liu S, Qi L, Qin H, Shi J, Jia J. Path aggregation network for instance segmentation. *arXiv*. Preprint posted online September 18, 2018. doi:10.48550/arXiv.1803.01534

24. Zhang L, Dong D, Zhong L, et al. Multi-focus network to decode imaging phenotype for overall survival prediction of gastric cancer patients. *IEEE J Biomed Health Inform*. 2021;25(10):3933-3942. doi:10.1109/JBHI.2021.3087634
25. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. Learning deep features for discriminative localization. *arXiv*. Preprint posted online September 18, 2018. doi:10.48550/arXiv.1512.04150
26. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: Visual explanations from deep networks via gradient-based localization. *2017 IEEE Int Conference Comput Vis*. 2017:618-626. doi:10.1109/ICCV.2017.74
27. Zhang L. Trisomy21Net. Accessed May 16, 2022. <https://github.com/dreamenwalker/Trisomy21Net/tree/master>
28. Robin X, Turck N, Hainard A, et al. Proc: an open-source package for r and s+ to analyze and compare roc curves. *BMC Bioinformatics*. 2011;12(1):1-8. doi:10.1186/1471-2105-12-77
29. Kagan KO, Cicero S, Staboulidou I, Wright D, Nicolaides KH. Fetal nasal bone in screening for trisomies 21, 18 and 13 and Turner syndrome at 11-13 weeks of gestation. *Ultrasound Obstet Gynecol*. 2009;33(3):259-264. doi:10.1002/uog.6318
30. Yazdi B, Sonek J, Oetting C, et al. Prefrontal space ratio in second- and third-trimester screening for trisomy 21. *Ultrasound Obstet Gynecol*. 2013;41(3):262-266. doi:10.1002/uog.12261

SUPPLEMENT.

eFigure 1. Overview of the Deep Learning Model for Screening Trisomy 21

eFigure 2. ROC Curve of the Image-Level Deep Learning Model

eFigure 3. Visualization of the Response Region to Show the Class-Specific Information for Multilevel Feature Maps

eFigure 4. Assessment of Screening Stability and Robustness of the DL Model